Biophysical and structural characterisation of functional bacterial amyloid secretion systems

Fisentzos Andreas Stylianou

Imperial College London
Department of Life Sciences

Thesis submitted for the degree of Doctor of Philosophy
Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.
Declaration of Originality

I hereby declare that the work reported in this thesis has been performed by myself, except where specifically acknowledged. I also confirm that all sources of information used in this thesis have been indicated, and due acknowledgement has been given to the work of others.
Abstract

Amyloids are characterised by their innate capacity to aggregate into insoluble fibrils, which are commonly recognised for their cytotoxicity and association to neurodegenerative diseases. Their unique physicochemical properties are exploited by bacteria for various functional roles, including the formation of extracellular matrix that is linked to biofilm construction and antimicrobial resistance. The curli fimbriae of *Escherichia coli* was the first functional bacterial amyloid (FuBA) to be discovered. Two operons encode a curli secretion system that is comprised of several distinct proteins including curli-forming subunits, chaperones, amyloid inhibitors, and an outer-membrane (OM) transporter. Amyloid secretion systems enable FuBA fibril formation, whilst minimising their cytotoxicity to the host-cell. A functional amyloid in *Pseudomonas* (Fap) operon was recently identified in the *Pseudomonas* genus encoding a novel FuBA secretion system. Unlike the curli system, detailed insight into the Fap system is lacking. The curli and Fap secretion systems export biochemically similar amyloid forming subunits; however, their FuBA secretion systems are genetically distinct.

The biophysical studies of this thesis sought to provide further insight into the structure and function of FapF, a uniquely structured OM transporter, as well as FapD, a periplasm-residing protein that is predicted to serve a proteolytic and chaperoning role within the Fap system. Previous reports demonstrate the evolutionary co-conservation of FapF and FapD, suggesting their functional co-dependence. In this thesis, a combination of biophysical techniques, including nuclear magnetic resonance (NMR) spectroscopy, are used to demonstrate the presence of a unique, asymmetric, parallel trimeric coiled-coil domain within the periplasmic N-terminus of FapF. Furthermore, a transient interaction between FapF and FapD was identified under solution conditions. Protocols that enable the near-native study of the FapF OM domain by solution-state NMR spectroscopy were also optimised. These studies pave the way for future research to enhance our mechanistic understanding of FuBA secretion systems, with aim to increase our capacity to modulate FuBA formation.
Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Stephen Matthews, for providing me with the opportunity to undertake a Ph.D. under his supervision. His insight and knowledge has been invaluable in assisting me through this research process.

I would also like to thank Dr Ernesto Cota for his scientific advice; Dr Yingqi Xu for his assistance with the technical aspects of NMR spectroscopy; and Dr Marc Morgan for his support with various biophysical techniques. My grateful thanks are extended to Dr Sarah Rouse, Dr Sophie McKenna, Dr Andrea Sauerwein, Dr William Hawthorne, Dr Nessim Kichik, and Dr Joseph Barritt for sharing their scientific insight and experience with me. I am indebted to Sarunas Driezis, Zhihao Wang, Sabine Jegouzo, Grace Wu, and Lee Sewell for their friendship and for creating a relaxed working environment.

I would also like to express my gratitude to the Medical Research Council for their financial support for this project.

I am especially grateful to Dr Mark Pfuhl, Dr Daniel Gutmans, Dr Nicola Pace, and Dr Elena Rostkova, of King’s College London, for providing me with the foundation for my studies in the field of structural biology.

I wish to thank my close friends Luke Whincop, Pedrom Talebi, Lewis Gable, Eessa Mansoor, Christos Panayi, and Christian Liasis for the many years of friendship.

I owe a special thanks to my parents, Sonia and Andy; my sister, Nichola; my future brother-in-law, Matt; my grandparents, Yiayia Kay, Papou Fivo, Yiayia Lulu, and Papou Mihali; and my friends, Teddy Mendonça and Spudley Bennett, for their love, support, and encouragement over the years.

Last but not least, I am especially grateful to my girlfriend Roisin Davis, for her love, for being my best friend, and for cooking the most delicious paellas.
Table of Contents

Copyright Declaration ................................................................. 2
Declaration of Originality ........................................................... 3
Abstract ...................................................................................... 4
Acknowledgements ....................................................................... 5
Table of Contents ......................................................................... 5
List of Figures ................................................................................ 10
List of Tables ................................................................................. 13
List of Abbreviations ..................................................................... 14

Chapter 1 - Introduction ............................................................... 17
1.1. Biofilms .................................................................................. 17
1.2. Functional Bacterial Amyloids .................................................... 20
1.3. Functional Bacterial Amyloid Secretion Systems .......................... 21
1.4. An Introduction to the Curli System .............................................. 22
1.5. Details of the Curli System .......................................................... 23
1.6. An Introduction to the Fap System ............................................... 26
1.7. Details of the Fap System ............................................................. 28
1.8. The FapF Transmembrane Domain: Structure and Function .......... 32
1.9. An Overview of the FapF N-terminal Domains .............................. 35
1.10. FapF and the Secretion Systems .................................................. 36
1.11. An Introduction to the C39 Peptidase Family ................................. 38
1.12. Structures of the C39 Peptidase Domain ....................................... 42
1.13. Substrates of the C39 Peptidase Domain ....................................... 45
1.14. FuBA and Beyond ................................................................. 48
1.15. Outline and Aims of this Thesis .................................................. 48

Chapter 2 – Materials and Methods ................................................ 50
2.1. Materials .................................................................................. 50
2.1.1. Media Components ............................................................... 50
2.1.2. PCR Primers ....................................................................... 51
2.1.3. Peptide Synthesis ................................................................. 52
2.1.4. Acquired Proteins ................................................................. 52
2.1.5. Acquired Plasmids .............................................................. 52
2.1.6. Gene Selection ..................................................................... 53
2.2. Molecular Biology and Cloning .................................................... 53
2.2.1. Ligation-Independent Cloning ............................................... 53
2.2.2. InFusion Cloning .................................................................. 54
2.2.3. Site-Directed Mutagenesis ...................................................... 54
2.2.4. Agarose Gel Electrophoresis ................................................... 55
2.2.5. Gel Extraction ..................................................................... 55
2.2.6. Plasmid Purification ............................................................... 55
2.2.7. Plasmid Sequencing .............................................................. 56
2.2.8. Plasmid Transformation into E. coli BL21 (DE3) ......................... 56
2.3. Bioinformatics Analyses ......................................................... 56
2.3.1. Multiple Sequence Alignment Generation and Secondary Structure Prediction... 56
2.3.2. Predicting the Presence of a Coiled-Coil Domain ....................... 56
2.3.3. Predicting the Oligomeric State of a Coiled-Coil .......................... 56
2.3.4. Signal Peptide Prediction ....................................................... 57
2.3.5. Polyproline Type II Secondary Structure Prediction .................................. 57
2.3.6. Structure Homology Modelling ................................................................... 57
2.3.7. ConSurf ........................................................................................................ 57
2.3.8. CH3Shift ....................................................................................................... 57
2.3.9. Molecular Analysis and Visualisation .......................................................... 57

2.4. Protein Expression and Purification ................................................................. 58
  2.4.1. Recombinant Protein Expression of Unlabelled (no isotopic labelling) FapF, and
        Unlabelled FapD ............................................................................................ 58
  2.4.2. Recombinant Protein Expression of Unlabelled Membrane Scaffolding Protein
        1D1 (MSP1D1) ............................................................................................... 58
  2.4.3. Recombinant Expression of 15N-labelled Protein ........................................ 59
  2.4.4. Recombinant Expression of [85 %-2H, U-15N]-FapFβ .................................. 59
  2.4.5. Recombinant Expression of {U-[1H, 15N], (Ala-β, Ile-δ1)-[13CH3]}-FapFβ ........ 59
  2.4.6. Small Scale Expression Trials ..................................................................... 60
  2.4.7. Purification of FapF 27-64 (including R57A; R57E; Y44A) and FapF 27-81 ...... 61
  2.4.8. Purification of FapD 50-246 (FapDWT) and FapD 50-246 (FapDT) ......... 62
  2.4.9. Purification of MSP1D1 ............................................................................ 62
  2.4.10. Refolding and Purification of FapF 107-430 (FapFβ) .................................. 63
  2.4.11. Detergent Exchange .............................................................................. 64
  2.4.12. Gel Filtration ............................................................................................ 64

2.5. Techniques For Protein Detection and Quantification ..................................... 65
  2.5.1. SDS-PAGE ................................................................................................ 65
  2.5.2. Western Blotting ....................................................................................... 65
  2.5.3. NanoDrop Measurement .......................................................................... 66
  2.5.4. Bradford Assay ........................................................................................ 66

2.6. FapD Stability Assay ....................................................................................... 67

2.7. Nanodisc Assembly .......................................................................................... 67

2.8. Biophysical Techniques .................................................................................. 68
  2.8.1. Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-
        MALS) ......................................................................................................... 68
  2.8.2. Solution-State NMR Spectroscopy .............................................................. 68
  2.8.3. Circular Dichroism (CD) Spectroscopy ..................................................... 68
  2.8.4. Isothermal Titration Calorimetry (ITC) ....................................................... 69
  2.8.5. Dynamic Light Scattering (DLS) ............................................................... 70

2.9. Background of the Biophysical Techniques Used in this Study ...................... 70
  2.9.1. Solution State Nuclear Magnetic Resonance Spectroscopy ......................... 70
  2.9.2. Circular Dichroism Spectroscopy .............................................................. 73
  2.9.3. Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering .... 74
  2.9.4. Isothermal Titration Calorimetry .............................................................. 76

Chapter 3 – Characterisation of the FapF N-Terminal Domains ......................... 78

3.1. Introduction to Coiled-Coils ............................................................................ 78

3.2. Results ............................................................................................................ 82
  3.2.1 FapF is Predicted to Possess a Conserved Coiled-Coil .................................. 82
  3.2.2. Biophysical Techniques Indicate the Presence of a Coiled-Coil at the FapF N-
        terminus ........................................................................................................ 85
  3.2.3. The FapF Coiled-Coil Domain is Asymmetric ........................................... 88
  3.2.4. Expression and Purification of FapF 27-64 .................................................. 93
  3.2.5. Analysis of the FapFcc Backbone Symmetry ............................................. 95
  3.2.6. Y44 is Not Essential for FapFcc Stability .................................................... 96
  3.2.7. FapFcc Possesses a Conserved Trimerisation Motif .................................. 100
  3.2.8. The RhxxhE Motif is Required for FapFcc Trimerisation ............................ 101
  3.2.9. Biophysical Characterisation of the FapF Linker Domain ......................... 104

3.3. Summary and Discussion ............................................................................. 109
  3.3.1. The FapFcc Structure ............................................................................... 109
List of Figures

Figure 1: Bacterial biofilm colonised with *Staphylococcus aureus* ........................................ 17
Figure 2: The biofilm lifecycle. .................................................................................................. 20
Figure 3: Functional bacterial amyloid (FuBA) and the amyloid structure. ......................... 21
Figure 4: The Curli secretion system. ...................................................................................... 26
Figure 5: Conservation of the Fap system. ................................................................................ 27
Figure 6: The Fap secretion system. .......................................................................................... 32
Figure 7: The structure of FapFβ. .............................................................................................. 33
Figure 8: FapFβ is structurally homologous to members of the Type 5a autotransporter family. .......................................................................................................................... 38
Figure 9: FapD is predicted to be a member of the C39-peptidase family. ............................... 39
Figure 10: The GG-motif. .......................................................................................................... 40
Figure 11: The C39 peptidase domains share a similar tertiary fold. ....................................... 42
Figure 12: The cryo-EM structure of the PCAT1-Cta complex. .............................................. 43
Figure 13: The PCAT1 substrate translocation mechanism. .................................................... 44
Figure 14: The C39-peptidase domains recognise a common set of substrate features. ........... 45
Figure 15: Coiled-coil formation is driven by the specific positioning and spacing of hydrophobic and polar residues within α-helices. ................................................................. 80
Figure 16: Coiled-coil geometries. ............................................................................................ 81
Figure 17: FapF multiple sequence alignment. ......................................................................... 83
Figure 18: Coiled-coil predictions of FapF27-64. ....................................................................... 84
Figure 19: Far-UV CD spectrum of the FapF27-64 peptide. .................................................. 86
Figure 20: FapF27-64 is thermostable. ...................................................................................... 87
Figure 21: SEC-MALS confirms the trimerisation of FapF27-64. ............................................ 88
Figure 22: The crystal structure of human lung surfactant protein D (hSP-D) trimeric fragment. ............................................................................................................................... 89
Figure 23: 2D ¹H-¹H NOESY spectrum of FapF27-64. ............................................................. 91
Figure 24: Tyr44 of FapFcc. ........................................................................................................ 92
Figure 25: Schematics of the domain organisation of FapF and the FapF deletion mutants utilised. ............................................................................................................................... 93
Figure 26: Expression trials of FapFcc. ...................................................................................... 94
Figure 27: Purification of FapFcc and Y44A. .............................................................................. 95
Figure 28: The FapFcc backbone possesses three-fold symmetry. .......................................... 96
Figure 29: Y44 assignment ........................................................................................................ 98
Figure 30: Far-UV CD spectrum and melting curve of FapFcc Y44A. .................................... 99
Figure 63: FapFβ precipitate.

Figure 64: The refolding efficiency of FapFβ is influenced by the refolding concentration of FapFβ, and the concentration of LDAO in the refolding buffer.

Figure 65: Refolded FapFβ was obtained under all FapFβ refolding conditions examined.

Figure 66: SEC-MALS analysis of FapFβ in buffer containing 0.1 % LDAO.

Figure 67: 1H-15N HSQC spectrum of FapFβ.

Figure 68: FapFβ detergent exchange.

Figure 69: Far-UV CD spectra of native FapFβ and refolded FapFβ.

Figure 70: 1D 1H-NMR spectrum of partially deuterated FapFβ.

Figure 71: 1H-15N TROSY HSQC spectrum of partially deuterated FapFβ.

Figure 72: CH3Shift chemical shift predictions for FapFβ methyl groups.

Figure 73: 1H-13C HMQC spectrum of {U-[2H,15N], (Ala-β, Ile-δ1)-[13CH3]}-FapFβ in 0.1 % LDAO.

Figure 74: 1H-13C HMQC spectrum of {U-[2H,15N], (Ala-β, Ile-δ1)-[13CH3]}-FapFβ in 0.5 % C8E4.

Figure 75: Comparison of the 1H-13C HMQC spectra of MFapFβL and MFapFβC.

Figure 76: Methyl-labelled FapFβ possesses a high spectral sensitivity.

Figure 77: Nanodisc assembly is a multistep process.

Figure 78: The insertion of FapFβ into MSP1D1-nanodiscs.

Figure 79: DLS confirms the formation of nanodiscs.

Figure 80: 1H-15N HSQC spectra of FapFβ in the presence of 0, 2, and 4 M urea. The addition of 4 M urea destabilises the global conformation of FapFβ.

Figure 81: Copyright permission.
## List of Tables

Table 1: Summary of the GG-motif tolerances

Table 2: The “GG-like motifs” within FapA-F (UK4)

Table 3: Media used for culturing *E. coli*

Table 4: PCR primers used for sub-cloning or mutagenesis
List of Abbreviations

ABC: ATP-binding cassette
AMR: Antimicrobial Resistance
AMS: ABC-Transporter Maturation and Secretion
AT: Autotransporter
ATP: Adenosine Triphosphate
C8E4: Tetraethylene Glycol Monooctyl Ether
CCExt: FapF_{27-81} (trimer)
CD: Circular Dichroism
CF: Cystic Fibrosis
CLA: Co-Lysis Assay
CLD: C39-Like Domain
COPD: Chronic Obstructive Pulmonary Disease
Cryo-EM: Cryogenic Electron Microscopy
DB: Denaturing Buffer
dCTP: Deoxycytidine Triphosphate
dGTP: Deoxyguanosine Triphosphate
DLS: Dynamic Light Scattering
DMPC: 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine
DNA: Deoxyribonucleic Acid
DSF: Differential Scanning Fluorimetry
DTT: Dithiothreitol
ECM: Extracellular Matrix
EDTA: Ethylenediaminetetraacetic Acid
EPS: Extracellular Polymeric Substances
ES: Extracellular Space
Fap: Functional Amyloid in *Pseudomonas*
FapDT: FapD_{50-246}
FapDWT: FapD_{39-246}
FapFcc: FapF_{27-64} (trimer)
FapFL: FapF_{63-106}
FapFB: FapF_{107-430}
FFaP: FapF_{27-64} Peptide
FSS: FuBA Secretion System
FuBA: Functional Bacterial Amyloid
FVC: FapFβ Variable Concentration
GG-motif: Double Glycine Motif
His6-tag: Hexahistidine Tag
HMOC: Heteronuclear Multiple Quantum Coherence
HPLC: High Performance Liquid Chromatography
HSQC: Heteronuclear Single Quantum Coherence
IM: Inner Membrane
ITC: Isothermal Titration Calorimetry
IVSA: In Vivo Secretion Assay
LB: Lysogeny Broth
LDAO: N, N-Dimethyldodecylamine N-oxide
Lol: Lipoprotein Localisation
LVC: LDAO Variable Concentration
MP: Membrane Protein
MS: Mass Spectrometry
MSP: Membrane Scaffolding Protein
MW: Molecular Weight
NBD: Nucleotide Binding Domain
NDB: Non-Denaturing Buffer
Ni-IMAC: Nickel Affinity-Immobilized Metal Affinity Chromatography
NMR: Nuclear Magnetic Resonance
NOE: Nuclear Overhauser Effect
NOESY: Nuclear Overhauser Effect Spectroscopy
NTD: N-Terminal Domain
OM: Outer-membrane
OMF: Outer Membrane Factor
OMP: Outer Membrane Protein
PCAT: Peptidase-Containing ATP-Binding Transporter
PEP: C39-Peptidase
PPII: Polyproline Type II
PRE: Paramagnetic Relaxation Enhancement
RCF: Relative Centrifugal Force
RTX: Repeat-In-Toxin
SCCA: Single-Channel Conductance Assays
SDM: Site-Directed Mutagenesis
SDS-PAGE: Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis
SEC: Size Exclusion Chromatography
SEC-MALS: SEC Coupled to Multiple Angle Light Scattering
T5a: Type 5a Secretion Systems
T5SS: Type 5 Secretion Systems
TCEP: Tris(2-Carboxyethyl)Phosphine Hydrochloride
TEV: Tobacco Etch Virus
$T_m$: Melting Temperature
TMD: Transmembrane Domain
TROSY: Transverse Relaxation Optimized Spectroscopy
WT: Wild Type
Chapter 1 - Introduction

1.1. Biofilms

One of the largest and growing threats to global public health is antimicrobial resistance (AMR), and one of the major contributors to AMR are bacterial biofilms (Stubbings & Labischinski, 2009). Biofilms are surface-attached biological systems comprised of organised and coordinated communities of micro-organisms that are embedded within a slime-like extracellular matrix (ECM) (Figure 1) (Donlan, 2002).

![Figure 1: Bacterial biofilm colonised with *Staphylococcus aureus*. Scanning electron microscopy (SEM) demonstrates the presence of biofilm on the skin lesion in a patient with atopic dermatitis. (A) SEM of a skin lesion; scale bar 100 µm. (B) Extracellular matrix material (indicated by the red arrow) and bacteria (indicated by the blue arrow) are visualised; scale bar 10 µm. (C) Extracellular matrix surrounds the individual bacteria; scale bar 2 µm. Figure modified from Sonesson et al. (2017) under creative commons license CC BY 4.0.](image)

The behaviour of bacteria within the biofilm environment is significantly different to that of planktonic bacteria. Within biofilms, bacteria possess diverse and circumstantial gene expression profiles; assemble and develop scaffolds that form part of a complex architecture; increase communication and co-operation processes; and form symbiotic relationships with various other micro-organisms (Costa et al., 2018; Falsetta et al., 2014; Li and Tian, 2012; Resch et al., 2005). The homeostatic conditions and cooperativity between microbes within the biofilm environment has been likened to the tissues of multicellular organisms that are formed by eukaryotic cells (Costerton et al., 1995). Biofilms are produced by various gram-negative and gram-positive bacteria in response to stress, and that provides bacterial communities with the protection to withstand antibiotics, disinfectants, immune responses, and a range of other physical, chemical and environmental stressors, thus, aiding their persistence that is evolutionarily advantageous (Bridier et al., 2011; Chu et al., 2018; Le et al., 2018; Yin et al., 2019).
The properties of biofilms conferring these protective properties includes: the extracellular polymeric substances (EPS) that adsorb and sieve various small molecules such as antimicrobial agents, thwarting their capacity to penetrate the biofilm; the trapping and build up of high concentrations of virulence factors, β-lactamases and other antibiotic degrading enzymes; the heterogeneity of biofilm bacterial populations, which increases the probability of resistance to various stresses, and includes the small populations of senescent, metabolically inactive bacteria (persister cells) that contribute to AMR; and the architectural and physical properties of the biofilm that provide bacteria with protection from physical and environmental stresses, such as the mechanical stress of water flow, or the host immune clearance mechanisms (Costa et al., 2018; Donlan and Costerton, 2002; Flemming and Wingender, 2010; Graf et al., 2019; Gunn et al., 2016; Hou et al., 2018; Lewis, 2012; Yin et al., 2019).

The majority ecosystems are predominantly comprised of bacterial populations that reside within the biofilm environment; thus, biofilms are associated with commensal bacterial environments, as well as the majority of bacterial-related problems in the clinical, industrial, and environmental sectors (Flemming and Wuertz, 2019; Hall-Stoodley et al., 2004; Rummel et al., 2017; Scofield et al., 2017). Within the clinical sector, bacterial biofilms are often associated with chronic and persistent infection by various types of pathogenic bacteria (Jamal et al., 2018). Some examples include the urinary tract infections of Escherichia coli (E. coli), and the opportunistic infections by Pseudomonas aeruginosa (P. aeruginosa) in cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD) (Eberly et al., 2017; Hassett et al., 2014; Høiby et al., 2010; Jamal et al., 2018). In the case of P. aeruginosa, which infects the lungs of CF patients, biofilms are formed that nullify the effectiveness of antibiotics (Høiby et al., 2010). The reduced effectiveness of antibiotics is in part due to the presence of biofilm persister cells that evade the antibiotics and result in the repeated relapse of P. aeruginosa infection (Pang et al., 2019). Biofilms are also able to colonise medical instruments and devices; thus, in the modern medical era of intravascular catheters and prosthetic heart valves, biofilm related infections are a significant health problem (Khatoon et al., 2018). Biofilm-associated diseases cause hundreds of thousands of deaths each year, and the annual cost of infection is in the billions ($) (Bryers, 2008). Biofilms are also implicated in various non-medical industries; for instance, within the water and food industry, various bacterial biofilms
result in biofouling, food spoilage and pipeline corrosion (Bédard et al., 2016; Galie et al., 2018). As a result, research interest into the bacterial secretion systems that contribute to the formation of biofilm have sprouted, with an increasing drive towards creating novel antimicrobial agents that target bacterial biofilm formation and AMR (Roy et al., 2018).

The process of bacterial biofilm formation involves a series of events (reviewed by Garrett et al. (2008)). To summarise, biofilm formation begins with the reversible attachment of microbial planktonic cells to a conditioned organic or inorganic surface; this step is dependent on energy availability, as well as environmental conditions. Upon the formation of sufficient bacteria-surface interactions, irreversible bacteria-surface adhesion occurs. This subsequently enables the bacterial populations to grow exponentially, and these bacterial populations consume the nutrients in their local environments. Microbial population growth is accompanied by alterations to the microbial gene expression profile. This includes the upregulated expression of EPS components and porins to enable the secretion of various biofilm-forming components; in contrast, the genes responsible for surface appendages and motility are simultaneously downregulated. Resultantly, this increases the adhesive and cohesive properties of the biofilm in order to provide it with additional stability and resistance to potentially harsh environmental conditions. The rate of the microbial population growth within the biofilm environment subsequently enters a stationary phase, during which quorum sensing occurs, and that results in the stimulation of the expression for alginate processing molecules. During the biofilm death phase, the alginate enzymes break down components of the ECM, and bacterial gene expression profiles are modified: motility-associated genes are upregulated, whilst porin-associated genes are downregulated. This enables the release of bacteria from the biofilm, enabling bacteria to colonise new surfaces, facilitating the spread of bacteria to enable the cycle of biofilm formation to continue (Garrett et al., 2008). The biofilm life cycle is illustrated in Figure 2.
Figure 2: The biofilm lifecycle. The main stages of biofilm formation are schematically represented (further details within the text) (Garrett et al., 2008; Khatoon et al., 2018).

1.2. Functional Bacterial Amyloids

The dry mass of a biofilm is predominantly comprised of the EPS that contains a plethora of self-generated sugars, nucleic acids, proteins, and functional bacterial amyloids (FuBA) (Flemming and Wingender, 2010). The biochemical properties of FuBA match the classical amyloid description; FuBA amyloid-forming subunits aggregate, adopting the classical orthogonal stacking of β-strands, to form amyloid fibrils that extend from the bacterial outer-membrane (OM) surface (Figure 3A and 3B) (Otzen, 2010; Otzen and Nielsen, 2008). The amyloid properties of FuBA are reinforced by their reactivity to the Congo red and ThT dyes that are recognised for their ability to bind and detect amyloid fibrils (Otzen, 2010; Otzen and Nielsen, 2008). The FuBA are a major biofilm component, with one study demonstrating that 85 % of the biofilm extracellular matrix is composed of FuBA (Reichhardt and Cegelski, 2014). The FuBA possess unique material properties that are reported to confer multiple important functions for biofilms. This includes the formation of a strong scaffolding matrix, enabling the expansion of bacterial biofilm populations; the
resistance of FuBA fibres to various proteases and denaturing conditions increases their protection to the environment; and the polymerisation of amyloids is a metabolically cheap process, enabling their rapid assembly (Blanco et al., 2012; Collinson et al., 1991; McLaurin et al., 2000; Taglialegna et al., 2016). Functional amyloids are also employed by the various kingdoms of life, including the Eubacteria, Fungi, and Animalia (reviewed by Pham et al. (2014)). For example, functional amyloids play a major role in the protective layer in the eggs of *A. polyphemus*, as well as the spider silk produced by various spiders such as *N. edulis* (Iconomidou et al., 2000; Kenney et al., 2002).

**Figure 3:** Functional bacterial amyloid (FuBA) and the amyloid structure. (A) Transmission electron microscopy image of *Pseudomonas* UK4 and its extracellular Fap fibres. Figure modified from Zeng et al. (2015) under creative commons license CC BY 4.0. (B) Pentameric structure of amyloid-β(1-42). PDB ID: 2BEG (Lührs et al., 2005).

### 1.3. Functional Bacterial Amyloid Secretion Systems

Unlike the unregulated formation of amyloid that is associated with neurodegenerative diseases, the production and assembly of bacterial FuBA incorporates complex secretion systems; these systems involve the cooperation of multiple proteins, the production of which are spatially and temporarily controlled (Barnhart and Chapman, 2006; Brombacher et al., 2006; Evans and Chapman, 2014; Taglialegna et al., 2016). In gram-negative bacteria, the FuBA secretion systems function to ensure that the transport of unstructured FuBA subunits proceeds from the intracellular space, to the periplasmic environment, and finally, to the extracellular environment (where the FuBA subunits self-assemble into amyloid fibrils); this process minimises the potential intracellular cytotoxic effects of the amyloid forming subunits by preventing the formation of amyloid intermediates or misfolded aggregates (Bhoite et al., 2019; Taylor and Matthews, 2015). Various secretion systems exist for gram-negative bacteria, including the well-studied curli system of
enteric bacteria such as that for the *Escherichia* genus, as well as the functional amyloid in *Pseudomonas* (Fap) system of the *Pseudomonas* genus (Barnhart and Chapman, 2006; Dueholm et al., 2010; Rouse et al., 2017). The Fap system is the central focus of this thesis; however, whilst both systems indicate low genetic conservation, the analogous functions between multiple proteins of each system permits the evaluation and comparison of both their FuBA secretion mechanisms (Dueholm et al., 2010; Rouse et al., 2017).

1.4. An Introduction to the Curli System

The curli system from *E. coli* represents the first and most extensively characterised FuBA secretion system (FSS) (Barnhart and Chapman, 2006; Olsén et al., 1989). The secreted FuBA from *E. coli* was named the curli fibre, and its abundance throughout the gram-negative phyla has been demonstrated, with its discovery in *Salmonella, Citrobacter, Enterobacter*, and additional *Escherichia*; various sequence homologs have also been identified in various bacterial species, many of which are opportunistic pathogens (Dueholm et al., 2012; Zogaj et al., 2003). The expression of curli is reportedly constrained to a sub-population of cells located at the biofilm air-colony interface, and that are able to grow under human physiological conditions (Kikuchi et al., 2005; Serra et al., 2013; Wu et al., 2012). Within biofilms, curli fibres provide strong scaffolds to support the formation of biofilm (Hobley et al., 2015; Oh et al., 2016). Furthermore, they have also been demonstrated to facilitate the colonisation and invasion of host tissues by binding to host cell proteins such as fibronectin (Barnhart and Chapman, 2006; Olsén et al., 1989). Beyond their contributions to biofilm formation, curli fibres have been reported to interact with host DNA (deoxyribonucleic acid), and that results in the triggering of auto-immune responses (Tursi et al., 2017). Furthermore, curli has been demonstrated to accelerate host neurodegeneration by cross-seeding with alpha-synuclein amyloid to increase amyloid formation (Chen et al., 2016; Friedland and Chapman, 2017). Since functional amyloids are a novel field of research, many other curli-host interactions, and their effects, may yet to be uncovered.
1.5. Details of the Curli System

The use of *E. coli* as one of the primary models for studying bacterial systems led to the extensive study of the curli FSS, positioning it as a model for the study of FuBA secretion systems (Barnhart and Chapman, 2006; Blount, 2015). The curli secretion system is encoded by a *csg* gene cluster comprised of *csgBAC* and *csgDEFG*, which are two divergently transcribed operons (Chapman et al., 2002; Hammar et al., 1995). Expression of the *csg* gene cluster is highly regulated and spatiotemporally controlled, with gene expression levels being correlated with environmental variables such as temperature, osmolarity, and oxygen (Barnhart and Chapman, 2006; Brombacher et al., 2006). All Csg proteins (except CsgD) contain a Sec secretion signal sequence, enabling their translocation into the periplasm (Barnhart and Chapman, 2006).

Once expressed, CsgA is transported into the periplasm where it is maintained in its unfolded and non-amyloidogenic state by CsgC (a chaperone), prior to secretion into the extracellular space (ES) (Bhoite et al., 2019; Evans et al., 2015). Recent studies by Evans et al. (2015) have indicated that CsgC is an effective inhibitor of CsgA fibrillation, even at low molar ratios of 500:1 of CsgA to CsgC. Of note, this regards the ability for CsgC to prevent the initial assembly of curli fibres, rather than its capacity to disassemble formed fibres. CsgC is also capable of inhibiting CsgB, however, at a significantly lower effectiveness (Evans et al., 2015). The crystal structure of CsgC was solved to reveal an immunoglobulin (Ig)-like β-sandwich structure, comprising an electrostatic surface that is responsible for the inhibition of curli amyloid formation (Taylor et al., 2011, 2016). Hammer et al. (2007) have reported that the nucleation and polymerisation of CsgA into amyloid fibres is initiated by CsgB, a protein that resides at the extracellular surface, functioning as a template for CsgA amyloid formation. Both CsgA and CsgB share a 30% sequence identity; this includes multiple highly conserved amyloidogenic asparagine and glutamine rich repeat sequences, for which CsgA contains five, and CsgB contains four (Hammer et al., 2007). Furthermore, the amyloid repeating units for each protein are comprised of similar oligopeptide sequences to those found in prion proteins such as PrP and Sup35 (Cherny et al., 2005). The crystal structure of the CsgA repeats, as well as the secondary structure predictions of CsgA and CsgB, suggest that CsgA and CsgB share the conserved β-sheet and turn based secondary structure that is able to
aggregate to form an amyloid cross-β structure (Barnhart and Chapman, 2006; Perov et al., 2019). Interestingly, both CsgA and CsgB are capable of self-associating, albeit with altered fibre properties (Shu et al., 2012). As well as a Sec-signal sequence, CsgA contains an N-terminal 22-amino-acid sequence (N22) (Robinson et al., 2006). The biogenesis of curli fibres was hindered by the removal of the N22 sequence from CsgA, as this sequence is required for recognition by CsgG (an outer membrane pore protein) for export (Robinson et al., 2006; Yan et al., 2020).

Transcription of the csgBAC operon is dependent on activation by csgD, a FixJ/LuxR family transcriptional regulator that is also involved in the modulation of various biofilm-related components; this includes the suppression of genes responsible for flagellum production and rotation, as well as the upregulation of the genes involved in the synthesis of cellulose (a major polysaccharide component of the biofilm ECM) and microbial adhesion factors (Brombacher et al., 2006; Flemming and Wingender, 2010; Ogasawara et al., 2011). Expression of csgD is higher at low temperatures (<32 °C), low osmolarity, and low growth rates; however, expression at 37 °C has also been reported (Brombacher et al., 2006).

CsgG is a lipoprotein localised to the bacterial outer membrane (OM) (via the lipoprotein localisation (Lol) pathway) that forms an ungated secretion channel to enable the exit of unfolded Csg amyloid substrates into the ES (Goyal et al., 2014; Robinson et al., 2006). The structure of CsgG was solved by Cao et al. (2014) to reveal that it forms an oligomer of nine CsgG subunits, with each monomer comprised of four β-strands and three α-helices; together, this forms a 36-stranded open transmembrane β-barrel assembly. The entrance (eyelet) of CsgG is constricted to approximately 12 Å due to the presence of 9 loops that are pointed towards the β-barrel centre. Together, these loops form a ring that prevents substrates larger than secondary structure elements such as β-strands from traversing the pore; thus, the eyelet constriction is consistent with the secretion of unfolded, soluble amyloid subunits. The CsgG periplasmic face is comprised of three α-helices, with helix α1 and α3 present on the outer surface, and helix α2 lining the inner surface, the latter of which is suggested to drive the formation of the CsgG nonamer (Cao et al., 2014).

CsgE resides within the periplasmic space, and deletion of CsgE within in vivo experiments has been reported to decrease the secretion of CsgA into the ES (Chapman et al., 2002). CsgE is reported to be a weak inhibitor of CsgA amyloid fibril formation, and that is effective at molar ratios of 1:1; however, the significantly
lower inhibition efficiency of CsgE compared to CsgC suggests that CsgA amyloid fibril inhibition is not the main function of CsgE (Evans et al., 2015; Nenninger et al., 2011). The structure-and-function analysis of CsgE has demonstrated that it differentially binds to both CsgA and CsgG (periplasmic face), functioning as a CsgA chaperone (Goyal et al., 2014). Interestingly, CsgE shares a similar structure to the CsgG monomer, and the interaction between CsgG and CsgE is reportedly driven by electrostatic interactions; this involves the negatively charged tail of CsgE and the positively charged periplasmic interface of CsgG (that includes CsgG helix α2) (Shu et al., 2016). Furthermore, CsgE is also capable of forming oligomers that are able to cap the periplasmic inlet of CsgG, a mechanism that is suggested to prevent the translocation of non-specific substrates (Goyal et al., 2014, Shu et al., 2016).

CsgF is translocated into the ES via CsgG, and it resides at the cell surface. CsgF functions as an assembly factor, and it is reportedly essential for the efficient polymerisation of curli fibre subunits, as well as the anchoring of curli fibres to the bacterial cell surface (Nenninger et al., 2009). Recent insight into the CsgF interactions from pulldown-assays and curli secretion assays has revealed that the N- and C-terminus of CsgF possess distinct functions, with the former binding to CsgG, and the latter binding to CsgB (CsgA nucleator) (Schubeis et al., 2018; Yan et al., 2020). An overview of the curli system is illustrated in Figure 4.
Figure 4: The Curli secretion system. The Csg components are translocated into the periplasm via the Sec pathway (Barnhart and Chapman, 2006). CsgC inhibits CsgA amyloid formation, and CsgE chaperones CsgA to CsgG for translocation into the extracellular space (Evans et al., 2015; Goyal et al., 2014). CsgB and CsgF are also translocated into the extracellular space via CsgG (Schubeis et al., 2018; Yan et al., 2020). CsgF is an assembly factor that differentially associates to CsgG and CsgB (a CsgA nucleator), enabling extracellular CsgA fibrillation to occur, while remaining anchored to the bacterial cell surface (Schubeis et al., 2018; Yan et al., 2020). See text for further details. CsgG PDB ID: 3X2R (Cao et al., 2014).

1.6. An Introduction to the Fap System

Functional amyloid in Pseudomonas (Fap) was first identified in the Pseudomonas sp. UK4 strain (later taxonomically labelled as P. fluorescens UK4; referred to as “UK4”), and that was first discovered within a biofilm that was present on a glass slide submerged within a water reservoir (Dueholm et al., 2010, 2014). Fap was later
discovered in various bacteria, encompassing over 40 genera and over four bacterial phyla (Figure 5) (Rouse et al., 2018a). This includes various pathogenic and non-pathogenic strains of *Pseudomonas*, such as *P. putida* and *P. protegens*, which are typical soil microorganisms, with the latter being exploited for bioremediation purposes (Rouse et al., 2018a; Wilkes and Aristilde, 2017). Similarly, *P. sp. UK4* is commonly found within soil and plant niches, and has potential use for exploitation as a biocontrol agent (Morrison et al., 2016). Within humans, reports have suggested that *P. fluorescens* may also contribute to the pathogenesis of inflammatory bowel diseases such as Crohn’s disease (Scales et al., 2014). The Fap system is also reported for the *P. aeruginosa* species, which commonly infects patients with CF, and this species also has a major association to various hospital acquired infections due to its antibiotic resistance (Pachori et al., 2019).

Figure 5: Conservation of the Fap system. The phylogenic tree indicates the conservation and distribution of Fap proteins across various orders. The presence of *fapA-F* within the various orders is indicated by the green-filled circles, and the order text coloured red corresponds to the sole presence of *fapD* and *fapF*. Figure taken from Rouse et al. (2018a) under creative commons license CC BY 4.0.

Unlike the two divergently transcribed operons of the curli system, the Fap system is comprised of a single *fapABCDEF* operon, suggesting simpler regulation mechanisms for Fap gene expression in comparison to the curli system (Dueholm et al., 2010). Transcriptional activity of the Fap operon was demonstrated to be optimal at high cell densities, temperatures at 30 °C, and NaCl concentrations at 200 mM;
transcription was also demonstrated under human physiological conditions (Dueholm et al., 2013b).

The Fap fibres possess a range of functions within biofilms. The overexpression of the Fap operon was demonstrated to significantly increase the formation of biofilms, even upon transplantation of the operon into E. coli (Dueholm et al., 2010). Furthermore, Fap fibres have been demonstrated to increase the adhesiveness and strength of biofilms (Dueholm et al., 2013b; Zeng et al., 2015). Zeng et al. (2015) have reported that Fap amyloid expression results in an over 20-fold increase in biofilm hydrophobicity and stiffness, likely owing to the strong amyloid material and the scaffold forming properties of the amyloid fibrils. Additionally, Fap amyloid was shown to increase the dehydration resistance of cells within biofilms (Zeng et al., 2015). Fap fibrils have also been demonstrated to transiently interact with quorum-sensing (QS) molecules and redox mediators; thus, retaining these molecules within proximity of the cells in conditions of high turbulence, such as water streams (Seviour et al., 2015). In addition, Herbst et al. (2015) have demonstrated that changes to the P. aeruginosa PAO1 phenotype occur upon its expression of amyloid fibrils. This includes the downregulation of virulence-associated genes, and the upregulation of genes responsible for alginate production. Interestingly, these changes led to a mucoid phenotype that mimics the phenotype observed for isolates of P. aeruginosa in chronic infections (Herbst et al., 2015). Although evidence for the direct role of Fap in human infection is limited, there is evidence to suggest that it is likely. For instance, upregulated transcription of the Fap operon occurs within murine models for infections stemming from acute burn and chronic surgical wounds, and that are commonly caused by P. aeruginosa (Turner et al., 2014). Furthermore, Fap amyloids were identified in corneal infection- and implant-based murine models (Kim et al., 2016). Together with the knowledge that various pathogenic species encode the Fap system, including P. aeruginosa, B. cepacia, B. gladioli, and the many more, suggests that the association between Fap fibrils and pathogenesis is only beginning to become unravelled (Rouse et al., 2018a).

1.7. Details of the Fap System

With the exception of FapA, all Fap proteins have been demonstrated to be vital for Fap fibre formation. Similar to the curli system, all Fap components contain an N-
terminal Sec-secretion signal to enable their translocation into the periplasm via the Sec-machinery (Dueholm et al., 2013b). FapC is the predominant component of the Fap fibres, and it contains conserved asparagine and glutamine rich repeat sequences that are postulated to correspond to the amyloid-core forming residues (Dueholm et al., 2013a, 2013b). The conservation of these residues is highlighted by the finding that CsgC is able to inhibit FapC fibril formation (Taylor et al., 2016). However, unlike CsgA, FapC possesses a conserved C-X-X-C motif (C corresponds to cysteine; X corresponds to any amino acid) at its C-terminus, with reports indicating that this motif is not incorporated into the core of the fibre, and it is instead postulated to stabilise interfibre interactions (Dueholm et al., 2013b; Rouse et al., 2018a). In addition, the width of the Fap fibre core is estimated at 4.5 nm, that is considerably larger than that the 3.0 nm width estimated for CsgA (Rouse et al., 2018a). Furthermore, the FapC N-terminus is comprised of 37 residues that do not share homology with the CsgA N22 sequence (Otzen, 2010). A FapF-targeting sequence (FapF is the OM transporter protein of the Fap system) is yet to be discovered (Otzen, 2010; Rouse et al., 2017).

FapB is secreted into the ES, and it possesses sequence homology with FapC; FapB possesses a high sequence conservation with three postulated amyloid repeat regions of FapC. However, the length of the residue sequences between the amyloidogenic repeats in FapB are reportedly shorter than those in FapC, and the variability of this region is also reduced compared to FapC (Dueholm et al., 2013a). FapB is also present within the Fap fibre at relatively low abundance compared to FapC (Dueholm et al., 2013b). Together, this suggests that FapB may function as a FapC nucleator, and/or a FapC tethering protein, akin to CsgB and/or CsgF, respectively (Dueholm et al., 2013b; Van Gerven et al., 2015).

Similarly, FapE is secreted into the ES where it is found present within the Fap fibre at relatively low abundance (compared to FapC) (Dueholm et al., 2013b). Furthermore, FapE also possesses a number of asparagine and glutamine residues within its C-terminus, akin to the amyloidogenic repeats of FapB and FapC (Rouse et al., 2017). Therefore, FapE may also function analogously to CsgB and/or CsgF. Interestingly, mass spectrometry (MS) fingerprinting indicated that FapE was the only member of the Fap system to demonstrate a truncation (of its N-terminus), suggesting that proteolytic processing of this terminus may occur (Rouse et al., 2017). However, the function of the FapE N-terminus remains to be elucidated.
FapA and FapD are predicted to be periplasmic proteins that possess chaperoning and/or regulatory roles, akin to CsgC and/or CsgE (Dueholm et al., 2010, 2013b). FapA and/or FapD are predicted to bind to Fap substrates (FapA/B/E) in order to maintain their unfolded states and enable their secretion into the ES (Dueholm et al., 2013b). In FapA knockouts, FapB replaces FapC as the dominant subunit of the Fap fibre, suggesting that FapA may function as a specific chaperone for FapC (Dueholm et al., 2013b).

FapD does not share any homology to any curli Csg components, and instead, shares sequence homology with the C39-peptidase family of cysteine proteases (Dueholm et al., 2013b). Interestingly, the substitution of the postulated active site cysteine to alanine was reported to abolish the secretion of FapC within an in vivo FapC secretion assay, suggesting that FapD is an active protease (Rouse et al., 2017). The C39-peptidase domains are commonly present at the N-terminus of ATP (adenosine triphosphate)-binding cassette (ABC)-transporter systems, functioning to recognise and cleave prepeptides at a distinctive double glycine motif (GG-motif) (Havarstein et al., 1995). None of the Fap components contain the complete canonical GG-motif; however, variability of this motif has been demonstrated, and it remains plausible that FapD recognises, binds, and cleaves variants of the GG-motif that are potentially present within one or more Fap proteins, processing them for secretion (Bobeica et al., 2019; Furgerson et al., 2008; Rouse et al., 2017). Alternatively, inactive variants of the C39-peptidase domain exist, named C39-like domains (CLDs). The CLDs possess structural homology with the C39-peptidases and have the capacity to interact with substrates to guide their translocation across their respective ABC-transporter and into the ES (in the absence of a cleavage event) (Lecher et al., 2012). Intriguingly, a recent study demonstrated that FapD and FapF associate; therefore, FapD may function akin to CsgE, chaperoning and facilitating the unfolded states of Fap substrates, via their “GG-like motifs”, for subsequent translocation into the ES via FapF (Chorev et al., 2020; Dirix et al., 2004; Goyal et al., 2014).

FapF forms the OM protein responsible for Fap subunit translocation into the ES (Rouse et al., 2017). The structure of the FapF C-terminal transmembrane domain (TMD) was recently solved (FapF107–430; FapFβ) by Rouse et al. (2017); FapFβ corresponds to a truncated form of FapF that does not possess its periplasmic N-terminal domains. Unlike CsgG, the FapF TMD consists of a 12-stranded trimeric β-barrel system, for which each β-barrel possesses a ~12 Å diameter pore, and that is
plugged by a short 13 residue α-helical plug. Multiple assays have indicated that full length FapF exists in an open conformation, suggesting that the N-terminus of FapF is involved in the regulation of FapF pore activity; this regulation may involve interactions between FapF and various Fap components (Rouse et al., 2017). Furthermore, a 38 residue long coiled-coil structure was solved for the FapF N-terminus (FapF_{27-64}), and the presence of a 42-residue predicted linker domain conjoins the N- and C-terminal domains (Rouse et al., 2017, 2018b). Of note, the structure of the FapF coiled-coil domain will be described in later sections of this thesis. Interestingly, a recent ecology study of the Fap genes indicated that gene homologs of *fapF* and *fapD* are frequently colocalised, unlike the other Fap genes, suggesting that FapF and FapD may function co-dependently to transport Fap substrates across the periplasm and into the ES (Rouse et al., 2018a). A model of the Fap system is illustrated in Figure 6. Both FapF (UK4) and FapD (UK4) are subjects of this thesis.
The Fap components are translocated into the periplasm. FapA and FapD may chaperone, process, or recognise FapB, FapC, FapE, and FapF, functioning to facilitate the translocation of unfolded Fap substrates through the FapF pore and into the ES (Dueholm et al., 2013b; Rouse et al., 2017). Within the ES, FapB and FapE may function as assembly factors and/or nucleators, akin to CsgB and CsgF, to enable FapC fibrillisation to occur whilst remaining anchored to the bacterial cell surface (Chorev et al., 2020; Dueholm et al., 2013b; Van Gerven et al., 2015). The structure and function of many Fap proteins remain to be elucidated. See text for further details.

The structure of FapFβ and FapF27-64 are shown connected by a hand-drawn linker domain (grey). FapFβ PDB ID: 5O65 (Rouse et al., 2017); FapF27-64 PDB ID: 6FUE (Rouse et al., 2018b).

1.8. The FapF Transmembrane Domain: Structure and Function

Structural insight into the components of the Fap system is limited. As previously described, two crystal structures for FapF were published: the FapF C-terminal transmembrane domain (PDB ID: 5O65); and the FapF N-terminal periplasmic coiled-coil (PDB ID: 6FUE) (Rouse et al., 2017, 2018b). Rouse et al. (2016) had initially attempted to solve the structure of full length FapF by X-ray crystallography, however, these attempts were stifled due to the presence of a predicted 42 residue N-
terminal periplasmic flexible linker domain. Limited proteolysis was subsequently employed to produce a stable truncation of FapF, FapF$_{107-430}$ (FapFβ) (Rouse et al., 2016). The structure of FapFβ was solved by X-ray crystallography, as previously described. Interestingly, the lumen of each FapFβ β-barrel is occluded by FapF$_{113-125}$, a 13-residue N-terminal α-helix termed the “plug”, and that is postulated to correspond to a closed state of FapF (Figure 7B and C) (Rouse et al., 2017). Of note, Rouse et al. (2017) confirmed that the trimer state of FapFβ represents the native oligomeric state of full length FapF by use of native mass spectrometry (MS) and single-channel conductance assays (SCCA).

**Figure 7: The structure of FapFβ.** (A) The domain boundaries of FapF, as approximated by Rouse et al. (2017), are illustrated as a schematic. Each domain is labelled according to its position in the amino acid chain of FapF, and is individually coloured; the predicted coil domain is coloured green, the variable linker is coloured grey, the helical plug is coloured purple, and the β-barrel is coloured blue. The SEC signal peptide is abbreviated to “SS” and coloured yellow (Rouse et al., 2017). (B) The structure of FapFβ (cartoon illustrated) reveals a 12-stranded β-barrel protein with a lumen that is plugged by a 13-residue α-helix (coloured red). The side view is illustrated. (C) Top-down view of the FapFβ structure, as described in (B). (D) The residues responsible for positioning and stabilising the α-helix within the β-barrel lumen are shown as sticks and labelled. (E) The side view of FapFβ is illustrated (grey) and the PTG motifs FapF$_{257-259}$ and FapF$_{283-285}$ are coloured red and green, respectively (Rouse et al., 2017). FapFβ PDB ID: 5O65.
The FapFβ plug was subjected to a detailed analysis by Rouse et al. (2017), and a summary of their findings is presented below. The FapFβ plug resides at the periplasmic inlet of FapF, for which the C-terminal end is positioned deeply within the lumen. A hairpin turn occurs at the Phe126 and Phe127 site of the plug, positioning the N-terminus of the plug towards the periplasmic space. The formation of the hairpin structure occludes the FapF pore, effectively preventing the translocation of molecules larger than water through the pore, as confirmed by molecular simulations (Rouse et al., 2017). Furthermore, experimental evidence produced by Rouse et al. (2017), in the form of in vitro conductance assays and in vivo FapC secretion assays, also demonstrated that the FapFβ pore exists in a closed state. The plug is stabilised within its respective β-barrel lumen by forming multiple interactions with residues of the interior face of the lumen (Figure 7D). This includes the cation-pi interactions that are formed between F127 and R181 that are both conserved amongst the Pseudomonas. A FapFβ F127A substitution mutant was demonstrated to restore the secretion of FapC within an in vivo secretion assay (IVSA), suggesting the importance of this interaction in maintaining the plug position within the FapFβ closed-state. In addition, the salt bridge formed between E122 and K425 is reported to be evolutionarily conserved; however, the substitution of either of these residues to alanine did not restore FapC secretion within the in vivo secretion assay, suggesting that this interaction may not be essential for plug-lumen tethering. Within the FapFβ lumen, E135 and R181 are two conserved residues that are able to form an interstrand salt bridge. Rouse et al. (2017) suggest that the localisation of the helical-plug may be dependent on the rearrangement of the electrostatic interactions between E122, K425, E135, and R181. The substitution of E135/R181 to an oppositely charged residue restores FapC secretion within the IVSA, suggesting that these residues possess an important role in tethering the plug to the FapFβ lumen (Rouse et al., 2017).

Another structural feature of FapFβ includes the presence of two conserved “PTG” motifs (FapF257-259; FapF283-285) that are reported to induce a twist on the base of a major extracellular facing loop (Figure 7E) (Rouse et al., 2017). The function of this motif is currently unknown, and its importance is questionable considering that mutagenesis of this motif did not disrupt the secretion of FapC within the IVSA (Rouse et al., 2017). However, it remains plausible that it possesses a structural role within FapF, or a role in anchoring the translocating Fap substrates to FapF (possibly
akin to CsgF) (Rouse et al., 2017; van Den Berg et al., 2015; Yan et al., 2020). Alternatively, it has been suggested to function as a lateral gate for the uptake of hydrophobic substrates, that could include quorum sensing molecules (Seviour et al., 2015; van Den Berg et al., 2015).

Lastly, the driving force for the translocation of Fap proteins through FapF and across the OM is yet to be elucidated. Like other OM transporters, FapF does not utilise any recognised external source of energy for transport (Klauser et al., 1993; Rouse et al., 2017). Rouse et al. (2017) have postulated that the Fap substrates form electrostatic interactions with the charged periplasmic face of FapFβ that initiates their threading into the hydrophilic β-barrel lumen; the subsequent translocation of the substrate from the β-barrel lumen to the ES is postulated to be entropically-driven by extracellular escape (Kang’ethe and Bernstein, 2013; Rouse et al., 2017; Van den Broeck et al., 2015).

1.9. An Overview of the FapF N-terminal Domains

The N-terminal periplasmic domains of FapF were lacking from the FapFβ crystal structure, however, the FapF_{27-64} crystal structure was later solved to reveal the presence of an asymmetric parallel trimeric coiled-coil (Rouse et al., 2017, 2018b). The structure and biophysical properties of the FapF coiled-coil were partially characterised by the work presented in this thesis, therefore, it will be described in detail throughout later chapters of this thesis. Instead, a brief introduction to the prior structural and functional characterisation of the FapF N-terminal domains (NTDs), that includes the coiled-coil and a linker domain, will be provided here.

In a previous study by Rouse et al. (2017), various biochemical and physiological properties of FapF and FapFβ were probed. This included their oligomeric state(s), pore conductivity, and their capacity to secrete FapC fibres; to examine each property, native MS, SCCA, and IVSA were employed, respectively. Firstly, native MS studies indicated that FapF (incorporated into detergent micelles) exists predominantly as a trimeric species. In contrast, the native MS results for FapFβ demonstrated the presence of an equilibrium of monomer-dimer-trimer mixtures at a 3:2:1 ratio, suggesting that the FapF NTDs are essential for stabilising the FapF trimer state. Secondly, the SCCA indicated that upon the application of an increasing voltage, FapF exhibited a proportional increase in ion-conductance, whereas FapFβ did not
produce any conductance changes. This suggests that FapF possesses an ion-conducting pathway (an unoccluded pore), and that FapFβ corresponds to a closed state. Lastly, the IVSA indicated that substitution of FapF for FapFβ within the fap operon was sufficient to abolish FapC secretion. Together with the conductance assays, this suggests that the FapF lumen is likely to exist in an open, unplugged state compared to FapFβ, and that the open state of FapF is dependent on, or regulated by, the presence of its periplasmic NTDs (Rouse et al., 2017).

Further insight into the dimensions and accessibility of the predicted flexible linker domain (LD) of FapF65-106 were provided by ion-mobility measurements (again, these studies were performed by Rouse et al. (2017)). These measurements indicated that FapF possesses a large cross-sectional area, and that corresponded to models of FapF containing extended conformations of the LD. Furthermore, the LD was shown to be amenable to trypsin digestion, indicating that the LD is likely to be solvent exposed, reinforcing the suggestion that it exists in an extended and/or flexible conformation. Rouse et al. (2017) suggested that the extended FapF conformation would likely span the full depth of the periplasm, potentially extending beyond, or interacting with the periplasmic peptidoglycan layer (Rouse et al., 2017). Despite the fact that both the Fap system and the curli system function to secrete FuBA into the ES, the structure and function studies for both FapF and CsgG contrastingly indicate that each OM protein operates in a unique way (Cao et al., 2014; Rouse et al., 2017).

1.10. FapF and the Secretion Systems

A range of secretion systems are employed by gram-negative bacteria to transport proteins across the inner- and outer-periplasmic membranes. The membrane proteins responsible for transporting proteins across the membrane are categorised and referred to as type I through type IX secretion systems (Bhoite et al., 2019; Green and Mecsas, 2016). Each category of membrane proteins is constructed based on various properties that includes their size, transport mechanism, and oligomeric state, to name a few (reviewed in Green and Mecsas (2016)) (Bhoite et al., 2019). FapF shares structural characteristics with the type 5 secretion systems (T5SS) (Rouse et al., 2017). In comparison to other secretion systems, the T5SS are considerably smaller, span only the outer-membrane, and do not utilise any recognised external source of energy for transport; this led to their naming as the “autotransporters” (ATs) (Green
and Mecsas, 2016; Klauser et al., 1993). The T5SS consist of two distinct domains; the OM inserted β-barrel domains (also known as β-domains), as well as the passenger domain(s) (also known as α-domains) that are transported into the ES (reviewed in Meuskens et al., 2019).

The structure, size, and domain organisation of members of the T5SS varies significantly, and the T5SS were subsequently divided into five sub-classes, named T5a through T5e (reviewed in Henderson et al., 2004) (Fan et al., 2016). The T5SS are translocated into the periplasm via the Sec pathway, and their β-barrel domains are subsequently inserted into the OM via the BamA complex. This enables the translocation and/or cleavage (autocatalytic in some cases) of their respective passenger domain, releasing it across the OM and into the ES (Fan et al., 2016). The FapFβ structure has previously been indicated to share structural features with the T5a subclass, that includes the presence of a C-terminal 12-stranded β-barrel domain, as well as the presence of an N-terminal α-helix that occludes the β-barrel pore in their respective closed states (Fan et al., 2016; Rouse et al., 2017). This was made evident by a previous study, in which a high degree of structural homology was demonstrated between the crystal structure of FapFβ and EspP (post autocatalytic cleavage of its passenger domain), a member of the type 5a subclass (Figure 8) (Barnard et al., 2007; Rouse et al., 2017). Of note, the bonding interactions responsible for stabilising the “helical plug” within the β-barrel lumen of the type 5a subclasses appears to vary protein-to-protein. For instance, hydrogen bonds are reported to dominate those for the helical plug of EspP, whereas, a mix of hydrogen bonds and salt bridges are responsible for stabilising the “helical plug” of NalP (Barnard et al., 2007, 2012; Oomen et al., 2004). In contrast to the type 5a secretion systems, the FapFβ helical plug forms a loop within the β-barrel lumen, and the N-terminus of the plug exits on the periplasmic FapFβ inlet, whereas, the N-terminus EspP “plug” faces towards the outlet/ES (Barnard et al., 2007; Rouse et al., 2017). Furthermore, FapF does not possess a passenger domain; instead, it is comprised of a periplasmic N-terminal extended linker domain and a coiled-coil domain (Rouse et al., 2017). FapF is also natively trimeric, whereas, the T5a ATs are monomeric (Fan et al., 2016; Rouse et al., 2017). The structural and functional properties of FapF are novel features that, together, do not fall within any of the current categorised secretion systems (Green and Mecsas, 2016; Rouse et al., 2017).
Figure 8: FapFβ is structurally homologous to members of the Type 5a autotransporter family. Two structures are available for the EspP protein: 3SLT (PDB ID) (blue), which corresponds to the full-length protein; and 2QOM (PDB ID) (green), which corresponds to a processed form. The reported backbone root-mean-square deviation (RMSD) resulting from superimposition of FapFβ and EspP (3SLT) is below 4 Å (Rouse et al., 2017). For clarity, all proteins are depicted in cartoon form; to highlight the plug domain, the β-barrel for each protein is transparent; and a single chain of FapFβ (chain B) is illustrated (Barnard et al., 2007, 2012; Rouse et al., 2017).

1.11. An Introduction to the C39 Peptidase Family

FapD shares sequence homology with the C39-peptidase family of cysteine proteases that belong to the clan CA, however, the structure and precise function of this protein remains to be elucidated (Barrett and Rawlings, 2001; Dueholm et al., 2013b; Rouse et al., 2017). In this section, a brief background on the C39-peptidase family will be provided. The C39-peptidase family, and protease families in general, are categorised in the MEROPS database based on their sequence homologies (Rawlings et al., 2014). Each family is labelled and numbered according to their catalytic residue type; for example, the “C” of the “C39-peptidase” family corresponds to a protease that contains a catalytic cysteine residue. Protease clans are comprised of one or more families that demonstrate conserved tertiary structures and/or conserved sequence homology of the catalytically active residues, as well as the sequence motifs that flank these residues. Two letters are used to identify each clan; the first letter corresponds to the catalytic residue type (Barrett and Rawlings, 2001; Rawlings et al., 2014).

The origin of the clan CA stems back to the isolation and characterisation of papain from *Carica papaya* (papaya latex), that is a cysteine protease designated to family C1 (Barrett and Rawlings, 1996; Konno et al., 2004). Papain forms the foundation of the CA clan that later grew to encompass an additional 43 families,
including the well-studied calpains (family C2) and the caspases (family C14) (Barrett and Rawlings, 2001). Members of the C39-family possess a conserved active site tetrad comprised of proximal glutamine, cysteine, histidine, and asparagine or aspartate; Cys, His, and Asp form the catalytic triad, and the Gln is postulated to form an oxyanion hole (Barrett and Rawlings, 1996; Kieuvongngam et al., 2020; Lin et al., 2015). Mutagenesis of the catalytic Cys or His residues result in a complete loss of C39-proteolytic activity, indicating their fundamentally important catalytic roles (Furgerson et al., 2008; Ishii et al., 2006). Interestingly, aligning the sequences of FapD and the C39-peptidase LahT150 indicates that FapD possesses the conserved active site residues (Figure 9).

![Multiple sequence alignment](image)

**Figure 9: FapD is predicted to be a member of the C39-peptidase family.** The multiple sequence alignment of FapD and a C39-peptidase domain, LahT150 (LahT), reveals conservation of the canonical C39-peptidase active site residues within FapD (Q70, C76, H160, F161; indicated by an asterisk) (Bobeica et al., 2019). The residues are numbered according to the FapD UK4 sequence. The sequence similarity is coloured according to a percentage equivalence, that accounts for the physicochemical properties for each residue; residues highlighted red are completely conserved, and residues coloured red are highly conserved. The secondary structure for FapD UK4 is also predicted and presented above the alignment; the coil, arrow, and gaps correspond to an α-helix, β-sheet, or no secondary structure, respectively. The predicted signal sequence of FapD UK4 is highlighted in blue. The conserved residues that are within 4 Å of the substrate bound to LahT147 (PDB ID: 6MPZ) are indicated by a black arrow (Bobeica et al., 2019).
The C39-peptidase family are defined by their ability to recognise and cleave at GG-motifs, the consensus and highly conserved sequence for which is L-S-X-X-E-L-X-X-I-X-G-G (where X corresponds to any amino acid), and the proteolytic cleavage occurs at the peptide bond immediately C-terminal of the G-G site (Dirix et al., 2004; Furgerson et al., 2008). The GG-motif was originally discovered by in silico analysis; mutagenesis and structural studies later confirmed the importance of this motif for its recognition and cleavage by the C39-peptidases (Bobeica et al., 2019; Dirix et al., 2004; Furgerson et al., 2008; Kieuvongngam et al., 2020). Of note, the nomenclature regarding the substrate cleavage site positions are designated between P1-P1’, with amino acids in the N-terminal direction being incremented as P2, P3, P3, and so on, whilst the C-terminal direction is incremented as P2’, P3’, P4’ and so on (Schechter and Berger, 1968). Regarding the nomenclature for residues of the GG-motif, residues N-terminal of the P1-P1’ site are denoted X-1 (P1), X-2 (P2), X-3 (P3) and so on, for which X is the corresponding residue (Figure 10) (Bobeica et al., 2019).

<table>
<thead>
<tr>
<th>Residue</th>
<th>L</th>
<th>S</th>
<th>X</th>
<th>X</th>
<th>E</th>
<th>L</th>
<th>X</th>
<th>X</th>
<th>I</th>
<th>X</th>
<th>G</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>-12</td>
<td>-11</td>
<td>-10</td>
<td>-9</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
</tbody>
</table>

Figure 10: The GG-motif. The canonical residues of the GG-motif are indicated, along with their respective positions N-terminal to the C39-peptidase cleavage site. Residue “X” corresponds to any amino acid (Dirix et al., 2004).

The C39-peptidase domains are located at the N-terminus of ABC-transporter maturation and secretion (AMS) proteins or peptidase-containing ATP-binding transporters (PCATs) (Bobeica et al., 2019; Havarstein et al., 1995; Lin et al., 2015). In gram-positive bacteria, the AMS/PCATs are comprised of an ABC transporter. The ABC-transporters are composed of two identical chains that form a symmetrical dimer. Each chain is comprised of an N-terminal C39-peptidase domain, a transmembrane domain (TMD) (formed of six helices), and a C-terminal nucleotide binding domain (NBD) (Kanonenberg et al., 2013; Kieuvongngam et al., 2020). The NBDs function to bind and hydrolyse ATP, resulting in conformational changes throughout the ABC-transporter that drive the translocation of unfolded (or partially structured) substrates across the OM and into the ES (Linton and Higgins, 2007). In gram-negative bacteria, the ABC transporter is conjoined to a membrane fusion protein (MFP) that forms a tubular barrel “tunnel-like” structure across the periplasm, and that is connected (associated) to the outer membrane protein (referred to as the
outer membrane factor (OMF)). Together, this complex forms the T1SS (Green and Mecsas, 2016; Kanonenberg et al., 2013).

Three groups of the T1SS ABC-transporters exist, and that are categorised by their distinct differences in their N-terminal sequences (Green and Mecsas, 2016). As mentioned above, one of these groups contains an N-terminal C39-peptidase (PEP) domain, which are known as the AMSs/PCATs (Bobeica et al., 2019; Green and Mecsas, 2016). The PEP domains function to recognise and cleave the GG-motif (also referred to as the leader sequence) between the P1-P1’ site of an unfolded cargo prepeptide (precursor peptide) (Havarstein et al., 1995; Lin et al., 2015). Subsequently, the mature substrate is released from the PEP domain in the direction of the TMD of the ABC-transporter (Kieuvongngam et al., 2020). The mature peptides are commonly members of the class II family of bacteriocin precursors (C2BPs; <10 kDa); cleavage of their prepeptides results in their activation, and it also enables their subsequent translocation into the ES via the T1SS machinery (Cotter et al., 2005; Nishie et al., 2011; Van Der Meer et al., 1993). To date, the exact role of the substrate leader peptides remains to be elucidated; however, studies have suggested that it may function to protect the host organism from the activity of the substrate, such as that of the C2BPs (Furgerson et al., 2008; Sahl and Bierbaum, 1998; Van Der Meer et al., 1993).

The second class of AMSs/PCATs possess N-terminal C39-like domains (CLDs) that are structurally homologous to the C39-peptidase domains, but lack their proteolytic activity due to the deficiency of a catalytic cysteine residue (Green and Mecsas, 2016; Lecher et al., 2012). Lecher et al. (2012) reported that despite the lack of catalytic activity, the CLDs remain essential for the secretion of their respective substrates. Lecher et al. (2012) also reported that the CLDs retain the capacity to recruit and tether unfolded substrates (that do not contain GG-motifs) for their secretion into the ES, as well as prevent the aggregation and/or degradation of unfolded substrates. Compared to the PEP domains, CLDs transport large repeat-in-toxin (RTX) substrates that are significantly larger (≥55 kDa) than the C39-peptidase substrates (Lecher et al., 2012). However, the differences between the CLDs and PEP domains extend beyond the loss of the catalytic cysteine residue. Interestingly, the introduction of a cysteine residue in the HlyB-CLD, at the location predicted for a cysteine to be present within the PEP domains, did not induce any PEP domain-like catalytic activity. Also, the substrate binding site of the CLD was found to exist at an
alternative site to the C39-peptidase (Lecher et al., 2012). Of note, the third class of T1SS ABC transporters does not contain any additional domains at their N-termini; their respective substrates contain C-terminal secretion signals that may aid their translocation via alternative mechanisms (Kanonenberg et al., 2013).

1.12. Structures of the C39 Peptidase Domain

The first structures of the C39-peptidase domains and CLDs were those isolated from the full length AMS/PCAT transporters, that includes ComA and HlyB-CLD (Ishii et al., 2010; Lecher et al., 2012). More recently, the structure of LahT147 in complex with its substrate (interacting via the GG-motif) was solved by X-ray crystallography (Bobeica et al., 2019). The tertiary structures for each of the PEP domains are highly homologous, sharing a fold wherein five α-helices typically surround a central region comprised of a six β-strand antiparallel β-sheet (Figure 11) (Bobeica et al., 2019; Ishii et al., 2010; Kieuvongngam et al., 2020).

![Figure 11: The C39 peptidase domains share a similar tertiary fold.](image)

Recently, the crystal structures for full length PCAT1 from Clostridium thermocellum, in both its ATP-bound and ATP-free states, that includes the N-terminal C39-peptidase domains, was solved (Lin et al., 2015). In addition, cryogenic electron microscopy (cryo-EM) was used to solve the structure of PCAT1 in complex with a GG-motif containing substrate (CtA; a 90-residue polypeptide) in the absence of ATP (Figure 12) (Kieuvongngam et al., 2020). Both papers provided substantial
insight into the structural and functional role of the C39-peptidase domains within the AMS/PCATs (Lin et al., 2015; Kieuvongngam et al., 2020).

Figure 12: The cryo-EM structure of the PCAT1-Cta complex. The cryo-EM structure of PCAT1 (illustrated as a cartoon) reveals a transporter consisting of three domains: an α-helical transmembrane domain (coloured orange/yellow), a NBD (coloured pink/magenta), and a C39-peptidase domain (coloured green/cyan). The Cta substrate is coloured blue/cyan, and the Cta GG-motif is depicted as a cylinder (PCAT1 PDB ID: 6V9Z). Figure modified from Kieuvongngam et al. (2020) under creative commons license CC BY 4.0.

In both the structures reported by Kieuvongngam et al. (2020) and Lin et al. (2015), the PCAT1 is a symmetrical dimer whose structure corresponds to that previously described for the AMS/PCAT ABC-transporters. The cytoplasmic N-terminal PEP domains interact with the TMDs and NBDs at opposite sides of the transporter in a symmetrical fashion. In the absence of ATP, the PEP domains of PCAT1 are docked to the lateral openings at cytoplasmic face of the ABC-transporter, and the C39-peptidase catalytic site is positioned facing towards the cytoplasmic inlet (≈10-12 Å diameter opening) of the TMD. The positioning of the PEP domain at this location is functionally optimal, as it guides translocation (exit) of processed substrates towards the TMD channel for secretion. The PEP domain is postulated to be weakly associated to the cytoplasmic face of the TMD, as only a small surface of the PEP domain is buried; this includes polar interactions and van der Waals contacts between the N-terminal loop 1 and helix α2 of the PEP domain to the TMD (Kieuvongngam et al., 2020; Lin et al., 2015). The conformation of the isolated PEP domains was consistent with the PEP domain of PCAT1 (unbound to substrate), suggesting that interaction to the ABC transporter does not trigger any conformational changes within the PEP domain (Lin et al., 2015). In contrast, binding of the Cta substrate to the PEP domain was demonstrated to induce a minor conformational
change in a loop that precedes the catalytic histidine residue. This conformational change was suggested to open a path for the substrate to translocate to the TMD cavity, effectively guiding the substrate for ES translocation (Kieuvongngam et al., 2020). In the ATP-bound form, the PEP domains disengage from the TMD; the substrate translocation mechanism of PCAT1 is outlined and illustrated in Figure 13 (Kieuvongngam et al., 2020; Lin et al., 2015). In the crystal structures for PCAT1, it is unclear whether one independent substrate is translocated, or whether two substrates can be translocated simultaneously (Lin et al., 2015). The structure PCAT1 in complex with Cta indicates that substrates occupy both PEP domains simultaneously, however, only one of the substrates was positioned for proteolytic cleavage by the PEP domain (Kieuvongngam et al., 2020). Furthermore, native MS also indicates that a single substrate (rather than two) is translocated per ATP binding/hydrolysis event of PCAT1 (Kieuvongngam et al., 2020). Intriguingly, the proteolytic activity of the PCAT1 PEP domain is dependent on the presence of full length PCAT1, as the catalytic activity of the isolated PCAT1-PEP domain is reported to be 80% lower than when in complex with the transporter. This was suggested to be due to the role of the TMDs in substrate orientation, however, this remains to be elucidated (Lin et al., 2015). The coupling of PEP domain activity to the full length PCAT may have evolved to minimise the cleavage of toxic prepeptides within the host-cell, thus, only peptides that are destined for translocation are activated (Kieuvongngam et al., 2020; Lin et al., 2015; Van Der Meer et al., 1993).

Figure 13: The PCAT1 substrate translocation mechanism. The postulated translocation mechanism for PCAT1 is illustrated. First, Cta is recruited to PEP domain(s) where it undergoes cleavage at its GG-motif, before being subsequently guided into the TMD cavity. The binding of ATP to the NBD triggers conformational changes throughout PCAT1, resulting in the dissociation of the PEP domains from the TMD and NBD, as well as the release of cargo peptide from the TMD (that is transferred to the ES). Finally, ATP hydrolysis restores PCAT1 to its original conformation, enabling another cycle of substrate translocation to occur. Figure taken from Kieuvongngam et al. (2020) under creative commons license CC BY 4.0.
1.13. Substrates of the C39 Peptidase Domain

The crystal structure of LahT147, a C39-peptidase domain from the PCAT1 family, was solved in complex with a GG-motif containing substrate (Bobeica et al., 2019). This provided the first structural insight for the interaction between a PEP domain and its substrate. The substrate is highly hydrophobic, and it is comprised of disorder, as well as a two-turn helix (helical residues are positions -10 to -4; residue positions are named according to their respective GG-motif positions). The substrate interacts within a hydrophobic groove of LahT147 that is proximal to the catalytic active site (Bobeica et al., 2019). The substrate and substrate binding sites for PCAT1 and LahT147 were later found to be conserved (Figure 14) (Kieuvongngam et al., 2020).

![Figure 14: The C39-peptidase domains recognise a common set of substrate features.](image)

The importance of helicity within the GG-motif was shown, with previous mutagenesis studies indicating that the presence of helix-breaking residues in the leader peptide abolish or significantly reduce the catalytic activity by LctT150 (a PEP domain) (Furgerson et al., 2008). Furthermore, mutagenesis studies also demonstrated the importance of hydrophobic residues within the leader peptide (Kieuvongngam et al., 2020). Kieuvongngam et al. (2020) highlighted the structural and positional similarities between the leucine residues at -7 and -12 of the GG-motif of LahT147 and PCAT1-PEP, demonstrating that they curl onto each other to form a structure described as a “hydrophobic knot” (Figure 14). Mutation of either residue in Cta to alanine reduced the affinity of Cta to PCAT1 by approximately twenty-fold. Also, the
hydrophobic residue at the GG-motif -4 position of the Cta and LahT150 substrate buries its side chain within a hydrophobic pocket of the PEP domain, and that is also important for the PEP-substrate interaction (Bobeica et al., 2019; Kieuvongngam et al., 2020).

The C-terminal Gly-1 and Gly-2 within the leader peptide have also been demonstrated to be essential for the catalytic activity by the PEP-domains. Reports demonstrate that the substitution of Gly-1 to Ala or Ser is tolerated; in contrast, the substitution of Gly -1 or Gly -2 to a charged residue completely nullifies the catalytic activity by PEP domains (Furgerson et al., 2008; Havarstein et al., 1995). With regards to the other residues of the leader peptide that are conserved, the PEP-domain catalytic activity demonstrates some tolerance to their variation. The substrate tolerance of C39-peptidases from multiple independent studies are summarised in Table 1 (Bobeica et al., 2019; Furgerson et al., 2008; Havarstein et al., 1995). It should be noted that the substrate mutants are able to increase or decrease the affinity of the PEP domain to the substrate; thus, the effects of substrate mutations on both catalytic activity, and interaction affinity (to the PEP domain) should be considered (Kotake et al., 2008).

Table 1: Summary of the GG-motif tolerances. The C39-peptidase leader peptide motifs (GG-motifs) that are experimentally determined to be tolerated or non-tolerated for cleavage are shown. Residues coloured blue are derived from the reported substrate tolerances of LcT150; residues coloured purple are derived from the reported substrate tolerances of LahT150; residues coloured black correspond to tolerances reported in both the LcT150 and LahT150 reports (Bobeica et al., 2019; Furgerson et al., 2008). The tolerated serine at G(-1) was previously reported (Havarstein et al., 1995).

<table>
<thead>
<tr>
<th>Residue Position</th>
<th>Canonical Residue</th>
<th>Tolerated Residue(s)</th>
<th>Non-tolerated Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>G</td>
<td>A/S</td>
<td>I/D/K</td>
</tr>
<tr>
<td>-2</td>
<td>G</td>
<td></td>
<td>E/K</td>
</tr>
<tr>
<td>-4</td>
<td>I</td>
<td>T</td>
<td>P/K/D</td>
</tr>
<tr>
<td>-7</td>
<td>L</td>
<td>A/E</td>
<td>K/D</td>
</tr>
<tr>
<td>-8</td>
<td>E</td>
<td>D/K/A</td>
<td>A/P</td>
</tr>
<tr>
<td>-11</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-12</td>
<td>L</td>
<td>V/A</td>
<td>P/K/D/F/W</td>
</tr>
</tbody>
</table>

Relating to the Fap system, this insight is valuable as it suggests that despite the lack of canonical GG-motifs within any of the Fap components (Table 2), FapD may
recognise (and potentially cleave) a variation of the GG-motif, a “GG-like motif.” For instance, unfolded or partially folded Fap substrates could possess a short α-helix and a hydrophobic knot, and that may be recognised by FapD (Bobeica et al., 2019; Furgerson et al., 2008; Kieuvongngam et al., 2020). The GG-like motifs present within the Fap operon of P. sp. UK4 are presented in Table 2. It is interesting to postulate that the GG-like motifs within the FapF linker domain (one of which is somewhat conserved between the Pseudomonas species; FapF	extsubscript{84-95}) represent binding and/or cleavage targets for FapD (Rouse et al., 2017). The cleavage of the FapF linker domain is predicted to induce a conformational change in FapF, generating a FapF conformation akin to FapFβ, thus, resulting in pore closure (Rouse et al., 2017). This cleavage and inactivation mechanism, if proven to be true, may exist to prevent the translocation of non-specific substrates into or out of the host bacterium after Fap fibril formation has occurred (Delcour, 1997; Goyal et al., 2014). Despite these suggestions, it should be noted that, to date, periplasmic C39-peptidases remain functionally and structurally uncharacterised. Therefore, it remains plausible that FapD possesses a unique structure and/or function that differs to the canonical PCAT/AMS associated PEP domains or CLDs.

Table 2: The “GG-like motifs” within FapA-F (UK4). The amino acid sequences displayed include those -12 to -1 (shown left to right) with respect to the GG/GA/GS sites within FapA-F. With reference to Table 1, residues in bold correspond to canonical GG-motif residues; residues coloured green correspond to tolerated residues; residues coloured red are non-tolerated; and for residues coloured black, no data is available to describe their tolerances. Hydrophobic residues at GG-motif positions -4, -7, and -12 are coloured cyan (Dirix et al., 2004; Kieuvongngam et al., 2020). Of note, substitution of residue Glu-8 to Ala was reported to be both tolerated and non-tolerated, however, here it is treated generously as a tolerable residue. Also, residues belonging to the FapF TMD (corresponding to PDB ID: 5O65) are not shown, as this structure is embedded within the OM, and therefore, unlikely to be a target of FapD (Rouse et al., 2017).

<table>
<thead>
<tr>
<th>#</th>
<th>FapA-12</th>
<th>FapB-12</th>
<th>FapC-12</th>
<th>FapD-12</th>
<th>FapE-12</th>
<th>FapF-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NPSQQILRATGG</td>
<td>QINSRAISLGG</td>
<td>FDDTKLNNAGA</td>
<td>AGQMAISAMPGG</td>
<td>NRDPGTCTVGG</td>
<td>GSTVAASSTGG</td>
</tr>
<tr>
<td>2</td>
<td>AAGHASGSSSLGG</td>
<td>NGILGNSQSAAGA</td>
<td>GSSLNSGKNGAGA</td>
<td>AGMAISAMPGGA</td>
<td>VAQGHAIAAPGG</td>
<td>AASSGFGATAG</td>
</tr>
<tr>
<td>3</td>
<td>NTLGNNASLGA</td>
<td>PSNLSIASQSGS</td>
<td>SGNIGTVAGGA</td>
<td>LVEQHTDFSGCA</td>
<td>AGRSVSANSNGG</td>
<td>STVAASSGTTGA</td>
</tr>
<tr>
<td>4</td>
<td>LGNSAGLQAGA</td>
<td>LPDERSPSSTGGS</td>
<td>TKTNNAGANGS</td>
<td>NQQNALQIQGG</td>
<td>AKSPAFASTGS</td>
<td>AQSGALIZEGGS</td>
</tr>
<tr>
<td>5</td>
<td>ERLAKLNMNIA</td>
<td>QVVTSDQATGGS</td>
<td>KGMLGANTGGS</td>
<td>NQQNALQIQGG</td>
<td>AASSGFGTATGGS</td>
<td>AQSGALIZEGGS</td>
</tr>
<tr>
<td>6</td>
<td>SREDFAVGSIGS</td>
<td>QRQSNTLTLGS</td>
<td>GQTSTCDAATGGS</td>
<td>VALNGNIALAGA</td>
<td>TTPQTVQSTS</td>
<td>TGTGVSAGLGS</td>
</tr>
<tr>
<td>7</td>
<td>SGSSLAGQQSGS</td>
<td>SGCTVCAATGGS</td>
<td></td>
<td></td>
<td></td>
<td>AQSGALIZEGGS</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.14. FuBA and Beyond

Further study of the Fap amyloid secretion system will inevitably lead to a plethora of insight into the FuBA secretion systems. This insight can be applied to combatting biofilm formation and AMR. Whilst the design and development of biofilm inhibitors and eradication agents is currently on-going, this field of research remains in its early stages of progress (Romero et al., 2013). Functional amyloid secretion system components may also find their use within the realm of biotechnology. Oxford Nanopore Technologies has exploited the CsgG channel for novel DNA sequencing technologies, and FuBA have recently been utilised for the production of novel nanomaterials (Brown and Clarke, 2016, Carter and Hussain, 2017; Knowles and Buehler, 2011). In order for the potential of Fap FuBA to be accessed, a comprehensive structural and functional characterisation of the Fap system proteins is required.

1.15. Outline and Aims of this Thesis

The aims of this project are listed in order of chapter chronology:

Aims i) Characterising the FapF N-terminal domains

Prior to this project, reports have indicated that the FapF N-terminal domain (NTD) may regulate the pore-open versus pore-closed configuration of FapF. Furthermore, the NTD of FapF is essential for FapF trimerisation, as its removal converts the predominant oligomeric state of FapF from a trimer to a monomer (Rouse et al., 2017). The work in this chapter sought to provide further insight into the structure and function of the FapF NTD, by use of various biophysical methods, in order to further our understanding of its role within the Fap secretion system.

Aims ii) Investigating the structure and function of FapD

The reported structure and function studies of the Fap system suggest a regulatory mechanism for Fap substrate translocation, and that may be driven by FapF-Fap protein interactions (Rouse et al., 2017). FapD is postulated to have a role in this regulation, as a previous ecology study suggested that FapF and FapD might have evolved to function co-dependently (Rouse et al., 2018a). The investigations within this chapter sought to provide further insight into the structure and function of FapD,
as well as the interactions between FapD and FapF, by use of various biophysical techniques.

**Aims iii) Enabling the study of FapFβ by solution-state NMR spectroscopy**

In order to elucidate the pore gating mechanism(s) of FapF, insight into the residue-specific dynamics and exchange-based processes for the various conformational states of FapF are required. This includes the effects of FapF-Fap protein interactions, as well as various Fap mutants. However, this information is not provided by the crystal structure of FapFβ, or any of the previously described assays (Rouse et al., 2017). Instead, solution state NMR spectroscopy represents a suitable technique to investigate and detail these processes (Markwick et al., 2008; Wiesner and Sprangers, 2015). With the eventual aim to elucidate the dynamics and exchange processes of FapFβ (and full length FapF), this study sought to optimise a protocol for the high-yield, cost-efficient production of FapFβ for isotopic labelling, as well as establish an isotopic-labelling strategy that is compatible with the study of FapFβ by NMR.
Chapter 2 – Materials and Methods

2.1. Materials

2.1.1. Media Components

Table 3: Media used for culturing *E. coli*

<table>
<thead>
<tr>
<th>Media</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysogeny broth (LB)</td>
<td>0.5 % w/v yeast extract, 1 % w/v tryptone, 0.5 % w/v NaCl</td>
</tr>
<tr>
<td>LB agar</td>
<td>LB components, 1.2 % w/v agar</td>
</tr>
<tr>
<td>Terrific broth (TB)</td>
<td>12 g/L tryptone, 24 g/L yeast extract, 1 g/L NaCl, 9 g/L Na₂HPO₄, 2.2 g/L KH₂PO₄, 2.6 g/L NH₄Cl, 0.7 g/L Na₂SO₄, 5 g/L glycerol, 2 mM MgSO₄, adjusted to pH 7.4.</td>
</tr>
<tr>
<td>Minimal media (M9)</td>
<td>M9 Salts, 1 g/L ¹⁴NH₄Cl, 4.0 g/L ¹²C-glucose, 1 mM MgSO₄, 10 mM CaCl₂, 1x micronutrient solution 1, 1x micronutrient solution 2, 1x Thiamine/riboflavin solution</td>
</tr>
<tr>
<td>M9 salts</td>
<td>6 g/L Na₂HPO₄, 3 g/L KH₂PO₄, 0.5 g/L NaCl, adjusted to pH 7.5.</td>
</tr>
<tr>
<td>Micronutrient solution 1 (1,000x; stock volume of 50 ml)</td>
<td>0.75 g CaCl₂xH₂O, 1.5 g Na₂EDTA, 1.25 g FeCl₃.</td>
</tr>
<tr>
<td>Micronutrient solution 2 (10,000x; stock volume of 50 ml)</td>
<td>1.2 g CuSO₄xH₂O, 0.9 g MnSO₄xH₂O, 0.14 g ZnSO₄xH₂O, 0.14 g CoCl₂</td>
</tr>
<tr>
<td>Thiamine/riboflavin solution (1,000x; stock volume of 10 ml)</td>
<td>Thiamine 0.1 g, D-riboflavin 0.1 g</td>
</tr>
</tbody>
</table>
2.1.2. PCR Primers

Table 4: PCR primers used for sub-cloning or mutagenesis.

<table>
<thead>
<tr>
<th>Construct or vector or mutant</th>
<th>Primer Direction</th>
<th>Primer Sequence (5’ to 3’)</th>
<th>Cloning Method or Mutagenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>FapF&lt;sub&gt;27-64&lt;/sub&gt; (FapFcc) Construct</td>
<td>Forward</td>
<td>TACTTCCAATCCATGGACGTGGATATC GAAACTCTCAAACAAGAACTGCTG</td>
<td>LIC</td>
</tr>
<tr>
<td>FapF&lt;sub&gt;27-64&lt;/sub&gt; (FapFcc) Construct</td>
<td>Reverse</td>
<td>TATCCACCTTTACTGTTACTGGTCTTC CACCTGCGGTACCCTTTG</td>
<td>LIC</td>
</tr>
<tr>
<td>FapF&lt;sub&gt;27-81&lt;/sub&gt; (CCExt) Construct</td>
<td>Forward</td>
<td>TACCTCCAATCCATGGACGTGGATATC GAAACTCTCAAACAAGAACTGCTG</td>
<td>LIC</td>
</tr>
<tr>
<td>FapF&lt;sub&gt;27-81&lt;/sub&gt; (CCExt) Construct</td>
<td>Reverse</td>
<td>TATCCACCTTTACTGTTACTGGTGAAGTC TGCGGCGGATTTGCCAG</td>
<td>LIC</td>
</tr>
<tr>
<td>FapD&lt;sub&gt;50-246&lt;/sub&gt; (FapDT) Construct</td>
<td>Forward</td>
<td>CATCATCATGGTACCGGCGTATCTTC AAGAAGGTC</td>
<td>Infusion</td>
</tr>
<tr>
<td>FapD&lt;sub&gt;50-246&lt;/sub&gt; (FapDT) Construct</td>
<td>Reverse</td>
<td>AGTGCAGGCCGCAAGCCTAGAAGAAGTCT GCCTGGGATG</td>
<td>Infusion</td>
</tr>
<tr>
<td>pET-28a Vector</td>
<td>Forward</td>
<td>GCTTCGCGCGCAGCTCGA</td>
<td>Infusion</td>
</tr>
<tr>
<td>pET-28a Vector</td>
<td>Reverse</td>
<td>GGTACCATGATGTGATGATGATGTA</td>
<td>Infusion</td>
</tr>
<tr>
<td>FapF&lt;sub&gt;107-430&lt;/sub&gt; (FapFB) Construct</td>
<td>Forward</td>
<td>TACTTCCAATCCAGATGATCCGGAG CCGCACAAGACGTC</td>
<td>LIC</td>
</tr>
<tr>
<td>FapF&lt;sub&gt;107-430&lt;/sub&gt; (FapFB) Construct</td>
<td>Reverse</td>
<td>TATCCACCCTTACTGTTAAGAGTAGTA CGGGAATTTCAGGCTGAAGTGAAG</td>
<td>LIC</td>
</tr>
<tr>
<td>FapDT (C76S) Mutant</td>
<td>Forward</td>
<td>GATTTCCAGCTCTGCTGGCTGCTG</td>
<td>Site-directed mutagenesis</td>
</tr>
<tr>
<td>FapDT (C76S) Mutant</td>
<td>Reverse</td>
<td>GGGTTTCTGTTCCACCAG</td>
<td>Site-directed mutagenesis</td>
</tr>
</tbody>
</table>
2.1.3. Peptide Synthesis

A peptide was synthesised corresponding to \( \text{FapF}_{27-64} \) (97.5 % purity; UK4) by ChinaPeptides. For analysis by circular dichroism spectroscopy, the peptide was dissolved in buffer containing 20 mM \( \text{PO}_4 \), 150 mM \( \text{NaCl} \), pH 6.8 (formulated in \( \text{ddH}_2\text{O} \)); for analysis by NMR spectroscopy, the buffer was formulated in 99.96 % \( \text{D}_2\text{O} \) (VWR).

2.1.4. Acquired Proteins

Purified his6-tagged TEV (tobacco etch virus) protease was kindly provided by Sarunas Dreizis. Purified native FapFβ (UK4) was kindly provided by Dr Sarah Rouse.

2.1.5. Acquired Plasmids

The pNIC28-Bsa4 plasmid was a gift from Opher Gileadi (Addgene plasmid #26103; http://n2t.net/addgene:26103; RRID:Addgene_26103) (Savitsky et al., 2010). The pNIC28-Bsa4 vector N-terminal fusion tag is “MHHHHHHSSGVDLGTELNYFQSM” and this was modified to “MHHHHHHSSGTENLYFQSM” by deletion of residues “GVDL”. The modified pNIC28-Bsa4 vector was kindly provided by Dr William Hawthorne.

The pMMB190-UK4\textit{fapA–F} vector was kindly provided by Professor Otzen (Dueholm et al., 2010).

The pMSP1D1 plasmid was a gift from Stephen Sligar (Addgene plasmid # 20061; http://n2t.net/addgene:20061 ; RRID:Addgene_20061) (Denisov et al., 2004).

The gene sequence encoding FapD\textsubscript{39-246} (FapDWT; UK4), with the addition of an N-terminal fusion tag “MGHHHHHHHGT” (sequence: CCATGGGCCACCACCA-CCACCACCACGRTACC), was synthesised by GenScript and subsequently cloned, using a restriction enzyme-based approach (inserting the gene between the NcoI and BamHI sites), into a pET-28a(-TEV) vector by GenScript.

Constructs corresponding to FapF\textsubscript{27-64} (UK4) mutants (Y44A, R57A, R57E) were kindly provided by Grace Wu.
2.1.6. Gene Selection

All Fap genes utilised throughout this thesis originate from the Fap operon of the *Pseudomonas* sp. UK4 strain. Despite the existence of more clinically relevant *Pseudomonas* strains, such as *aeruginosa*, the UK4 strain was selected for study as its full genome sequence is available, which has led to its classification as the model organism for studies of the Fap system (Dueholm et al., 2014). Furthermore, both structural and functional insight is available for the protein components of the Fap secretion system from *P.* sp. UK4 (Rouse et al., 2017).

2.2. Molecular Biology and Cloning

2.2.1. Ligation-Independent Cloning

Truncations of the FapF (UK4) gene, corresponding to FapF<sub>107-430</sub> (FapFβ), FapF<sub>27-64</sub>, and FapF<sub>27-81</sub> were each sub-cloned from the pMMB190-UK4fapA–F vector and into the modified pNIC28-Bsa4 vector via the ligation-independent cloning (LIC) method (Li and Elledge, 2012; Savitsky et al., 2010). Each respective gene was inserted downstream of the modified pNIC28-Bsa4 vector N-terminal fusion site.

Polymerase chain reaction (PCR) was performed to amplify the gene of interest using KOD Hot Start DNA polymerase (Sigma-Aldrich) according to the manufacturers’ instructions, using the required pair of PCR primers from Table 4. The pNIC28-Bsa4 vector (1 µg) was linearised by BsaI-HF digestion (NEB) according to the manufacturers instructions. The PCR product and linearised vector were visualised by agarose gel electrophoresis. The bands that corresponded to their expected size were extracted (by excision) from the gel and independently purified using the GeneJET Gel Extraction Kit (ThermoFisher) according to the manufacturers’ instructions. The following LIC steps were performed according to those previously described by Hawthorne (2016). The gene insert (50 ng) was treated with 1 unit of T4 DNA polymerase (3,000 units/ml; supplied in NEB buffer 2.1; NEB), 10 mM deoxycytidine triphosphate (dCTP), 2.5 mM dithiothreitol (DTT), 1 mg/ml Bovine Serum Albumin (BSA), and ddH<sub>2</sub>O was used to make up a total reaction volume of 10 µl. The vector (150 ng) was treated with 10 units of T4 DNA polymerase (3,000 units/ml; supplied in NEB buffer 2.1; NEB), 10 mM deoxyguanosine triphosphate (dGTP), 2.5 mM DTT, 1 mg/ml BSA, and ddH<sub>2</sub>O was
used to make up a total reaction volume of 100 µl. The vector and insert mixtures were incubated at 22 °C for 30 minutes, followed by incubation at 75 °C for 20 minutes. After T4 DNA polymerase treatment, a 1:2 molar ratio of vector to insert (1 µl and 2 µl, respectively) were mixed and incubated at 22 °C for 5 minutes. Next, 1 µl ethylenediaminetetraacetic acid (EDTA) (25 mM) was added to the mixture and the sample was incubated for 5 minutes at 22 °C. For transformation, 1 µl of the reaction mixture was added to 50 µl of NEB 5-alpha Competent E. coli (High Efficiency) (NEB5a; NEB) and transformation was performed according to the manufacturers’ instructions. The cell suspension was subsequently plated onto LB-agar containing 50 mg/L kanamycin. The plasmid DNA was extracted and purified from transformed NEB5a colonies by use of a GeneJET Plasmid Miniprep Kit (ThermoFisher) and sequenced.

2.2.2. InFusion Cloning

FapD50-246 (FapDT) was sub-cloned into the pET-28a vector according to the protocols described in the In-Fusion HD Cloning Kit (Takara Bio) (Bird et al., 2014). Polymerase chain reaction (PCR) was performed to independently amplify FapDT from the FapDWT gene, as well as amplify and linearise the pET-28a vector (excluding the FapDWT gene; termed inverse PCR), by using KOD Hot Start DNA polymerase (Sigma-Aldrich) according to the manufacturers’ instructions. The required pairs of PCR primers are presented in Table 4. The PCR product and linearised vector were visualised by agarose gel electrophoresis, and extracted from the gel for purification as previously described. The purified linearised vector (100 ng), purified insert (20 ng), and 2 µl 5x In-Fusion HD Enzyme Premix (Takara Bio) were mixed and ddH2O was used to make up a total reaction volume of 5 µl. The mixture was incubated at 50 °C for 15 minutes, and subsequently placed on ice. For transformation, 2.5 µl of the reaction was added to 50 µl of NEB5a (NEB). The transformation, LB-agar plating, plasmid DNA extraction, and sequencing were performed as previously described for the LIC protocol.

2.2.3. Site-Directed Mutagenesis

Site-directed mutagenesis was performed using the Q5 Site-Directed Mutagenesis Kit according to the manufacturers’ instructions (NEB). PCR was performed using
the Q5 Hot Start High-Fidelity 2X Master Mix, using the required primer pair from Table 4. Half of the PCR product was visualised by agarose gel electrophoresis to confirm the PCR success. If successful, the PCR product (that is not subjected to agarose gel electrophoresis) was subsequently treated by kinase, ligase, and DpnI. The reaction product was transformed into NEB5a, according to the Q5 Site-Directed Mutagenesis Kit Protocol. The transformants were plated on LB-agar, and their plasmid DNA was extracted and sequenced, as previously described for LIC. Of note, the FapF_{27-64} mutants described in this study (Y44A, R57A, and R57E) were all kindly provided by Grace Wu.

### 2.2.4. Agarose Gel Electrophoresis

The separate DNA species of various molecular masses, 6x Gel Loading Dye, Purple (NEB) was mixed with the sample of interest for gel electrophoresis. The composition of the agarose gels were as follows: 1 % (w/v) agarose, TAE buffer (40 mM, Tris-acetate pH 8.0, 1 mM EDTA, and 0.1 µl/ml SYBR Safe (ThermoFisher Scientific). Agarose-gels were submerged in TAE buffer and 100 V was applied for 90 minutes, or until the dye front had sufficiently migrated. In order to estimate sample DNA molecular weights, a 2-Log DNA Ladder (NEB) and/or a Quick-Load 1 kb DNA Ladder (NEB) were used.

### 2.2.5. Gel Extraction

When required, linear DNA was extracted from an agarose gel and purified using the GeneJET Gel Extraction Kit (ThermoFisher). This was performed according to the manufacturers’ instructions.

### 2.2.6. Plasmid Purification

The GeneJET Plasmid Miniprep Kit was used to purify plasmid DNA; this was performed according to the manufacturers’ instructions.
2.2.7. Plasmid Sequencing

To sequence and confirm the identity of plasmid constructs, samples were sent to Source Bioscience (UK) or GATC (Eurofins UK), and sequencing was performed using T7 forward and reverse primers.

2.2.8. Plasmid Transformation into \textit{E. coli} BL21 (DE3)

To enable high yield protein expression, purified constructs were transformed into BL21 (DE3) competent \textit{E. coli} (NEB), according to the manufacturers’ instructions.

2.3. Bioinformatics Analyses

2.3.1. Multiple Sequence Alignment Generation and Secondary Structure Prediction

Protein amino acid sequences were aligned using Clustal Omega (default parameters) (Madeira et al., 2019). The PSIPRED v4.0 server was used for the prediction of protein secondary structure (Buchan and Jones, 2019; Jones, 1999). The ESPript 3.0 server was used for illustrating protein multiple sequence alignments together with their respective secondary structure predictions. The similarity colouring scheme was set to percentage equivalence (default parameters) (Robert and Gouet, 2014).

2.3.2. Predicting the Presence of a Coiled-Coil Domain

To predict the presence of a coiled-coil domain within the FapF (UK4) gene, the FapF amino acid sequence was input into the MarCoil web server (default parameters) (Delorenzi and Speed, 2002; Zimmermann et al., 2018).

2.3.3. Predicting the Oligomeric State of a Coiled-Coil

The FapF\textsubscript{27-64} amino acid sequence was input into the LOGICOIL web server (whole protein sequence input format selected) to predict its oligomeric state (Vincent et al., 2013).
2.3.4. Signal Peptide Prediction

To predict the presence and location of signal peptides within the Fap genes, the SignalP 5.0 server was utilised (selecting the gram-negative organism group) (Armenteros et al., 2019).

2.3.5. Polyproline Type II Secondary Structure Prediction

To predict polyproline type II (PPII) secondary structure propensity in proteins, their corresponding amino acid sequences were input into the PPIIPred web server for analysis (O’Brien et al., 2020).

2.3.6. Structure Homology Modelling

The I-TASSER online server was utilised to predict the three-dimensional (3D) structure of the FapD<sub>39-246</sub> amino acid sequence (Yang and Zhang, 2015). The 6MPZ (PDB ID) PDB structure was used as a structural template (Rouse et al., 2017).

2.3.7. ConSurf

The ConSurf webserver was used to identify the evolutionarily conserved residues of FapD (using default parameters) (Ashkenazy et al., 2016). The conservation scores for each residue were plotted and colour coded onto the FapD homology model that was generated by I-TASSER (described previously).

2.3.8. CH3Shift

CH3Shift was used to back-calculate the <sup>1</sup>H and <sup>13</sup>C chemical shifts of the FapFβ methyl groups from the FapFβ crystal structure (PDB ID: 5O65) (Sahakyan et al., 2011). Dr Sarah Rouse kindly provided the 5O65 (PDB ID) structure with hydrogens added.

2.3.9. Molecular Analysis and Visualisation

UCSF Chimera (v1.11.2) and PyMOL v1.7.4 was used for protein analytical and visualisation purposes. Unless otherwise stated, UCSF Chimera was used for protein visualisation and analysis (Pettersen et al., 2004; PyMOL).
2.4. Protein Expression and Purification

2.4.1. Recombinant Protein Expression of Unlabelled (no isotopic labelling) FapF, and Unlabelled FapD

All FapF and FapD constructs were transformed into BL21 (DE3) Competent *E. coli* (NEB), plated onto LB-agar containing 50 mg/L kanamycin, and incubated at 37 °C. Of note, all of the media described in this chapter contain 50 mg/L of kanamycin. After overnight incubation at 37 °C, a single colony was picked to inoculate 10 ml of LB media (pre-culture). Cells were grown at 37 °C with shaking at 200 rpm. Once an OD600 of 0.4-0.6 absorbance units (AU) was achieved, 0.1 ml was extracted and used to inoculate 500 ml LB in a 2.5 L Thomson’s Ultra Yield Flasks (Thompson Instrument Company) (TUY; overnight culture). Cells were grown overnight at 37 °C at 200 rpm until an OD600 of 0.8 AU was achieved. Next, 1 L of LB in a 2.5 L TUY flask was inoculated with the quantity of overnight culture required to result in a starting OD600 of 0.10 AU (day culture). Cells were grown at 37 °C, 200 rpm, until an OD600 of 0.6 or 0.8 AU was achieved. For protein overexpression at 37 °C, cells at an OD600 of 0.8 AU were induced by addition of 1 mM IPTG (final concentration). For overexpression at temperatures lower than 37 °C, cells at an OD600 of 0.6 AU were grown at their required expression temperatures (200 rpm) for an additional 20 minutes, and protein overexpression was induced by the addition of 1 mM IPTG. After induction, cells were harvested by centrifugation at 4,000 RCF for 20 minutes at 4 °C. The supernatant was discarded, and the cell pellets that were not immediately required for purification were stored at -80 °C.

2.4.2. Recombinant Protein Expression of Unlabelled Membrane Scaffolding Protein 1D1 (MSP1D1)

To overexpress unlabelled MSP1D1, the protocol detailed for the unlabelled protein expression was modified as follows: The pre-culture and overnight culture procedures remain identical, however, the LB media used for the day culture was substituted for 500 ml of Terrific Broth (TB). The day cultures were allowed to grow to OD600 2.5-3.0 AU before the addition of 1 mM IPTG. Overexpression was carried out for 3 hours at 37 °C before cells were harvested by centrifugation, and subsequently stored at -80 °C, as described previously.
2.4.3. Recombinant Expression of $^{15}$N-labelled Protein

In order to overexpress $^{15}$N-labelled protein, the expression protocol detailed for unlabelled protein expression was modified as follows: 0.5 ml of pre-culture was inoculated into a 500 ml flask containing 100 ml of M9 minimal media (using $^{14}$NH$_4$Cl). This culture is grown at 37 °C, 200 rpm, overnight. Upon reaching an OD$_{600}$ of 0.8 AU, the overnight culture was centrifuged at 2,000 RCF for 20 minutes at 20 °C. The supernatant was discarded and the cell pellet was resuspended in 100 ml fresh M9; this was used to inoculate 900 ml of M9 (in a 2.5 L TUY flask) to result in a starting OD$_{600}$ of 0.10 AU, and the day culture was made up to 1 L using fresh M9. Of note, $^{14}$NH$_4$Cl was substituted for $^{15}$NH$_4$Cl in the day culture. Protein overexpression and cell harvesting was carried out as previously described.

2.4.4. Recombinant Expression of [85 %-$^2$H, U-$^{15}$N]-FapFβ

In order to overexpress [85 %-$^2$H, U-$^{15}$N]-protein, the expression protocol detailed for $^{15}$N-labelled protein was modified as follows: the overnight and day cultures were formulated in 100 % D$_2$O (Sigma-Aldrich) rather than H$_2$O. Once the OD$_{600}$ of the day culture reached ~0.7 AU, the culture was incubated for an additional 20 minutes at 30 °C, 200 rpm. Protein overexpression was subsequently induced by the addition of 1 mM IPTG, and the cell culture was incubated at 30 °C, 200rpm, for ~12 hours. After incubation, the cells were harvested by centrifugation, and the cell pellets were stored at -80 °C, as described previously.

2.4.5. Recombinant Expression of {U-[^2$^2$H, $^{15}$N], (Ala-β, Ile-δ1)-[^13$^3$CH$_3$]}-FapFβ

The expression of {U-[^2$^2$H,$^{15}$N], (Ala-β, Ile-δ1)-[^13$^3$CH$_3$]}-FapFβ was performed according to Tugarinov et al. (2006), with modifications. Of note, M9 was formulated in either H$_2$O or D$_2$O, and that is indicated in brackets. Transformation and preculture steps were performed as previously described. A 100 ml flask containing 20 ml of M9 (H$_2$O) was inoculated to achieve a starting OD$_{600}$ of 0.1 AU. Once an OD$_{600}$ of 0.6 AU was reached, the culture was centrifuged at 4,000 RCF 12 °C for 15 minutes and the supernatant was discarded. The cell pellet was resuspended in 100 ml M9 (D$_2$O) prepared with d$_{7}$-D-glucose, instead of $^{12}$C-glucose. The culture was incubated at 37 °C, 200 rpm, until an OD$_{600}$ of 0.5 AU was achieved. The culture was
subsequently diluted with 200 ml M9 (D2O; d7-D-glucose (Cambridge Isotope Laboratories); and 1g/L 14NH4Cl) and the precursors for Ala-β and Iso-δ1 methyl-labelling were subsequently added to the M9 mixture. Of note, the precursors were obtained from Cambridge Isotope Laboratories. The quantity of precursors added (per 1 L of cell culture) are as follows: 2.5 g succinic acid (D4, 98 %); 0.12 g alpha-ketoisovaleric acid, sodium salt (U-D7, 98 %); 0.06 g alpha-ketobutyric acid, sodium salt (methyl-13C, 99 %: 3,3-D2, 98 %); 0.8 g L-alanine (3-13C, 99 %; 2-D, 96 %); and 2 g d7-D-glucose (1,2,3,4,5,6,6-D7, 97-98 %). Of note, the E. coli metabolic pathways that are involved this methyl labelling strategy are summarised by Schütz and Sprangers (2020). Once an OD600 of 0.5 AU was achieved (~1 hour of growth), the culture was incubated for an additional 20 minutes at 30 °C, 200 rpm. Protein overexpression was subsequently induced by the addition of 1 mM IPTG, and the cell culture was incubated at 30 °C, 200 rpm, for ~8 hours. After incubation, the cells were harvested by centrifugation, and the cell pellets were stored at -80 °C, as described previously.

2.4.6. Small Scale Expression Trials

For small scale expression trials, all procedures up to the inoculation of an overnight culture are identical to those described for unlabelled protein expression. Once a preculture OD600 of ~0.6 was achieved, a 2.5 L TUY flask containing 1 L of LB was inoculated to result in a starting OD600 of <0.10 AU. Cells were cultured at 37 °C, 200 rpm, until an OD600 of 0.6 or 0.8 AU was achieved. Next, 100 ml of culture was aliquoted and transferred to a number of empty 250 ml flasks. Each of the 250 ml flasks were subsequently placed inside incubators set at different temperatures, and protein overexpression was induced as described within the unlabelled protein expression protocol. The expression conditions that were examined include 37 °C expression for 4 hours, and 30 °C, 25 °C, or 18 °C expression for 14 hours. After expression, a total of 50 ml from each 100 ml culture was harvested by centrifugation at 4,000 RCF for 20 minutes at 4 °C. The cell pellets were resuspended in 2 ml of their respective lysis buffer, that are later detailed, and cooled on ice for 10 minutes. After cooling, cells were lysed by sonication. Cell lysate was centrifuged at 13,000 RCF for 15 minutes. The soluble and insoluble fractions were separated and resuspended in SDS-PAGE loading buffer for analysis by SDS-PAGE.
2.4.7. Purification of FapF<sub>27-64</sub> (including R57A; R57E; Y44A) and FapF<sub>27-81</sub>

The purification procedures for FapF<sub>27-64</sub> and FapF<sub>27-81</sub> are identical. Cell pellets corresponding to FapF<sub>27-64</sub> or FapF<sub>27-81</sub> were resuspended in lysis buffer (LyB; 20 mM PO₄, 150 mM NaCl, 10 mM imidazole, pH 7.4) supplemented with 0.5 mM phenylmethylsulfonyl fluoride (PMSF); of note, 4 ml of lysis buffer was used per 1 g of cell pellet. The cell resuspension was placed on ice for 15 minutes. After cooling, cells were lysed by sonication. Cell lysate was centrifuged at 17,000 RCF, 4 °C, for 1 hour and the soluble fraction was collected. The supernatant was applied to a gravity flow column containing Ni-NTA Superflow resin (Qiagen) for nickel affinity-immobilized metal affinity chromatography (Ni-IMAC), and that was equilibrated in LyB. Of note, 1 ml of resin was used per 40 ml of supernatant. Once all of the supernatant was applied to the Ni-IMAC column, the resin was washed with LyB that contained incrementally increasing concentrations of imidazole, starting from 20 mM, to 30 mM, and finally 40 mM, in order to remove non-specific proteins from the column. Ten column volumes (CV) of each wash buffer were applied to the column. The protein of interest was eluted from the column by application of 5 CV of elution buffer (LyB containing 250 mM imidazole), and dialysed into LyB at 4 °C in order to reduce the sample imidazole concentration. Of note, Spectrum Spectra/Por 3 RC Dialysis Membrane Tubing 3500 Dalton MWCO (Fisher Scientific) is used for all dialysis procedures described. Next, the FapF<sub>27-64</sub> or FapF<sub>27-81</sub> were mixed with TEV protease 10:1 (w/w) (to remove the hexahistidine tag (his6tag)) and incubated overnight at 4 °C. Subsequently, the sample was applied to a Ni-IMAC gravity column equilibrated with LyB (a process referred to as “Reverse Ni-IMAC”). Cleaved samples eluted in the flow through and this was collected. SDS-PAGE analysis was used to confirm the success of TEV-cleavage. To increase sample purity, the sample of interest was concentrated to 2 ml using a 3 kDa molecular weight cut-off (MWCO) concentrator (FisherScientific). The concentrated sample was applied to a HiLoad 16/600 Superdex 75 pg column (GE) or a ProSEC 16/60 3-70 HR SEC Column (Generon), that was equilibrated in buffer containing 20 mM PO₄, 150 mM NaCl, pH 6.8, for size exclusion chromatography (SEC). The SEC UV trace indicated the fractions that contain protein, and SDS-PAGE analysis confirmed the presence of the
protein of interest. The required fractions were collected and stored at 4 °C until further analysis was necessary. Of note, the FapF<sub>27-81</sub> samples that were used in experiments involving FapD were not subjected to TEV protease digestion or the reverse Ni-IMAC step. Instead, the first Ni-IMAC eluent was concentrated to 5 ml using a 3 kDa MWCO concentrator and subjected to SEC. The required SEC fractions were dialysed against buffer containing 50 mM HEPES, 200 mM NaCl, pH 7.5, and the required reducing agent (specified within the text). If required, samples were concentrated using a 3 kDa MWCO concentrator (FisherScientific).

2.4.8. Purification of FapD<sub>39-246</sub> (FapDWT) and FapD<sub>50-246</sub> (FapDT)

The FapD lysis and purification protocol shares similarities to that described for FapF<sub>27-64</sub> and FapF<sub>27-81</sub>, and their differences are outlined below. Cell pellets corresponding FapDWT or FapDT were resuspended in lysis buffer (DLyB; 50 mM HEPES, 200 mM NaCl, 5 % (v/v) glycerol, 10 mM imidazole, pH 7.5) supplemented with 0.5 mM PMSF. The cells were lysed, centrifugated, and applied to a Ni-IMAC column as previously described, however, the Ni-IMAC column was equilibrated with DLyB instead of LyB. The wash steps were consistent with that described previously, with the exception that the wash buffers for FapD were formulated with DLyB instead of LyB. FapD was eluted from the column by the addition of 5 CV of DlyB that contained 250 mM imidazole. The elution fraction was supplemented with 2 mM DTT. To increase the purity of FapD, 5 ml of the Ni-IMAC elution fraction was applied to a SEC column equilibrated in buffer containing 50 mM HEPES (pH 7.5), 200 mM NaCl, 5 % (v/v) glycerol, and 2 mM DTT. The SEC fractions containing FapD, as determined by SDS-PAGE analysis, were stored at 4 °C until further analysis. If required, samples were concentrated using a 10 kDa MWCO concentrator (FisherScientific).

2.4.9. Purification of MSP1D1

The purification of membrane scaffolding protein 1D1 (MSP1D1) was performed according to Ritchie et al. (2009), with additional purification steps that are as follows: After the purification of MSP1D1 by Ni-IMAC, the sample was dialysed at 4 °C overnight against buffer 1a (20 mM Tris-HCl, 0.1 M NaCl, 0.5 mM EDTA, pH 7.4). The sample was subsequently mixed with TEV protease 10:1 (w/w) and dialysed
overnight at 4 °C against buffer 1b (20 mM Tris-HCl, 0.1 M NaCl, pH 7.4). Subsequently, the sample was applied to a Ni-IMAC gravity column equilibrated in buffer 1b for reverse Ni-IMAC. Cleaved samples eluted in the flow through and this fraction was collected. To increase sample purity, the sample was concentrated to 5 ml using a 5 kDa MWCO concentrator (FisherScientific) and applied to a HiLoad 16/600 Superdex 75 pg column (GE) equilibrated in buffer 1b for SEC. The SEC fractions containing MSP1D1 were collected, flash frozen in liquid nitrogen, and stored at -80 °C until required.

2.4.10. Refolding and Purification of FapF\textsubscript{107-430} (FapFβ)

The cell pellets corresponding to FapFβ were resuspended in lysis buffer (20 mM Tris, 300 mM NaCl, pH 8.0) supplemented with 0.5 mM PMSF. The cells were lysed and centrifuged, as previously described. After centrifugation, the supernatant was discarded, and the insoluble pellet (containing FapFβ inclusion bodies) was collected. The insoluble pellet was dissolved in denaturing buffer (DB; 20 mM Tris-HCl, 300 mM NaCl, 8 M urea, pH 8.0) by stirring at 20 °C for 8 hours. Of note, the volume of DB utilised per gram of pellet is specified within the text. Next, the sample was centrifuged at 17,000 RCF, 16 °C, for 2 hours and the soluble fraction was collected. This step removes any insoluble matter that did not dissolve in DB. The soluble fraction was applied to a Ni-IMAC column equilibrated in DB, in order to purify and concentrate FapFβ. After applying the soluble fraction to the column, the column was washed with 10 CV of DB supplemented with 20 mM imidazole. FapFβ was eluted from the column by application of 5 CV of elution buffer (DB supplemented with 250 mM imidazole; DBe), and a fraction was collected for SDS-PAGE analysis. The concentration of FapFβ in the eluent was approximated by Nanodrop measurement. If required, the concentration of FapFβ was diluted to the required concentration using DBe. The FapFβ sample was refolded at 20 °C by dropwise dilution, with gentle stirring, into a 20-fold greater volume of non-denaturing buffer (NDB; 20 mM Tris-HCl, 300 mM NaCl, pH 8.0) containing 5 % (v/v) LDAO (N,N-Dimethyldodecylamine N-oxide). The rate of dilution was maintained at approximately 0.25 ml/min by use of a Peristaltic Pump P-1 (GE). After dilution, the sample was stirred at 20 °C and incubated for 12 hours (or the duration specified within the text). Next, the sample was dialysed at 4 °C against NDB
containing 0.5 % (v/v) LDAO to remove the residual imidazole and urea present in the sample. After dialysis, the sample was concentrated and purified by application to a 1 ml Histrap HP column (GE) for Ni-IMAC. The column was washed with 10 CV of NDB supplemented with 0.1 % (v/v) LDAO and 20 mM imidazole, and FapFβ was eluted from the column by using 5 CV of NDB supplemented with 0.1 % (v/v) LDAO and 250 mM imidazole. The eluent was centrifuged at 10,000 RCF for 5 minutes. The soluble fraction was applied to a HiLoad 16/600 Superdex 200 pg column equilibrated in NDB supplemented with 0.1 % (v/v) LDAO, in order to remove soluble aggregates. SDS-PAGE analysis was use to confirm the presence of purified FapFβ, and the required SEC fractions were stored at 4 °C until required.

2.4.11. Detergent Exchange

The exchange of FapFβ from NDB containing 0.1 % (v/v) LDAO to NDB containing 0.5 % (v/v) C8E4 (tetraethylene glycol monooctyl ether) was performed according to Rouse et al. (2016). Briefly, Purified FapFβ was applied to a 1 ml Histrap HP equilibrated in NDB supplemented with 0.1 % (v/v) LDAO. The column was first washed with 10 CV of NDB supplemented with 0.1 % (v/v) LDAO and 0.5 % (v/v) C8E4. Next, the column was washed with 10 CV of NDB supplemented with 0.5 % (v/v) C8E4, and FapFβ was eluted from the column by applying 5 CV of NDB supplemented with 0.5 % (v/v) C8E4 and 250 mM imidazole. The eluent was centrifuged at 10,000 RCF for 5 minutes, and the soluble fraction was loaded onto a HiLoad 16/600 Superdex 200 pg column equilibrated in NDB supplemented with 0.5 % (v/v) C8E4; or the sample was concentrated to 1 ml (using a 50 kDa MWCO concentrator; Amicon Ultra-15 Centrifugal Filter Unit (Millipore)) for loading onto a Superdex 200 10/300 GL (GE). SDS-PAGE analysis was use to confirm the presence of purified FapFβ, and the required SEC fractions were stored at 4 °C until required.

2.4.12. Gel Filtration

The ÄKTA pure, ÄKTA prime, or ÄKTA start systems (GE) were paired with SEC columns for protein purification. For preparative columns or analytical columns, flow rates between 0.8-1.0 ml/min or 0.25 ml/min were utilised, respectively; fraction sizes were 2 ml or 0.5 ml, respectively. Of note, the fraction volume indicated on the SDS-PAGE gels throughout this thesis refer to the initial fraction volume. For
example, “74 ml” corresponds to fraction volumes 74.00-75.99 ml, for a preparative column.

2.5. Techniques For Protein Detection and Quantification

2.5.1. SDS-PAGE

Samples for SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) analysis were mixed with 2x Laemmli buffer (126 mM Tris, 20 % (v/v) glycerol, 4 % (w/v) SDS, 0.02 % bromophenol blue, pH 6.8) and supplemented with 1 mM DTT (SDS-PAGE loading buffer). The required samples were incubated at 100 °C for 5 minutes. After incubation, samples were loaded into precast SERVAGel TG PRiME gels (Generon) containing 12 % or 14 % polyacrylamide, or 4-20 % polyacrylamide gradient. The running buffer was comprised of 250 mM Tris, 1.92 M glycine, and 1 % (w/v) SDS. A voltage was applied across required precast gel in accordance with the manufacturers’ instructions. In order to visualise proteins, Quick Coomassie Stain was used for staining gels, and gel images were subsequently collected. In order to estimate sample protein molecular weights, a PageRuler Unstained Broad Range Protein Ladder (Thermo Scientific), or a PageRuler Prestained Protein Ladder, 10 to 180 kDa (Thermo Scientific), or a Blue Prestained Protein Standard, Broad Range (11-190 kDa; NEB) was used for comparison. Of note, lanes labelled “M” at the top of each SDS-PAGE gel figure corresponds to the ladder or “marker” lane.

2.5.2. Western Blotting

Western blots were prepared using the protocol described for SDS-PAGE analysis, without staining. The Trans-Blot Turbo (TM) System (BioRad) was used to transfer the proteins from unstained gels onto a 0.2 μm nitrocellulose membrane from the Trans-Blot Turbo Mini Nitrocellulose Transfer Packs. The nitrocellulose membrane was blocked using TBS buffer (20 mM Tris, 150 mM NaCl, pH 7.5) containing 5 % (w/v) skimmed milk powder and left shaking for an hour (unless otherwise stated, all the steps described are performed at 20 °C). After blocking, the buffer was discarded and the membrane was washed and lightly shaken for 5 minutes with TBST buffer (TBS buffer plus 0.05 % (v/v) Tween20). Next, the membrane was incubated with TBST buffer, 0.1 % (w/v) skimmed milk powder, and a 3000:1 ratio of the volume of
TBST to Monoclonal Anti-polyHistidine-Preoxidase clone HIS-1 (Sigma Aldrich; prepared according to the manufacturers’ instructions), that was left shaking for 1 hour at 20 °C. After 1 hour, the membrane was washed and lightly shaken with TBST buffer for 5 minutes and repeated four times, followed by rinsing and shaking with TBS buffer for five minutes, a procedure that is repeated twice. The antibodies were detected using the SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Scientific), performed in accordance with the manufacturers instructions, and subsequently imaged by the ChemiDoc XRS+ System (Bio-rad). In order to estimate sample protein molecular weights, a BenchMark His-tagged Protein Standard Ladder (Thermo Scientific) was used for comparison.

2.5.3. NanoDrop Measurement

With the exception of FapF27-64 Y44A, all other protein concentrations were measured by their absorbance at 280 nm using a NanoDrop ND-1000 UV-Vis Spectrophotometer (A280 nm). The ProtParam tool on the ExPASy Server was used to determine protein extinction coefficients (Gasteiger et al., 2005). Corresponding values were input into the Beer-Lambert Law for protein concentration calculation (Pace et al., 1995; Swinehart, 1962).

2.5.4. Bradford Assay

To estimate the concentration of FapF27-64 Y44A, the Bradford assay was employed. To begin, 1 ml of Bradford Coomassie reagent (Bio-Rad) was mixed with 50 µl of protein sample, or 50 µl of buffer without protein (blank), within a cuvette. This mixture was incubated for 5 minutes at 22 °C (according the manufacturer’s instructions). Next, the absorbance at 595 nm was measured, and the absorbance of the blank was subtracted from that of the protein sample. This process was repeated for various concentrations of bovine serum albumin (BSA; Bio-Rad) in order to generate a standard curve, from which the concentration of FapF27-64 Y44A could be estimated (Simonian and Smith, 2006).
2.6. FapD Stability Assay

To determine the stability of FapD, FapD was purified according to the methods previously described. Aliquots of purified FapD were supplemented with an additional 3 mM DTT (5 mM DTT total), and samples were subsequently concentrated to 25 µM using a 5 kDa MWCO concentrator. For the assays conducted with 5 µM of FapD, the concentrated 25 mM sample was diluted using buffer containing 50 mM HEPES, 200 mM NaCl, 5 % (v/v) glycerol, and 5 mM DTT. Samples were incubated at 22 °C for up to 6 days. Fractions were collected at incremented time intervals (specified within the text), centrifuged at 13,000 RCF, and an aliquot from the soluble fraction was collected for SDS-PAGE analysis.

2.7. Nanodisc Assembly

The reconstitution of FapFβ into MSP1D1 nanodiscs was performed according to the instructions published in Ritchie et al. (2009). Briefly, DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine; Avanti Polar Lipids) in a chloroform stock solution was desiccated. Next, DMPC was solubilised in a volume of buffer (20 mM HEPES, 150 mM NaCl, 100 mM sodium cholate, pH 7.5) that results in a 50 mM DMPC stock solution. In order to assemble the MSP1D1 nanodisc, purified FapFβ that was exchanged into 0.5 % (v/v) C8E4 was mixed with DMPC and purified MSP1D1 at a ratio of 1:4:320 (60 µM:240 µM:19.2 mM). This mixture was incubated for 1 hour at 20 °C. Empty nanodiscs comprised of MSP1D1 and DMPC were prepared as described above, in the absence of FapFβ. To remove detergent from the sample, 100 % w/v of Bio-Beads SM-2 (Sigma Aldrich) was added to the sample. The mixture was incubated at 26 °C and shaken at 350 rpm for for 14 hours using a Eppendorf ThermoMixer C (Eppendorf AG). The sample was subsequently centrifugated at 3,000 RCF for 10 minutes to pellet the Bio-Beads. The supernatant was collected and subjected to SEC, using a Superdex 200 10/300 GL (GE) equilibrated in 20 mM HEPES, 150 mM NaCl, pH 7.5. The SEC fractions containing FapFβ and/or MSP1D1 were identified by SDS-PAGE analysis.
2.8. Biophysical Techniques

2.8.1. Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS)

To perform size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), a 1260 infinity (Agilent Technologies) was coupled to a MiniDawn TREOS (Wyatt Technology), an OptiLab T-rEX (Wyatt Technology), and a Superdex 75 10/300 GL (GE; S75A) or a Superdex 200 10/300 GL (GE; S200A). For molecular weight determination, FapF\textsubscript{27-64}, FapF\textsubscript{27-64}R57A, FapF\textsubscript{27-64}R57E, and FapF\textsubscript{27-81}, were subjected to the S75A; FapF\beta was subjected to the S200A. The column was equilibrated in buffer corresponding to the SEC buffer for each respective sample. Sample volumes for injection were 0.1 or 0.5 ml (specified within the text), and the flow rate was set to 0.25 ml/min. The refractive index increment (dn/dc) was set to 0.185 ml/g, and ASTRA software (Wyatt Technology) was used to determine the sample weight-average molar mass (Zhao et al., 2011). The dn/dc for samples in buffer containing 0.1 % (v/v) LDAO was set to 0.148 ml/g (Strop and Brunger, 2005).

2.8.2. Solution-State NMR Spectroscopy

All solution state nuclear magnetic resonance (NMR) data was acquired using the Avance III HD 800 MHz or Avance III 600 MHz spectrometers equipped with triple-resonance cryoprobes. A deuterium signal was used for NMR locking and shimming; samples that were not prepared in 100 % D\textsubscript{2}O were supplemented with 5 % D\textsubscript{2}O (99.96 %; VWR). Samples for \textsuperscript{1}H-\textsuperscript{1}H Nuclear Overhauser Effect Spectroscopy (NOESY) and \textsuperscript{1}H-\textsuperscript{13}C heteronuclear multiple quantum coherence (HMQC; J-coupling set to 125 Hz) experiments were exchanged into buffer formulated in 99.96 % D\textsubscript{2}O by using an appropriate MWCO concentrator as previously described. The concentration of each NMR sample, and the temperature and number of scans for each experiment, are specified within the text. The resulting NMR data was processed and analysed using TopSpin 3.2 (Bruker).

2.8.3. Circular Dichroism (CD) Spectroscopy

All far-UV circular dichroism (CD) Spectroscopy data was obtained on a Chirascan CD Spectrometer (Applied Photophysics). Samples were loaded (200 µl)
into a Hellma absorption cuvette (spectral range 200-2500 nm, pathlength 1 mm, chamber volume 350 µl). CD measurements were carried out at wavelengths of 200 nm to 260 nm, with a step size of 0.5 nm or 1.0 nm, a bandwidth of 1 nm, a time per point of 2s, and the collection of three or five repeat scans. For each protein sample, CD spectra were collected for their respective buffers by using an identical experimental set up. The averaged buffer measurements were subtracted from their corresponding and averaged protein sample measurements. The resulting spectrum was smoothed by a Savitsky-Golay filter (window 3 or window 5) (Savitzky and Golay, 1964). The data from the CD spectra was input into the BeStSel webserver in order to predict the protein secondary structure compositions (Micsonai et al., 2015, 2018). The spectral deviation (normalised root mean square deviation) between the experimentally derived data and the data calculated by BeStSel was always below 0.01. For CD melting curves, the cell temperature was controlled by a Peltier regulator and increased by 1 °C per minute (tolerance set to 0.2 °C), ranging from 20 °C to 90 °C. Measurements at 222 nm were collected across the temperature range to produce a melting curve, from which the \( T_m \) was derived by calculation of the maximum of the first derivative (Greenfield, 2006). Of note, the \( \Delta_{22}^{22} \) ratio used to inform the presence of a coiled-coil structure was based on the following sources: Crooks et al. (2011); Kwok and Hodges (2004); Lau et al. (1984); Zhou et al. (1992).

### 2.8.4. Isothermal Titration Calorimetry (ITC)

The isothermal titration calorimetry (ITC) data was collected using a MicroCal iTC200 (Malvern Pananalytical). An ITC experiment was performed to investigate the thermodynamic binding parameters of the FapDT and FapF\(_{27-81}\) interaction. Purified FapDT and purified FapF\(_{27-81}\) were dialysed three-times at 4 °C with gentle stirring against 2 L of buffer containing 50 mM HEPES, 200 mM NaCl, 5 % (v/v) glycerol, 1 mM TCEP (tris(2-carboxyethyl)phosphine hydrochloride), pH 7.5. The FapF\(_{27-81}\) was concentrated to 500 µM, and 40 µl was aspirated into the ITC syringe. FapDT was concentrated to 25 µM and 300 µl was loaded into the ITC cell chamber. ITC experiments were performed at 298 K, with a stirring speed of 750 rpm, and the reference power was set to 7. A total of 40 injections were performed, the first of which was a test injection (0.4 µl injected for a duration of 0.8 s), and after 540 s, 39 injections followed, each comprised of a 1 µl injection (duration of 2.0 s, 5 s filter
period) with a 180 s interval between each of the 39 remaining injections. A control experiment was performed, for which 500 µM FapF<sub>27-81</sub> was titrated into the cell containing only the ITC buffer, using identical experimental parameters to those described above. The data was processed and analysed using Origin 7.0 software (OriginLab). The heat of dilution was subtracted (point-by-point) from the heat of the binding for the interaction experiment, and the results were integrated and fitted using a nonlinear least-squares algorithm with a one-site binding model, and that includes the derivation of thermodynamic parameters. Anomalous data points were removed to enhance the fitting accuracy.

2.8.5. Dynamic Light Scattering (DLS)

The dynamic light scattering (DLS) data was collected using a Zetasizer Nano-ZS 3600 instrument (Malvern Instruments Ltd., Malvern, UK), equipped with a He-Ne laser (633 nm light source; 173 degree backscatter measurement). A DLS experiment was performed to determine the hydrodynamic size of FapFβ-MSP1D1-nanodiscs and empty MSP1D1-nanodiscs. For analysis, each sample was diluted to 1.0 mg/ml and loaded into a 45 µl Ultra-Micro Cell 105.251-QS LP 3x3 mm (Hellma Analytics). The sample was allowed to equilibrate for 5 minutes at 298 K prior to the experiment that was carried out at 298 K. Three repeats measurements were performed for each sample. The measurement parameters were automatically optimised: the measurement position is 4.20 mm, the attenuator was set to 8, and the duration of each experiment was 80 s. To analyse the light scattering data, Zetasizer Nano software v3.30 (Malvern Panalytical) was utilised, for which a “General purpose” (non-negative least squares) analysis model was used for calculating the sample size distributions.

2.9. Background of the Biophysical Techniques Used in this Study

2.9.1. Solution State Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy represents an invaluable method for probing the structure, interactions, and dynamics of proteins, and together, this information can yield insight into the function(s) of a protein (Markwick et al., 2008). For detailed descriptions and discussions on NMR theory, the reader is guided to reviews such as Breukels et al. (2011) and Kwan et al. (2011). In this thesis, a brief description of the phenomena of
NMR, based predominantly on the reviews of Breukels et al. (2011), Kwan et al. (2011), and Putheenveetil and Vinogradova (2019), is provided.

Fundamentally, and over-simplistically, atomic nuclei are able to transition between different energy states, and these transitions give rise to NMR spectra. The isotopes of various nuclei possess a magnetic dipole, including $^1$H, $^{13}$C, and $^{15}$N, and that when placed within a magnetic field, adopt orientations of different energy (Kwan et al., 2011). The nuclei described above each possess a nuclear spin of $\frac{1}{2}$; nuclear spin is a description for the total angular momentum of a nucleus. For nuclei with spin $\frac{1}{2}$, two spin states exist: spin $-\frac{1}{2}$ ($\alpha$), for which the nuclear spin is aligned with the magnetic field (low energy state); and spin $\frac{1}{2}$ ($\beta$), for which the nuclear spin is aligned oppositely to the magnetic field (high energy state) (Breukels et al., 2011). In order to observe NMR signals, specific frequencies of electromagnetic (EM) radiation are applied to the nuclei (in the magnetic field) to induce transitions between these two energy levels, in accordance with Planck’s law (Breukels et al., 2011; Kwan et al., 2011). The precise frequency of EM radiation that is required to excite the magnetic nuclei is referred to as the Larmor (resonance) frequency. The resonance frequency of a magnetic nucleus varies depending on the strength of the applied magnetic field, the type of NMR nucleus, as well as the extent of magnetic nuclear shielding (Breukels et al., 2011). The intensity of the nuclear resonance is plotted against its respective resonance frequency to produce a 1D NMR spectrum (Breukels et al., 2011; Kwan et al., 2011). To correct for the difference in resonance frequency between the varying magnetic field strengths of NMR spectrometers, the resonance frequency is reported in terms of chemical shift (using parts per million (ppm) units) with reference to tetramethylsilane (0 ppm) (Breukels et al., 2011). Unlike other spectroscopic methods, the lifetime of the excited state in NMR spectroscopy extends into the range of milliseconds to seconds (Kwan et al., 2011). Various NMR experiments have come to exploit this phenomenon by transferring the excitation state between nuclei in a through-bond or through-space manner (Breukels et al., 2011). Complex heteronuclear NMR experiments are regularly employed to measure the resonance frequency of multiple, correlated nuclei. The chemical shifts of correlated nuclei can be visualised on multidimensional spectra (Breukels et al., 2011; Kwan et al., 2011). The correlation of multiple nuclear frequencies enables, in theory, the unambiguous assignment of each spectral peak to its corresponding atom of a molecule or protein (Breukels et al., 2011). Ultimately, the through-bond and through-
space information provided by NMR experiments allows us to determine the local environment (position) of an atom or amino acid within a protein, enabling the elucidation of a protein's three-dimensional structure, and therefore, potentially its function (Kwan et al., 2011; Puthenveetil and Vinogradova, 2019).

Within structural biology, the 1D $^1$H-NMR and 2D $^1$H-$^{15}$N HSQC (heteronuclear single quantum coherence) spectra are ubiquitously employed due to their relatively high sensitivity and versatility (Kwan et al., 2011). The peaks within the former spectrum correspond to proton signals, and the latter spectrum presents signals that correspond to the $^1$H-$^{15}$N amide group. Regarding the $^1$H-$^{15}$N HSQC, each spectral peak is comprised of both an intensity, as well as the chemical shifts of the correlated $^1$H and $^{15}$N nuclei (Breukels et al., 2011; Kwan et al., 2011). In theory, the peaks observed on this spectrum include: one per protein backbone amide (with the exception of proline); one per tryptophan indole; and two per Asn and Gln side chain (Kwan et al., 2011). The $^1$H-$^{15}$N HSQC spectrum is often described as the “protein fingerprint” due to its ability to crudely identify whether a protein is comprised of folded and disordered regions, with the former indicated by a range of peak dispersion in the $^1$H dimension, and the latter by narrow dispersion in this dimension (Gupta and Bhattacharjya, 2014; Kwan et al., 2011; Puthenveetil and Vinogradova, 2019). The range of peak dispersion observed for a folded protein, particularly the peaks downfield of 8.5 ppm, results from the various chemical environments found in $\alpha$-helices and $\beta$-sheet environments. This is in contrast to the narrow chemical shift range observed for unfolded proteins, due to their near-uniform exposure to solvent. Of note, $\beta$-sheet-rich proteins possess a more varied chemical environment compared to $\alpha$-helix-rich proteins, therefore, display relatively more peak dispersion (Kwan et al., 2011). Furthermore, the change in isotropic chemical shift due to motions on the microsecond to millisecond timescale (chemical exchange processes) can also be determined by NMR spectroscopy, as described by Kleckner and Foster (2011). For instance, slow-exchange processes (on the NMR time scale) occurs when the exchange rate ($k_{ex}$) of two conformations is significantly smaller than the chemical shift difference ($\Delta \omega$); this is expected to increase the number of spectral peaks (relative to the number of spectral peaks expected). When two conformations undergo intermediate exchange on the NMR time scale ($k_{ex} \sim \Delta \omega$), peak line widths are broadened; conformations undergoing fast exchange ($k_{ex} \gg \Delta \omega$) will display a single, intense peak (Kleckner and Foster, 2011; Kwan et al., 2011). The $^1$H-$^{15}$N NMR
spectrum is also able to crudely inform the molecular weight (MW) of a protein. In NMR spectroscopy, signal broadening increases as the molecular weight (MW) of a protein increases, due to factors that contribute to nuclear relaxation (Puthenveetil and Vinogradova, 2019). Broadly speaking, nuclear relaxation is the process by which the high-energy state of a nuclear spin returns to its equilibrium energy state, and that results in NMR signal deterioration (Pfuhl and Driscoll, 2000). NMR signal deterioration is particularly obvious for proteins above >25 kDa; also, the extent of NMR signal broadening can therefore inform on protein sample aggregation (Puthenveetil and Vinogradova, 2019).

The natural abundance of the $^{15}$N isotope is 0.4 % (and 1.1 % for $^{13}$C), which is insufficient for most protein NMR studies (Kwan et al., 2011; Pfuhl and Driscoll, 2000). In order to produce isotopically label proteins for NMR studies, recombinant proteins are typically expressed by culturing E. coli in minimal media containing nutrients enriched in the required isotope, such as $^{15}$NH₄Cl and/or $^{13}$C-glucose (Kwan et al., 2011). Unlike the 2D $^1$H-$^{15}$N HSQC spectrum, a 1D $^1$H-NMR spectrum can be obtained without the need for isotopic labelling, as protons account for over 99.9 % of hydrogen isotopes (Kwan et al., 2011; Pfuhl and Driscoll, 2000). The $^1$H-NMR spectrum is similar to that of the $^1$H-$^{15}$N HSQC, in terms of its ability to broadly inform on the folding and size of a protein, based on the peak dispersion and peak line widths, respectively (Kwan et al., 2011). In addition, methyl group protons that are in proximity to aromatic rings can be upfield shifted due to aromatic ring currents, thus, the occurrence of peaks upfield of 0 ppm also informs on the presence of a folded domain (Haigh and Mallion, 1979; Perkins and Wüthrich, 1979). However, the $^1$H-NMR spectrum amide region is often overcrowded, and additional peaks corresponding to the non-labile protons of various buffer components may complicate spectral analysis (Kwan et al., 2011). Therefore, the 2D $^1$H-$^{15}$N HSQC is the preferred experiment to determine whether a protein is amenable to NMR studies using the $^{13}$C, $^{15}$N-based heteronuclear NMR experiments (Breukels et al., 2011).

### 2.9.2. Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy is commonly employed technique within the field of structural biology due to its ability to rapidly evaluate the secondary structure of a protein. CD spectroscopy, which is a form of light absorption
spectroscopy, measures the differential absorption of left- and right-circularly polarised light by an aqueous sample containing the molecules of interest. Chiral molecules exhibit circular birefringence, which produce elliptically polarized light; this includes all amino acids except glycine (Adler et al., 1973; Greenfield, 2006b). Importantly, the phenomenon of CD is highly sensitive to the secondary structure composition of a protein; the different secondary structures of protein molecules exhibit unique CD absorptive properties, and therefore, possess distinctive CD absorption profiles (evident within the far-ultraviolet (far-UV) range). The general CD absorption profile of an α-helix gives rise to two bands of negative ellipticity at 222 nm and 208 nm, as well as a band of positive ellipticity at 193 nm. A β-sheet will generally give rise to a band of negative ellipticity at 218 nm, and a band of positive ellipticity at 195 nm. The CD absorption profile of a random coil is characteristically comprised of minimal ellipticity above 210 nm, and a band of negative ellipticity near 195 nm (Greenfield, 2006b). Various softwares have been designed to deconvolute CD spectra; this includes the highly accurate BeStSel web server (Greenfield, 2006b, Micsonai et al., 2015, 2018). As well as being able to estimate the secondary structure of a protein, CD spectroscopy is capable of monitoring the conformation of a protein under various conditions; this includes the effects of heat and denaturing agents, as well as the effects of a mutation or binding interaction (Greenfield, 2006b). The lack of residue-specific detail that is obtained from CD spectroscopy, in comparison to techniques such as NMR spectroscopy and X-ray crystallography, is compensated for by its ability to rapidly analyse samples within a few hours, and its low sample requirements (≤ 20 µg of protein) (Greenfield, 2006b; Kelly and Price, 2000). For additional details on the underlying theory and applications of CD spectroscopy, the reader is guided to reviews by Greenfield (2006b) and Kelly and Price (2000).

2.9.3. Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering

The reliable measurement of a protein’s molecular weight (MW) is essential for biomolecular research as it allows one to determine the presence of a protein of interest, as well as its oligomeric state and suitability for downstream experimentation. Briefly, analytical size exclusion chromatography (SEC) is a method that is utilised in order to separate macromolecules according to their relative sizes
and diffusion coefficients as they flow through a packed column (nominally speaking). Protein molecular weights (MWs) are often estimated by comparing the SEC retention volume of a sample (containing the protein of interest) with a calibration curve. In general, the calibration curve is comprised of a number of stable globular proteins of various MWs, and that are relatively inert with respect to their propensity for column binding. However, the above method is not considered an “absolute” technique for protein MW determination due to the various disparities between the proteins comprising the calibration curve and the protein of interest. This includes the differences between their conformations and surface properties; the resulting MW calculations can often be erroneous (Some et al., 2019).

To increase the reliability (and versatility) of SEC for MW determination, it is combined with differential refractive index (dRI) and multi-angle light scattering (MALS) detectors; the collective instrumentation is referred to as SEC-MALS. The dRI detector is a universal concentration detector; the presence of an analyte within a solution modifies the solution refractive index, which is monitored by the dRI detector, and that informs on the analyte concentration. In addition, the analyte scatters light (originating from a laser source) at multiple angles relative to the incident laser beam, the proportion of which is measured by the MALS detector(s). Together, the dRI and MALS measurements allow the MW of a macromolecule to be determined independently of its SEC retention volume, as described within Some et al. (2019). Of note, the use of SEC simply enables the various macromolecular species that are present within solution to be separated for individual analysis by the dRI and MALS detectors. The calibration of dRI and MALS instruments are performed independently of the SEC column, without the reliance on reference standards; thus, SEC-MALS is classified as an “absolute” technique for protein MW determination. In addition, SEC-MALS is capable of calculating the average MW of solute particles across an elution fraction, thus, it is able to inform on sample homogeneity or heterogeneity (Some et al., 2019; Ye, 2006). For additional details on the underlying theory and applications of SEC-MALS, the reader is guided to Some et al. (2019).
2.9.4. Isothermal Titration Calorimetry

The thermodynamic and binding parameters of a binding reaction are often determined by use of isothermal titration calorimetry (ITC). ITC is a label-free technique that directly measures the change in enthalpy ($\Delta H$) that results from a biomolecular interaction. For a detailed review of the theory and methodology underlying ITC, the reader is guided to Velazquez-Campoy et al. (2004). Briefly, the ITC instrument is comprised of a sample cell and a reference cell, both of which are identical, constructed of an inert and highly conductive material, and placed within an adiabatic jacket (Ciulli and Abell, 2005; Velazquez-Campoy et al., 2004). The sample cell is filled with the sample of interest (protein A), and the reference cell is filled with water or buffer; the second sample of interest (protein B) is contained within a syringe. Over the course of an ITC experiment, protein B is titrated into the sample cell (in a defined manner), and mixed. This results in heat changes due to protein A - protein B interaction(s) (as well as dilution due to titration, and stirring); sensitive circuits are responsible for detecting the temperature within the sample cell. Subsequently, the ITC instrument regulates the amount of power delivered to the sample cell in order to restore the equilibrium temperature between the two cells; the required power (relative to a reference power setting) is plotted as a function of time, generating peaks that correspond to each injection event. The magnitude of the peaks progressively decrease after each successive injection; this occurs due to the decreasing quantity of uncomplexed protein, which continues until all binding sites are saturated, and that may generate a titration curve. In order to calculate the enthalpy of binding, the peaks are integrated and plotted as a function of the molar ratio of protein B to protein A. The production of a sigmoidal titration curve yields a number of parameters that includes the binding enthalpy ($\Delta H$), which is proportional to the amplitude of the titration curve; the stoichiometry (N), which is given by the inflection point of the curve; and the association constant ($K_a$), which is yielded by the slope of the curve (Callies and Daranas, 2016; Ghai et al., 2012; Velazquez-Campoy et al., 2004). The change in Gibbs free energy ($\Delta G$) in an open system comprising a constant temperature and pressure is attained by determination of $K_a$ using the following equation: $\Delta G = -RT\ln(K_a)$ (where $R$ is the universal gas constant (8.31 J mol$^{-1}$ K$^{-1}$), and $T$ is the experimental temperature (K)). The equilibrium dissociation constant ($K_d$), that represents the affinity of a protein-protein interaction, is inversely
proportional to $K_a$, and is attained by taking the reciprocal of $K_a$ ($K_a = 1/K_d$). The change in the entropy of binding is attained with measurement of experimental temperature ($T$) and the $\Delta H$, which can be used to solve the following equation: $\Delta G = \Delta H - T\Delta S$ (Callies and Daranas, 2016; Velazquez-Campoy et al., 2004). The ability of ITC to rapidly yield accurate thermodynamic information in the absence of a label has led to its widespread use across the field of structural biology, including its use for the study of protein interactions, protein-membrane interactions, and enzyme kinetics (Ghai et al., 2012).
Chapter 3 – Characterisation of the FapF N-Terminal Domains

3.1. Introduction to Coiled-Coils

The previously reported secondary structure predictions for the periplasmic N-terminus of FapF (excluding the SEC secretion signal) indicate the presence of a 33 residue-long α-helical domain (FapF_{30-62}), and that is connected to the terminal TMD (FapF_{107-430}) by a 45 residue-long disordered linker domain (FapF_{63-106}) (Rouse et al., 2017). The periplasmic N-terminus of FapF has previously been shown to be important for the secretion of FapC into the ES, the trimerisation of FapF, and the open/closed state of the FapF pore (Rouse et al., 2017). The work in this chapter sought to provide further insight into the structure and function of the periplasmic N-terminus of FapF by use of various biophysical methods. In this chapter, FapF_{27-64} is found to self-assemble into a trimeric parallel coiled-coil. The structure of the FapF coiled-coil revealed the presence of a conserved trimerisation motif, and the importance of this motif for the trimerisation of the FapF coiled-coil was investigated (Rouse et al., 2018b). Therefore, a brief introduction to the coiled-coil structure, function, as well as a few relevant examples of the factors that govern their oligomeric states, is presented. For a more thorough review of these topics, the reader is guided to reviews by Lupas and Gruber (2005), Truebestein and Leonard (2016), and Woolfson (2005).

A coiled-coil is a protein structure in which at least two α-helices self-assemble, by twisting around each other, to result in a left-handed supercoil structure (Truebestein and Leonard, 2016). The coiled-coils are a ubiquitous structural motif: approximately 10% of eukaryotic proteins, 5% of prokaryotic proteins, and 5% of archaea proteins contain a coiled-coil structure (Liu and Rost, 2001). At the molecular level, coiled-coils accomplish a range of key functional roles, including the dictation of protein oligomeric states, as well as the mediation of protein-protein interactions (Burkhard et al., 2001; Hákansson et al., 1999; Hitchcock-DeGregori and Singh, 2010; Kuhn et al., 2014). As a result, a wide array of biological tasks such as transcription regulation and cell division are dependent on the coiled-coil structures (Barbara et al., 2007; Wang et al., 2012). Beyond their various biological functions, coiled-coil proteins have potential applications within biotechnology, biomaterials, and medicine.
Lapenta et al., 2018; Woolfson, 2010; Yu, 2002). For instance, one study was able to produce coiled-coils (via sequence-based structure design) that were utilised as the primary probes for the detection of adenomatous polyposis coli tumour-suppressor protein. In this example, coiled-coils were favoured over the use of antibodies due to their small size, low production costs, and high stability (coiled-coils do not require the formation of disulphide bonds) (Sharma et al., 1998). In another study, the potential usefulness of coiled-coils as inhibitors for HIV infection was demonstrated (Bianchi et al., 2005).

Coiled-coil domains possess a distinct periodic seven-residue amphipathic heptad repeat, conventionally labelled as (abcdefg) (reviewed by Truebestein and Leonard (2016)). The first (a) and fourth (d) residues are frequently occupied by aliphatic hydrophobic amino acids that, on average, are distributed 3.5 residues apart (Figure 15A) (Truebestein and Leonard, 2016). This distribution is consistent with the supercoil characteristics of coiled-coils, as the undistorted α-helix typically possesses approximately 3.6 residues per turn. The association of at least two α-helices enables the burial of hydrophobic residues, shielding their hydrophobic side chains from aqueous solvent. As a result, hydrophobic residues arrange into the “knobs-into-holes” motif, for which the knobs constitute the core hydrophobic side chains of one α-helix, and that are localised into a space (“holes”) between the side chains of an associated α-helix (Truebestein and Leonard, 2016). Of note, the aromatic hydrophobic side chains of Phe, Trp, and Tyr, are less common in coiled-coils, likely owing to their large steric constraints (Woolfson, 2005). The fifth (e) and seventh (g) heptad residues are frequently occupied by charged residues that enable the formation of interhelical salt bridges, and that are also important for the self-assembly of coiled-coils (including their orientations) (Apostolovic et al., 2010; Grigoryan and Keating, 2008). The second (b), third (c), and sixth (f) heptad residues usually correspond to the solvent exposed regions of coiled-coils, and resultantly, they are mostly occupied by polar or charged residues that form intrahelical salt-bridges, and that confer additional stability to the coiled-coil structure (Woolfson, 2005). The interhelical interactions between residues within parallel coiled-coil dimers and trimers are shown in Figure 15B (Apostolovic et al., 2010). A range of coiled-coil oligomers exist, with 86.8 % of all coiled-coils constituting dimers, 8.1 % constituting trimers, 4.4 % are tetramers, and less than 1 % correspond to the higher oligomeric states (CC+ database, accessed 10 March 2020; Testa et al., 2009). Furthermore, coiled-coils can
adopt parallel orientations, for which the peptide backbones are aligned in the same direction, or the anti-parallel orientation, for which the peptide backbones are aligned in opposite directions. In addition, coiled-coils can be comprised of homo- or hetero-oligomers (Apostolovic et al., 2010).

**Figure 15:** Coiled-coil formation is driven by the specific positioning and spacing of hydrophobic and polar residues within α-helices. (A) The hydrophobic residues (heptad a and d position) form layers throughout the coiled-coil structure. Figure modified from Truebestein and Leonard (2016) under creative commons license CC BY 4.0. (B) Illustration of the hydrophobic and polar interactions (depicted as a helical wheel) for a parallel dimeric coiled coil and a parallel trimeric coiled coil. Adapted from Apostolovic et al. (2010), with permission from The Royal Society of Chemistry.

One of the current and greatest challenges in the field of structural biology is the accurate prediction of a proteins three-dimensional structure from its one-dimensional amino acid sequence (Kammerer et al., 2005; Li et al., 2016; Zhang, 2008). Some progress has been made to identify the factors that influence the oligomeric state and helical packing of coiled-coils. For instance, the oligomeric state of coiled-coils has been demonstrated to strongly correlate to the knobs-into-holes packing arrangement of the core a and d sites (Harbury et al., 1993; Woolfson, 2005). Three categories of packing arrangement have been described: perpendicular, parallel, and acute packing; and that are defined by the angle of the Cα-Cβ bond vectors of each knob relative to the Cα-Cα vector of its corresponding hole (Figure 16) (described in Harbury et al. (1993); reviewed in Woolfson (2005)). In perpendicular packing, the d and d’ side chains pack alongside each other and point in the direction of the helical axis, for
which the knob vector makes an approximate 90 degree angle with the hole vector (Harbury et al., 1993). This arrangement favours the non-β-branched residues such as Leu, whilst disfavouring the β-branched Ile and Val (Woolfson, 2005). The perpendicular packing arrangement is observed for dimeric coiled-coils such as the leucine zipper GCN4 (a bZIP transcription factor), for which all heptad d positions are comprised of leucine residues (Landschulz et al., 1998; Woolfson, 2005). Dimeric coiled coils are also characterised by parallel packing at the a and a’ positions, for which the Cα-Cβ bond vectors are positioned facing outwards relative to the helical interface (Harbury et al., 1993; Woolfson, 2005). In this position, β-branched residues are common due to their capacity to project their hydrophobic side chains towards the hydrophobic interface of the coiled-coil (Woolfson, 2005). The tetrameric coiled-coils are characterised by perpendicular packing at their a and a’ sites, and parallel packing at their d and d’ sites (Woolfson, 2005). Trimeric coiled-coils are characterised by acute packing geometries, defined by an approximate 60 degree angle between the knob Cα-Cβ vectors and the corresponding Cα-Cα hole vectors, for which heptad position a resembles heptad position d of the perpendicular packing arrangement, and heptad position d resembles heptad position a of the parallel packing arrangement (Harbury et al., 1993; Woolfson, 2005). A mix of β-branched and non-β-branched residues are tolerated at these positions (Woolfson, 2005).

Figure 16: Coiled-coil geometries. The knobs-into-holes packing of three types of coiled-coil geometries (perpendicular, parallel, and acute) is schematically represented. The Cα-Cβ and Ca-Ca vectors are shown (not to scale) (Harbury et al., 1993).

Another oligomeric state determinant for the coiled coils was recently postulated in a report by Kammerer et al. (2005). The motif is comprised of R1-h2-x3-x4-h5-E6 (RhxxhE) for which: R1 = Arg; E6 = Glu, h1 = Ile/Leu/Val/Met; h5 = Leu/Ile/Val; x = any amino acid residue; the polar and non-polar residues align with their
corresponding heptad positions. A copy of the RhxxhE motif is frequently observed within the short (50 amino acids or less) parallel trimeric coiled-coils, and rarely in other coiled-coil topologies, that includes parallel dimers, as well as antiparallel dimers and antiparallel trimers. The RhxxhE motif is comprised of conserved arginine-glutamate salt-bridge networks that are postulated to enable the optimal side-chain packing of core hydrophobic residues, which adopt the acute packing arrangement described above; these features have been demonstrated to be crucial for the assembly of short coiled-coil trimers (Kammerer et al., 2005). However, of note, the sole presence of a trimerisation motif is not sufficient to guarantee coiled-coil trimer formation, with other factors such as the frequency of β-branched and non-β-branched residues at the a and d heptad positions also playing a major role in the determination of a coiled-coil oligomeric state (Ciani et al., 2010; Kammerer et al., 2005; Woolfson, 2005; Xu and Minor Jr, 2009).

The prediction of a coiled-coil oligomeric state remains a difficult and complex task. This is reinforced by the fact that a number of single mutants have been demonstrated to alter the oligomeric state and strand orientation of coiled-coils (Kammerer et al., 2005; Yadav et al., 2006). As more sequence-to-structure based data is gathered, the accuracy of coiled-coil structure predictions will continue to improve, enhancing our capacity to exploit coiled-coils for applications within medicine (Li et al., 2016; Sharma et al., 1998).

3.2. Results
3.2.1 FapF is Predicted to Possess a Conserved Coiled-Coil

The α-helical N-terminus of FapF UK4, FapF29−63, is highly conserved amongst the Pseudomonas (Figure 17). Of note, FapF1−24 corresponds to the FapF signal sequence; all descriptions of FapF in this chapter exclude this sequence. Truncating FapF to remove its periplasmic N-terminus disrupted the trimer state of FapF, suggesting that FapF29−63 may be responsible for increasing the propensity of FapF trimer formation and increasing the trimer stability (Rouse et al., 2017). One way this might occur is through the self-association of the FapF29−63 into a trimeric coiled-coil structure, and that might drive the trimerisation of FapF.
Figure 17: FapF multiple sequence alignment. The alignment of the FapF UK4 amino acid sequence with its homologs amongst some of the members of the *Pseudomonas* species are shown. The residues are numbered according to the FapF UK4 sequence. The sequence similarity is coloured according to a percentage equivalence, that accounts for the physico-chemical properties for each residue; residues highlighted red are completely conserved; and residues coloured red are highly conserved. The secondary structure for FapF UK4 is also predicted and presented above the alignment; the coil, arrow, and gaps correspond to an α-helix, β-sheet, or no secondary structure, respectively. The predicted signal sequence of FapF UK4 is highlighted in blue. Of note, the secondary structure presented for FapF UK4 is derived from PDB ID: 5O65 and PDB ID: 6FUE, respectively; the secondary structure for all other FapF residues are predicted using the PSI-PRED server (Rouse et al., 2017, 2018b).
To predict whether a coiled-coil domain is present within FapF, the MARCOIL server was utilised (Delorenzi and Speed, 2002). The N-terminus of FapF is predicted with high probability to be a coiled-coil; more specifically, the probability of coiled-coil formation at FapF$_{27-64}$ is 100% (Figure 18A). Five heptad repeats are predicted for FapF$_{27-64}$ by MARCOIL, which are shown in Figure 18B.

![Figure 18: Coiled-coil predictions of FapF$_{27-64}$](image)

The oligomeric state of the predicted FapF coiled-coil was analysed by the LOGICOIL algorithm. The LOGICOIL program predicts, for a given primary sequence, the probability of various coiled-coil oligomeric states and helix orientation(s). LOGICOIL is currently regarded as the paramount program for predicting the coiled-coil oligomeric state due to its higher accuracy compared to other programs (Li et al., 2016; Vincent et al., 2013). LOGICOIL generates a probability score for the various oligomeric states: a score above one indicates a higher probability for that oligomeric state occurring, and vice-versa for a score below
1 (Vincent et al., 2013). After input of the FapF<sub>27-64</sub> sequence, LOGICOIL indicates a slightly higher probability for the homo-dimer state (1.16) compared to the homo-trimer (1.10) or homo-tetramer (1.04) states. However, the difference between the probabilities of each oligomeric state is small, thus, the prediction does not provide a clear outcome. Looking beyond the predictions of LOGICOIL, FapF<sub>27-64</sub> possesses a RhxxhE motif (with regards to sequence- rather than structure-based homology; FapF<sub>57-62</sub>), and that may drive the FapF<sub>27-64</sub> to form a trimer (Kammerer et al., 2005; Vincent et al., 2013). Furthermore, FapF<sub>57-62</sub> is conserved in sequence amongst species of the *Pseudomonas*, highlighting its potential structural or functional importance (Figure 17). Additionally, the predicted heptad sequence of FapF<sub>27-64</sub> reveals a mix of hydrophobic β-branched and non-β-branched unbranched residues at the heptad a and d positions, and that also points towards the formation of a coiled-coil trimer. However, experimental work is required to confirm the presence of α-helicity and elucidate the oligomeric state of FapF<sub>27-64</sub>, and that is conducted within this chapter.

### 3.2.2. Biophysical Techniques Indicate the Presence of a Coiled-Coil at the FapF N-terminus

To determine the secondary structure composition of FapF<sub>27-64</sub>, a FapF<sub>27-64</sub> peptide was subjected far-UV circular dichroism (CD) spectroscopy. To analyse the FapF<sub>27-64</sub> peptide (FFaP) by CD, a FFaP sample was prepared by dissolving a 1 mg lyophilised stock into phosphate buffer and diluted to a final concentration of 0.20 mg/ml. The CD profile displayed negative bands at 208 nm and 222 nm that is a strong indicator for the presence of α-helical secondary structure (Figure 19A) (Greenfield, 2006b). The BeStSel webserver was used to deconvolute and quantify the secondary structure composition of the sample. This deconvolution protocol reported higher predictive accuracy compared to other protocols available (Micsonai et al., 2015, 2018). The BeStSel predictions indicated a secondary structure composition that is 64 % α-helical, 16 % “others” (that includes unordered regions or/and additional helical content), and 9 % β-stranded (Figure 19B). As predicted, this indicates that the FFaP is predominantly α-helical. In addition, the likelihood of an α-helix existing within a coiled-coil structure, or in isolation, can be inferred from the ratio of ellipticity at 222/208 nm. A ratio greater than 1.0 typically indicates the presence of a coiled-coil
structure, and a ratio below approximately 0.9 usually indicates the presence of an isolated helix (Crooks et al., 2011; Kwok and Hodges, 2004; Lau et al., 1984; Zhou et al., 1992). The Δ222/208 ratio calculated is 1.07, supporting the presence of a coiled-coil structure. Data from this work was published in Rouse et al. (2017).

![Figure 19: Far-UV CD spectrum of the FapF_{27-64} peptide.](image)

Figure 19: Far-UV CD spectrum of the FapF_{27-64} peptide. (A) The far-UV CD profile for FapF_{27-64} is typical of a protein containing a high proportion of α-helical content. (B) The secondary structure composition of the FapF_{27-64} peptide was predicted based on the CD profile in (A) by the BeStSel server, and the results are depicted as a pie chart.

The stability of the FFaP was studied by collecting far-UV CD spectra over a temperature range of 20 to 80 °C. The CD spectra collected over this range reveals a decrease in helicity (determined by the decrease in 222 and 208 nm band absorption) as the sample temperature is increased (Figure 20). In addition, the Δ222/208 ratio that is calculated for each CD thermal spectrum decreases as a function of temperature. Raising the temperature of the sample to 40 °C reduced the Δ222/208 ratio to 0.99, suggesting that the coiled-coil is partially destabilised; after raising the sample temperature to 70 °C, the Δ222/208 ratio had further decreased to 0.84, suggesting that the coiled-coil had been destabilised, and as a result, dissociated into individual FapF_{27-64} chains. However, even at a sample temperature of 80 °C, partial secondary structure content is retained, suggesting that the FFaP is highly stable. Data from this work was published in Rouse et al. (2018b).
Figure 20: FapF<sub>27-64</sub> is thermostable. The far-UV CD data of the FapF<sub>27-64</sub> peptide was collected as a function of temperature, ranging from 20 °C up to 80 °C.

To characterise the oligomeric state of the FFaP peptide, SEC-coupled multiple angle light scattering (MALS) was performed. The FFaP analysed was dissolved in phosphate buffer at 2 mg/ml and 0.1 ml was injected into a column. The elution profile reveals a single monodisperse peak with an average molecular mass of 13.5 kDa (±5 %), corresponding to a trimer (theoretical MW of the FFaP monomer is 4.5 kDa) (Figure 21). Together with the CD data, this suggests that FFaP forms a trimeric coiled-coil. The crystal structure of FFaP was later solved by Dr Sarah Rouse, confirming that the FFaP forms a parallel trimeric coiled-coil (referred to as “FapFcc” to describe the trimeric state of FapF<sub>27-64</sub>) (Rouse et al., 2018b). The structure of the FapFcc is later discussed. Data from this work was published in Rouse et al. (2017).
Figure 21: SEC-MALS confirms the trimerisation of FapF$_{27-64}$. The SEC-MALS data for the FapF$_{27-64}$ peptide indicates the presence of a single species with an average MW of ~13.5 kDa (the average MW value is derived from the solid red line corresponding to the molar mass) that corresponds to a FapF$_{27-64}$ trimer. The MW of the sample is consistent across the RIU peak (solid black line) indicating that the sample is monodisperse.

3.2.3. The FapF Coiled-Coil Domain is Asymmetric

Solving the crystal structure of FapFcc was a collaborative effort that incorporated the structural insight supplied by solution NMR spectroscopy (provided by this work), with crystallographic data that was obtained and processed by Dr Sarah Rouse. Dr Sarah Rouse had successfully crystallised the FapFcc peptide and X-ray crystallographic data was collected to 1.8 Å. However, after multiple stages of refinement there remained ambiguity in the positions of Tyr44 sidechain density (personal communication with Dr Sarah Rouse) (Rouse et al., 2018b).

In a previous study, the crystal structure of human lung surfactant protein D (hSP-D) was published. The hSP-D structure includes a trimeric α-helical coiled-coil, and interestingly, a single tyrosine sidechain was projected into the centre of the coiled-coil representing a deviation from three-fold symmetry, the first example of its kind (Figure 22) (Håkansson et al., 1999). FapF$_{27-64}$ contains a single tyrosine residue (Tyr44), and the initial attempts at FapFcc refinement (performed by Dr Rouse) assumed the three-fold symmetry of FapFcc, in which all the Tyr44 side chains are surface exposed (personal communication with Dr Sarah Rouse) (Rouse et al., 2018b). In order to improve the refinement statistics and gather insight into the structure of FapFcc, this work aimed to determine whether the FapFcc Tyr44 side chain possessed asymmetry, akin to the asymmetric tyrosine of the hSP-D coiled-coil.
The crystal structure of human lung surfactant protein D (hSP-D) trimeric fragment. The N-terminal parallel trimeric coiled-coil domain is coloured grey (residues 203-227 are transparent); the remaining C-terminal region of hSP-D is coloured sea green (shown as cartoon) (Håkansson et al., 1999). Tyr228 residues (chains A, B and C) are shown as sticks and coloured red, illustrating its asymmetry.

To solve this question, the 2D $^1$H-$^1$H nuclear Overhauser effect spectroscopy (NOESY) NMR experiment was employed. The NOESY experiment is capable of providing information on through space proton-proton correlations (measured by the nuclear Overhauser effect; NOE) up to approximately 5 Å, over two dimensions, and that does not require $^{15}$N- or $^{13}$C-isotopically labelled protein. The intensity of NOE correlations (herein referred to as the “NOE”; cross-peaks) between two protons is inversely proportional to $r^6$ ($r$ is distance between the two protons) (reviewed in Breukels et al. (2011) and Wider (1998)). Therefore, the nature of the 2D $^1$H-$^1$H NOESY NMR experiment enables it to inform on the presence of inter- and intramolecular correlations (1H-1H distances) between Tyr44 Hε and Hδ sidechain protons, and that has the capacity to shed insight into the symmetry or asymmetry of Y44 within the FapFcc. Tyrosine Hε and Hδ sidechain atoms are typically characterised by their distinct downfield chemical shifts between 6.5 and 8.5 ppm ($^1$H; Biological Magnetic Resonance Data Bank), together with the presence of relatively intense intramolecular NOE crosspeaks as a result of the sub~2.7 Å proximity between Hε and Hδ protons (Ulrich et al., 2007). Since FapFcc possesses no other aromatic-containing residues, the Y44 Hε and Hδ protons were expected to be easily identified.
on the $^1$H-$^1$H NOESY spectrum due to their distinct spectral characteristics, in comparison to the aliphatic resonances.

The $^1$H-$^1$H NOESY experiment was recorded and as expected, intense diagonal- and cross-peaks corresponding to Y44 Hε and Hδ sidechain protons were located in a relatively peak sparse region of the spectrum (6.3 to 7.3 ppm), downfield of the densely packed aliphatic peaks (0 to 5 ppm) (Figure 23A and B). Six diagonal peaks corresponding to the Y44 side chain protons are expected (assuming rapid flipping of the Tyr44 ring, relative to the NMR time scale), and six strong and distinctive peaks of unique chemical environments could be identified (labelled 1 to 6 in Figure 23B) (Campbell et al., 1975; Wagner et al., 1976; Wüthrich and Wagner, 1978). Of note, peaks 7 and 8 (indicated in Figure 23B), that are of relatively low signal intensity, may correspond to slow conformational exchange motions of one or more tyrosine side chain protons, of a small spin population, on the NMR time scale (Kleckner and Foster, 2011). Despite this, the strong intensities of peaks 1-6 peaks suggests that Y44 does not undergo any significant exchange motions on the NMR timescale. The assignment of these peaks to Tyr44 is later confirmed by site-directed mutagenesis (section 3.2.6).
Figure 23: 2D $^1$H-$^1$H NOESY spectrum of FapF$_{27-64}$. (A) The aliphatic and aromatic regions of the 2D $^1$H-$^1$H NOESY spectrum recorded for FapF$_{27-64}$ peptide are shown. The square box indicates the position of the Y44 side chain peaks on the diagonal in the spectrum, displayed in (B). Sample conditions: 500 µM, 293 K, 40 scans, performed on the Avance III HD 800 MHz spectrometer with cryoprobe. (B) The three sets of Y44 side chain peaks on the diagonal are labelled “A” (Tyr A; dotted blue), “B” (Tyr B; dotted cyan), and “C” (Tyr C; dotted purple), and numbered. Intra-residue cross-peaks are identified using a dot of a colour corresponding to their respective diagonal peaks; inter-residue NOE crosspeaks are identified using a red dot. Peaks 7 and 8 (referred to in the text) are indicated.
The Y44 crosspeaks possessing the greatest signal intensities correspond to intramolecular proton-proton NOEs within each Tyr44 side chain. Three sets of resolved Y44 spin systems were identified (labelled Tyr A, Tyr B and Tyr C), that correspond to the greatest through-space correlations between the diagonal Y44 protons at peak positions: 1 and 3 (Tyr A), 2 and 4 (Tyr B), and 5 and 6 (Tyr C). The remaining crosspeaks of relatively lower signal intensities between the Y44 side chain protons correspond to intermolecular Y44 Hε and Hδ proton NOEs. The Tyr A, B, and C protons each possess distinct chemical environments, indicating that the Tyr44 side chains are asymmetric. Interestingly, the spin systems corresponding to Tyr B possess strong intermolecular correlations (high intensity crosspeaks) to Tyr A, and weak correlations (low intensity crosspeaks) to Tyr C, suggesting that Tyr B is proximal (within ~5 Å) to both Tyr A and C. In contrast, no through space correlations between Tyr A and C are evident, suggesting that the distance between these Tyr side chains is greater than 5 Å. The unique proximity of the Tyr B to both Tyr A and C suggests that the side chain of Tyr B is projected and buried into the centre of FapFcc, akin to hSP-D. In contrast, Tyr A and Tyr C are not projected into the FapFcc centre, and must therefore be surface exposed. Taking this data into account, Dr Sarah Rouse subsequently processed the FapFcc diffraction data to model the single buried Y44 side chain. As a result, the refinement statistics were improved. Furthermore, the molecular simulations performed by Dr Rouse confirmed the stability of the single core-buried Y44 side chain within the refined model, and the FapFcc crystal structure was subsequently published (PDB ID: 6FUE) (Rouse et al., 2018b). The asymmetric Tyr44 is illustrated in Figure 24. Data from this work was published in Rouse et al. (2018b).

Figure 24: Tyr44 of FapFcc. Side-views and top-down (truncated) views of the FapFcc crystal structure (illustrated in cartoon form; grey), displaying the localisation and orientation of Y44 (residue and side chain shown as sticks; coloured red). PDB ID: 6FUE (Rouse et al., 2018b).
3.2.4. Expression and Purification of $\text{FapF}_{27-64}$

To enable additional NMR studies, the $\text{FapF}_{27-64}$ was sub-cloned for recombinant expression in *E. coli*. $\text{FapF}_{27-64}$ was sub-cloned (using the LIC cloning method) from a pMMB190 vector containing the *P. sp. UK4* fapABCDEF operon into the pNIC28-Bsa4 expression vector (a pET28a-derivative), for cytoplasmic expression and biophysical characterisation. Of note, this vector is ideal for recombinant overexpression as it possesses a T7 promoter, a lactose operon regulatory system, and kanamycin resistance (Studier and Moffatt, 1986). The vector also includes an N-terminal hexahistidine affinity tag (his6tag) and a downstream TEV protease cleavage site that will be used for a second Ni-IMAC-based purification step (reverse Ni-IMAC) (Savitsky et al., 2010). The FapFcc construct was transformed into *E. coli* BL21 (DE3), a T7 overexpression host that is protease deficient (Jeong et al., 2009).

Schematics of the constructs used in this study are shown in Figure 25.

**Figure 25: Schematics of the domain organisation of FapF and the FapF deletion mutants utilised.** The domain boundaries of FapF, as approximated by Rouse et al. (2017), are illustrated as a schematic. Each domain is labelled according to its position in the amino acid chain of FapF, and is individually coloured as in Figure 7A (Rouse et al., 2017). The FapF deletion mutants utilised within this chapter are also illustrated, and are colour coded according to the descriptions for Figure 7A. The N-terminal fusion tag of the two constructs utilised within this study, $\text{FapF}_{27-64}$ and $\text{FapF}_{27-81}$, is abbreviated to “NFT”; its corresponding amino acid sequence is “MHHHHHHSSGTENLYFQSM” (written N- to C-terminus; detailed within section 2.1.5.).

Expression trials for FapFcc were conducted by examining its expression into the soluble and insoluble fractions under three conditions: 18 °C for 14 hours, 30 °C for 14 hours, and 37 °C for 4 hours. Analysis by SDS-PAGE indicates that FapFcc (theoretical molecular weight (MW) is 6.9 kDa) is expressed in both the soluble and insoluble fractions over all conditions tested (Figure 26). The highest level of both
soluble and insoluble expression resulted from a 14 hour incubation at 30 °C. This condition was taken forward for FapFcc expression in large volumes. Of note samples for SDS-PAGE analysis were prepared by this author, the SDS-PAGE gel was prepared by Zena Stead, and the SDS-PAGE gel image was produced by this author.

**Figure 26: Expression trials of FapFcc.** SDS-PAGE gel analysis of soluble “S” and insoluble “I” fractions from an expression trial across a range of temperature conditions (18 °C, 30 °C, 37 °C) indicates the relative quantity of FapFcc expressed for each condition (corresponding gel band is indicated by the red arrow). Soluble expression for FapFcc appears to be greatest at 30 °C. Of note, the samples for SDS-PAGE analysis were prepared by myself; the SDS-PAGE gel was prepared by Zena Stead; and the gel was subsequently imaged by myself.

Recombinant FapFcc was purified in a four-step procedure, comprising Ni-IMAC, TEV protease digestion, reverse Ni-IMAC, and SEC purification steps. The eluted product of the first Ni-IMAC purification step produced a highly heterogeneous mixture containing FapFcc, as identified by SDS-PAGE analysis (Figure 27A). To enhance the sample purity and minimise any possible interference from the his6tag on downstream biophysical experiments, the his6tag was removed by TEV-protease digestion, and a reverse Ni-IMAC step was incorporated in order to separate FapFcc (post-TEV cleavage; his6tag removed) from the non-specific proteins (impurities). SDS-PAGE analysis indicates that this step significantly improved the purity of FapFcc (Figure 27B). A final preparative size exclusion chromatography step was incorporated to maximise sample purity and remove any higher oligomeric species (aggregates) potentially present. The sample was concentrated to 1 mg/ml and 2 ml was injected into the SEC column. The SEC UV trace is shown in Figure 27C. SDS-PAGE analysis of the eluent indicates that a high level of sample purity is achieved (Figure 27D). Approximately 5 mg of purified FapFcc was obtained from 1 L of culture grown in LB media.
Figure 27: Purification of FapFcc and Y44A. (A) An SDS-PAGE gel of an example Ni-IMAC purification for FapFcc (and FapFcc mutants). The soluble fraction after cell lysis (Sol), Ni-IMAC flow through (FT), wash steps (20; 30; 40; the number corresponds to wash buffer imidazole concentration (mM)), and elution (“E”), are shown. A large proportion of FapFcc does not bind to the column, as indicated by the presence of FapFcc in the FT, which is possibly due to saturation of the column resin binding sites; little or no FapFcc is lost during the wash steps. FapFcc is successfully eluted from the column, albeit, at a low purity. (B) SDS-PAGE gel analysis of FapFcc before (“E”) and after his6tag removal (TEV) by TEV protease cleavage indicates the successful removal of the FapFcc N-terminal his6tag. The flow-through (FT) was collected after reverse Ni-IMAC, indicating a significant increase in FapFcc purity. (C) SEC (UV 280 nm) trace of the FapFcc after application of the FapFcc reverse Ni-IMAC FT fraction to a SEC column. (D) SDS-PAGE gel analysis of the specific SEC fractions collected for FapFcc, indicating that FapFcc is eluted in high purity between 72 to 76 ml. (E) SDS-PAGE gel analysis of the specific SEC fractions collected for Y44A, indicating that Y44A is eluted in high purity between 74 to 78 ml.

3.2.5. Analysis of the FapFcc Backbone Symmetry

To determine whether the FapFcc backbone possesses three-fold symmetry, a recombinant $^{15}$N-labelled sample of FapFcc WT was prepared (as described in section 3.2.4.) and used to collect a 2D $^1$H-$^{15}$N HSQC spectrum (herein referred to as “HSQC”). The FapFcc HSQC spectrum exhibits poor chemical shift dispersion that is characteristic of $\alpha$-helices (Figure 28) (Kwan et al., 2011). The extent of amide backbone symmetry or asymmetry of the FapFcc can be gauged by comparing the number of peaks present on the HSQC spectrum to the number of peaks expected for
FapFcc that possesses three-fold backbone symmetry (47 peaks; 1 chain) or complete backbone asymmetry (141 peaks; 3 chains). The spectrum displays 51 peaks, which is four more than expected for a single chain. Most of the peaks counted on the HSQC spectrum are broad and consistent in size, which likely results from the overlap (chemical equivalence) of identical residues from each FapFcc chain. This indicates that the majority of the FapFcc backbone possesses three-fold symmetry. Of note, the seven peaks downfield of 127 ppm (\(^{15}\)N) were not counted, as these likely correspond to degradation products (Kwan et al., 2011). Together with the \(^1\)H-\(^1\)H NOESY data for FapF\(_{27-64}\), this data suggests that the asymmetry of the FapFcc is localised to the side chains of Tyr44. Data from this work was published in Rouse et al. (2018b).

![Figure 28: The FapFcc backbone possesses three-fold symmetry. \(^1\)H-\(^{15}\)N HSQC spectrum for [U-\(^{15}\)N]-FapFcc. Sample conditions: 100 µM, 298 K, 348 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe (Rouse et al., 2018b).](image)

### 3.2.6. Y44 is Not Essential for FapFcc Stability

To probe the importance of Y44 for the stability of FapFcc, as well as assign the postulated Y44 peaks on the \(^1\)H-\(^1\)H NOESY spectrum, a FapFcc Y44A mutant was subject of this investigation. Site-directed mutagenesis was performed by Grace Wu who substituted Tyr44 to Ala within the FapFcc construct. Grace Wu kindly provided
the FapFcc construct containing Y44A (referred to as “Y44A”). Y44A was expressed in M9 media and purified according to the methods previously described for FapFcc. Of note, Y44A does not contain any aromatic residues, thus, does not possess a SEC UV trace at 280 nm (Simonian and Smith, 2006). Despite this, Y44A was identified in fractions corresponding to the elution volumes recorded for FapFcc by SDS-PAGE gel analysis (Figure 27E).

In order to identify the peaks corresponding to Tyr44 on the $^1$H-$^1$H NOESY spectrum of FapFcc previously shown, the purified Y44A sample was exchanged into deuterated buffer and concentrated to approximately 2 mg/ml (concentration determined by the Bradford assay) in order to perform the 2D $^1$H-$^1$H NOESY NMR experiment (Simonian and Smith, 2006). The majority of the aliphatic diagonal and cross-peaks displayed considerable overlap between the spectra for the WT (FapFcc) and Y44A, suggesting that Y44A did not affect the global conformation of the FapFcc (Figure 29A). The diagonal and cross-peaks in the aromatic region of the spectrum that are predicted to correspond to the Y44 Hε and Hδ sidechain protons are no longer evident in the NOESY spectrum for Y44A, confirming their assignment (Figure 29B).
Figure 29: Y44 assignment. Substitution of the FapFcc Y44 to Ala does not disrupt the FapFcc conformation and enables the assignment of Y44 residues. (A) The aliphatic and aromatic regions of the 2D $^1$H-$^1$H NOESY spectrum recorded for FapFcc (positive peaks are blue, negative peaks are cyan) and Y44A (positive peaks are red, negative peaks are pink) are shown. The square box indicates the position of the Y44 side chain peaks on the diagonal, displayed in (B). Sample conditions: 600 µM, 293 K, 80 scans, performed on the Avance III HD 800 MHz spectrometer with cryoprobe. (B) Close-up of the Y44A $^1$H-$^1$H NOESY spectral region indicated in (A).
In order to compare the secondary structure composition of the WT and Y44A, purified Y44A was diluted to approximately 0.21 mg/ml and subjected to far-UV CD spectroscopy. The CD spectrum of the WT and Y44A share a similar CD profile, with negative bands similarly present at 208 nm and 222 nm that indicate the presence of helicity (Figure 30A) (Greenfield, 2006b). Notably, the Δ222/208 nm is slightly reduced for Y44A compared to WT, calculated at 1.01 and 1.07, respectively. This suggests that the Y44A mutation has only a minor destabilising effect on FapFcc, however, the substitution is not sufficient to destabilise the FapFcc complex. To assess the stability of Y44A and compare it to the WT, thermal unfolding profiles were recorded by measuring the CD signal at 222 nm over a temperature range of 20 to 90 °C for both samples. The ellipticity changes were measured at 222 nm as this wavelength is a strong indicator of helicity. Furthermore, measurements at 222 nm possess a higher signal-to-noise than the lower wavelengths (DiNitto and Kenney, 2012). The melting profiles for both the WT and Y44A reveal sigmoid curves that are typically exhibited by coiled coils due to the mechanics of coiled-coil unfolding (Figure 30B) (Hamed and Keten, 2014; McAllister et al., 2008). The melting temperature ($T_m$), defined as the temperature at which the concentration of folded protein is equal to the concentration of unfolded protein, was calculated. Of note, the descriptions of $T_m$ in this study do not include any analysis of the reversibility of protein unfolding (Cimmperman et al., 2008; Greenfield, 2006a). The $T_m$ of the FapFcc wildtype (WT) and Y44A is 76.5 °C and 73.5 °C, respectively. Together, the above data indicates that Y44A has little effect on the FapFcc secondary structure composition, and that Y44 possesses a minor role in stabilising the FapFcc conformation. Data from this work was published in Rouse et al. (2018b).

Figure 30: Far-UV CD spectrum and melting curve of FapFcc Y44A. (A) The CD profiles for recombinant FapFcc (WT; black) and Y44A (pink), are overlapped. The CD profiles for both samples are typical of a protein containing a high proportion of α-helical content. (B) The CD melting curve profiles of the recombinant WT and Y44A (coloured according to (A)).
3.2.7. FapFcc Possesses a Conserved Trimerisation Motif

The FapFcc possesses a RhxxhE site that is located at its C-terminus (FapF\textsubscript{57-62}), and the MSA indicates that this sequence is conserved amongst the FapF homologs of the \textit{Pseudomonas}, suggesting that it is important for the structure and/or function of FapF (Figure 17). To determine whether the RhxxhE site of FapFcc is structurally conserved, the structure of the RhxxhE site from both ccCor1 and FapFcc were superimposed, as shown in Figure 31A, indicating that the equivalent side chains of the two structures share strong structural homology (Kammerer et al., 2005). The FapFcc RhxxhE site may also possess a conserved salt bridge (between R57 and E62), as well as the presence of a conserved water-mediated hydrogen bond network between the R57 carbonyl oxygen and the side chain carboxylate of E62 (Figure 31B). Lastly, the hydrophobic core residues of the FapFcc RhxxhE site (Val58 and Val61) adopt the acute packing geometry, and that is also a structurally conserved feature of the RhxxhE motif (Kammerer et al., 2005). The RMSD between the corresponding atoms of the RhxxhE motifs was 0.345 Å. Together, this indicates that FapFcc possesses the RhxxhE motif.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure31.png}
\caption{FapF\textsubscript{27-64} contains a structurally conserved RhxxhE motif. (A) The RhxxhE motif of ccCor1 (purple) is overlapped with that of FapF\textsubscript{27-64} (cyan) (depicted in cartoon form; spheres correspond to H\textsubscript{2}O molecules). The individual residues for chain B of ccCor1 and chain F for FapF are distinctly coloured according to each residue (as indicated in the figure), highlighting their structural conservation; ccCor1 PDB ID: 2AKF; FapF\textsubscript{27-64} PDB ID: 6FUE (Kammerer et al., 2005) (B) The Arg-Glu electrostatic bond and Arg-H\textsubscript{2}O-Glu hydrogen bonding network (depicted as dotted lines; length shown in angstroms) within the FapF\textsubscript{27-64} RhxxhE motif are shown. Image produced using PyMOL. FapF\textsubscript{27-64} PDB ID: 6FUE (Rouse et al., 2018b).}
\end{figure}
3.2.8. The RhxxhE Motif is Required for FapFcc Trimerisation

To gain further insight into the importance of the RhxxhE motif for FapFcc trimerisation and stability, two mutant FapF27-64 constructs were designed to disrupt the RhxxhE motif. The first construct includes an Arg57 to Ala substitution. Alanine was chosen to knock out the electrostatic and hydrogen bond interactions of R57, and the relatively small size of its side chain would minimise the probability of steric-related problems compared to other bulkier residues (Kammerer et al., 2005; Taylor, 1986). The second construct contains an Arg57 to Glu substitution. This was chosen to investigate the effects of an opposite charge mutant within the RhxxhE motif. Of note, the site-directed mutagenesis was performed by Grace Wu, who substituted Arg57 to Ala in one FapF27-64 construct, and Arg57 to Glu in a second FapF27-64 construct (the FapF27-64 construct was previously described). Grace Wu kindly provided both constructs. The FapF27-64 R57A and FapF27-64 R57E mutants (referred to as “R57A” and “R57E”, respectively) were expressed in LB media and purified using the methods previously described for FapF27-64.

To investigate the effects of R57A and R57E on the oligomeric homogeneity (or heterogeneity) of the FapFcc, SEC was employed, and the UV traces for each mutant were compared to that of the wild type (WT). Interestingly, the SEC profiles for R57A and R57E (their presence was confirmed by SDS-PAGE analysis) demonstrate significant broadening over later elution volumes, which includes the presence of two shoulder peaks of sequentially lower absorbance, in comparison to the WT (Figure 32A, B, and C). The main SEC peak (possessing the highest UV absorbance value) elutes over a volume of approximately 10 ml; R57A and R57E elute over a volume of approximately 13 ml and 16 ml, respectively. The broadening and resolution of these late eluting shoulder peaks was noticeably higher in the SEC profile for R57E compared to R57A. These results indicate that the R33A and R33E mutants increase the heterogeneity of the FapFcc oligomeric state compared to FapFcc WT, with the heterogeneity more pronounced for R33E, compared to R33A. Furthermore, the SEC profiles for R57A and R57E indicate that the pinnacle of their major peak elutes 1.0 ml and 1.5 ml later than the pinnacle of the WT elution peak, respectively. Together, the SEC data indicates that each mutant increases the oligomeric heterogeneity of the FapFcc, and that is more pronounced for R57E compared to R57A. It also suggests that the trimer state of the FapFcc is somewhat destabilised, with a small proportion...
of lower MW species (potentially a dimer and monomer of FapF27-64) being generated, and that these states (or more) exist within an equilibrium (Hong et al., 2012). Data from this work was published in Rouse et al. 2018b.

Figure 32: SEC analysis of the R57A and R57E mutants. (A) The normalised SEC (UV 280 nm) traces for FapFcc (WT; solid thick line), R57A (dotted line), and R57E (solid thin line), are overlaid. (B) SDS-PAGE gel analysis of the specific SEC fractions collected for R57A, indicating that R57A is eluted in high purity between 74 to 80 ml. (C) SDS-PAGE gel analysis of the specific SEC fractions collected for R57E, indicating that R57E is eluted in high purity between 74 to 84 ml.

To probe the oligomeric state of each mutant, the SEC elution fractions between 75.0-75.5 ml and 75.5-76.0 ml were collected for R57A and R57E, respectively, and their average molecular weights (MWs) were determined by SEC-MALS. The average molecular masses obtained were 12.9 kDa ± 0.5, kDa and 11.7 kDa ± 0.5 kDa, for R57A and R57E, respectively (Figure 33A and B). The MW of monomeric R57A and R57E is 4.69 kDa and 4.75 kDa, respectively. Thus, the MWs for R57A and R57E correspond to 2.8 FapFcc monomers and 2.4 FapFcc monomers, respectively. The SEC-MALS data also reveals that both R57A and R57E are somewhat polydisperse, with the average MW of R57A ranging from 15.5 kDa to 11.2 kDa, and R57E ranging from 11.2 to 10.5 kDa. Together, this data indicates that the trimer state of the FapFcc is somewhat disrupted by R57A and R57E, and both mutants induce an equilibrium of various oligomeric states likely comprising the FapF27-64 monomer-dimer-trimer states. Furthermore, these results indicate that R57E is more disruptive to the FapFcc trimer state than R57A; these results are in agreement with the SEC profiles previously described for each mutant. Data from this work was published in Rouse et al. (2018b).
Figure 33: SEC-MALS analysis of the R57A and R57E mutants. (A) The SEC-MALS data indicates that the R57A mutant possesses an average MW of ~12.9 kDa (the average MW value is derived from the solid red line corresponding to the molar mass). The MW of the sample varies across the RIU peak (solid black line) indicating that the sample is polydisperse. The sample concentration is 8.5 mg/ml, and 0.5 ml was loaded into the S75A column. (B) The SEC-MALS data indicates that the R57E mutant possesses an average MW of ~11.7 kDa (the average MW value is derived from the solid red line corresponding to the molar mass). Similar to R57A, minor sample polydispersity is evident. The sample concentration is 8.5 mg/ml, and 0.5 ml was loaded into the S75A column.

In order to characterise the effects of R57A and R57E on the secondary structure composition of FapFcc, each mutant was subjected to far-UV CD spectroscopy. Identical samples to those used for SEC-MALS were analysed; each sample was diluted to 0.22 mg/ml and 0.23 mg/ml, for R57A and R57E, respectively. Both R57A and R57E reveal a CD profile comparable to that obtained for the WT, with negative bands present at 208 nm and 222 nm indicating the presence of α-helicity (Figure 34A) (Greenfield, 2006b). Notably, the Δ222/208 nm (previously defined) for both mutants is significantly reduced compared to the WT. The Δ222/208 nm is 1.07, 0.99, and 0.95 for the WT, R57A, and R57E, respectively, suggesting that the FapFcc coiled-coil state is disrupted by either mutant. The thermal stabilities for R57A and R57E were subsequently determined using the methods previously described. The melting profiles for the WT, R57A, and R57E reveal sigmoid curves, as previously described (Figure 34B). The $T_m$ of the WT, R57A, and R57E is 76.5 °C, 65.5 °C, and 53.5 °C, respectively. These results indicate that both mutants reduce the stability of FapFcc, R57E more so than R57A, and that is in accordance with the results of SEC and SEC-MALS. These results highlight the importance of the RhxxhE motif on FapFcc trimer formation and stability. Data from this work was published in Rouse et al. (2018b).
Figure 34: Far-UV CD spectra and melting curves of the R57A and R57E mutants. (A) The CD profiles for recombinant FapFcc (WT; black), R57A (green), R57E (blue) are overlapped. The CD profiles for all samples are typical of a protein containing a high proportion of α-helical content. (B) The CD melting curve profiles of the recombinant WT and mutants (coloured according to (A)).

3.2.9. Biophysical Characterisation of the FapF Linker Domain

The periplasmic N-terminus of FapF is predicted to contain a 44-residue disordered linker domain (FapF₆₃₋₁₀₆; FapFL) (Figure 17) (Rouse et al., 2017). The structure and function of FapFL remains to be elucidated. The first construct to encompass the disordered linker, which is probed in this thesis, includes FapF₂₇₋₈₁. This construct was designed as it includes the highly conserved region of the disordered linker (FapF₆₃₋₇₉), unlike FapF₈₀₋₁₀₀, which is highly variable amongst the Pseudomonas (Figure 17). Furthermore, the N-terminus of FapFL contains a high frequency of proline residues (FapF₆₅₋₇₀, residues PATPAP) that may form a polyproline type II (PPII) helix (Figure 35) (O’Brien et al., 2020). Interestingly, PPII helices are often located at protein surfaces that are involved in protein-protein interactions, thus, FapF₆₅₋₇₀ may represent an interaction site for Fap substrates (Adzhubei et al., 2013).
**Figure 35: PPII predictions of FapF.** The PPIIPred web server indicates a high probability of a PPII structure at the FapF periplasmic N-terminus (FapF\(_{65-70}\)). The probability of a PPII structure was plotted against its respective FapF residue on a chart. The FapF signal sequence (FapF\(_{1-24}\)) is not shown. The PPII probabilities for FapF residues 60 to 75 are enlarged, and their corresponding residues are indicated below the sub-chart. PPII probability scores between 0.31 and 0.55 were obtained for “PATPAP” (FapF\(_{65-70}\)), with a higher score (up to 1) corresponding to a higher probability of PPII formation, and vice-versa for scores closer to 0 (O’Brien et al., 2020).

In order to begin characterising recombinant FapF\(_{27-81}\), FapF\(_{27-81}\) was first sub-cloned (using the LIC cloning method) from a pMMB190 vector containing the *P. sp. UK4 fapABCDEF* operon into the pNIC28-Bsa4 expression vector. The resulting plasmid was subsequently transformed into *E. coli* BL21 (DE3) for cytoplasmic expression trials as described previously. SDS-PAGE analysis indicates that the highest level of both soluble expression resulted from a 14 hour incubation at 30 °C (Figure 36). This expression condition was taken forward for FapF\(_{27-81}\) expression in large volumes. Of note samples for SDS-PAGE analysis were prepared by this author, the SDS-PAGE gel was prepared by Zena Stead, and this author imaged the SDS-PAGE gel.

**Figure 36: Expression trials of FapF\(_{27-81}\).** SDS-PAGE gel analysis of soluble “S” and insoluble “I” fractions from an expression trial of FapF\(_{27-81}\) across a range of temperature conditions (18 °C, 30 °C, 37 °C) indicate the relative quantity of FapF\(_{27-81}\) expressed for each condition (corresponding gel band is indicated by the red arrow). Soluble expression of the FapF\(_{27-81}\) was greatest at 30 °C. Of note, the samples for SDS-PAGE analysis were prepared by myself; the SDS-PAGE gel was prepared by Zena Stead; and the gel was subsequently imaged by myself.
The purification protocol of FapF<sub>27-81</sub> was performed as described for FapFcc, which includes multiple stages of Ni-IMAC, a TEV protease digestion step, as well as a round of SEC (Figure 37A, B, C, and D). SDS-PAGE analysis of the reverse Ni-IMAC flow-through indicates that, unlike the FapFcc, FapF<sub>27-81</sub> elutes with a low sample purity (Figure 37B). The final purification step, preparative SEC, significantly improved the sample purity (as indicated by SDS-PAGE analysis) (Figure 37C and D). Approximately 2 mg of purified FapF<sub>27-81</sub> is obtained from 1 L of culture grown in LB media.

SEC-MALS was employed to determine the average MW of the FapF<sub>27-81</sub> in order to deduce its oligomeric state. The FapF<sub>27-81</sub> possesses an average molecular mass of 20.0 kDa (±7 %), and that corresponds to the MW of a FapF<sub>27-81</sub> trimer (theoretical MW is 19.6 kDa), indicating that the addition of FapF<sub>65-81</sub> does not disrupt the

---

**Figure 37: The expression and purification of FapF<sub>27-81</sub>.** (A) SDS-PAGE gel analysis of an example Ni-IMAC purification for FapF<sub>27-81</sub>. The soluble fraction after cell lysis (Sol), Ni-IMAC flow through (FT), wash steps (20; 30; 40; the number corresponds to wash buffer imidazole concentration (mM)), and elution (E), are shown. A small proportion of FapF<sub>27-81</sub> does not bind to the column, which may be due to the saturation of the binding sites on the column resin; little or no FapF<sub>27-81</sub> is lost during the wash steps. The FapF<sub>27-81</sub> is successfully eluted, albeit, at a low purity. (B) SDS-PAGE gel analysis of FapFcc before (E) and after his6tag removal (TEV) by TEV protease cleavage indicates a partial success in the removal of the FapF<sub>27-81</sub> his6tag. The flow-through (FT) was collected after reverse Ni-IMAC, indicating a subtle increase in the purity of FapF<sub>27-81</sub> (indicated by a red arrow) compared to the first Ni-IMAC eluent. (C) SEC (UV 280 nm) trace of the FapF<sub>27-81</sub> reverse Ni-IMAC FT fraction to a SEC column. (D) SDS-PAGE gel analysis of the specific SEC fractions collected for FapF<sub>27-81</sub>, indicating that FapF<sub>27-81</sub> is eluted in high purity between 74 to 88 ml.
trimeric state of the FapF periplasmic NTD (Figure 38A). The SEC-MALS data also reveals that FapF
27-81 is near monodisperse, with the average MW of the FapF
27-81 ranging from 21.9 kDa to 19.7 kDa (trimeric FapF
27-81 will herein be referred to as “CCExt”).

**Figure 38: SEC-MALS and far-UV CD analysis of FapF\textsubscript{27-81}.** (A) The SEC-MALS data indicates that the FapF
27-81 possesses an average MW of ~20.0 kDa (the average MW value is derived from the solid red line corresponding to the molar mass) that corresponds to a FapF
27-81 trimer. The monodispersity of the sample MW across the RIU peak centre (~11.4-11.6 ml) (solid black line) suggests the presence of a single species. The presence of higher-MW impurities (as shown in Figure 37D) may drive the higher MW species observed at ~11.2 ml. (B) The far-UV CD spectra for FapF
27-81 and FapFcc are overlapped. The FapF
27-81 CD profile is typical of a protein containing a high proportion of α-helical content. The concentration of the CCExt sample is 0.23 mg/ml. (C) The secondary structure composition for FapF
27-81 was predicted based on the CD profile in (B) by the BeStSel server, and the results are depicted as a pie chart.

Far-UV CD spectra were subsequently collected for the CCExt in order to determine its secondary structure composition. Similar to the FapFcc, negative bands present at 208 nm and 222 nm that are characteristic of α-helical proteins (Figure 38B) (Greenfield, 2006b). Additionally, the CCExt presents a minor increase in molar ellipticity between 230 nm and 205 nm. This increase likely corresponds to the presence of additional disorder, and/or PPII helix structure compared to the FapFcc. Of note, the PPII helix is typically characterised by a positive band at approximately 217 nm and a negative band at 200 nm; however, it was not possible to unambiguously discern its presence here, as these wavelengths are overlapped with other secondary structure features (Bochicchio and Tamburro, 2002). The BeStSel webserver was used to deconvolute and quantify the secondary structure composition of the CD spectrum, which indicated that the secondary structure of CCExt is comprised of 38 % α-helix, 30 % “others”, and 22 % β-strand. These results indicate an increase in the proportion of sample disorder compared to FapFcc, which is
consistent with the secondary structure predictions for FapF (Figure 38C) (Micsonai et al., 2018)

To determine whether FapF<sub>65-81</sub> is disordered, a <sup>15</sup>N-labelled sample of CCExt was expressed and purified (as described previously), and a 2D <sup>1</sup>H-<sup>15</sup>N HSQC was recorded. Disordered regions of a protein generally undergo fast exchange on the NMR time scale, therefore, their corresponding peaks on the HSQC spectrum will exhibit narrow line widths and strong signal intensities (Kleckner and Foster, 2011; Mollica et al., 2016). By comparing the HSQC for the CCExt and FapFcc, it should be possible to determine the peaks (if any) that correspond to FapF<sub>65-81</sub>. Thirteen additional peaks are expected for CCExt compared to FapFcc (a total of 58 peaks), assuming three-fold symmetry (Kleckner and Foster, 2011). Intriguingly, the HSQC for both FapFcc and CCExt overlap near identically, with 51 peaks counted for both spectra (excluding the peaks downfield of ~127 ppm (<sup>15</sup>N) that likely correspond to protein degradation products) (Figure 39) (Kwan et al., 2011). The extensive peak overlap observed between the two HSQC spectra indicates that the conformation and symmetry of the amide backbone of the coiled-coil domain (FapF<sub>27-64</sub>) is unaffected by the addition of FapF<sub>65-81</sub>. The lack of any obvious additional peaks on the CCExt HSQC, compared to the HSQC for FapFcc, suggests that FapF<sub>65-81</sub> is undergoing intermediate exchange (due to exchange-broadening), and/or peaks belonging to FapF<sub>65-81</sub> are present in the central region of the spectrum (~7.5 – 8.5 ppm (<sup>1</sup>H), ~120 ppm (<sup>15</sup>N)), where there is a dense clustering of peaks, which makes the identification of new peaks more arduous. It also remains possible that FapF<sub>65-81</sub> may have undergone partial degradation. Further experimental evidence is required to clarify the stability, structure, and flexibility of FapF<sub>65-81</sub>, and this is later discussed.
3.3. Summary and Discussion

3.3.1. The FapFcc Structure

Prior to this project, a previous study had indicated that the periplasmic NTD of FapF is important for FapF trimerisation and pore regulation (Rouse et al., 2017). This study aimed to provide further insight into the structure and function of the FapF NTD, which would ultimately contribute to further our understanding of Fap biogenesis. The prior in silico analysis of FapF revealed the presence of a ~33 residue long α-helix at the N-terminus of FapF that is highly conserved amongst the *Pseudomonas* (Rouse et al., 2017). FapF27-64 was confidently predicted to form a coiled-coil spanning five heptad repeats, and based on predictions by LOGICOIL, it was unclear whether this formed a dimeric or trimeric coiled-coil. In contrast, an inspection of the predicted FapF27-64 heptad repeats indicated the presence of both β-branched and non-β-branched hydrophobic residues at the heptad a and d sites, as well as the presence of a postulated RhxxhE trimerisation motif at the FapF27-64 C-terminus; together, this suggested that FapF27-64 is likely driven towards a trimer state (Kammerer et al., 2005; Woolfson, 2005). To provide further insight into the structure
of FapF<sub>27-64</sub>, a truncated form of FapF, FapF<sub>27-64</sub>, in both recombinant and peptide form, was biophysically characterised. The initial CD spectroscopy and SEC-MALS analysis for FapF<sub>27-64</sub> indicated that its secondary structure composition is predominantly α-helical, and its MW corresponds to a FapF<sub>27-64</sub> trimer. Together, this indicates that FapF<sub>27-64</sub> forms a trimeric coiled-coil (FapFcc).

The structural characterisation of the FapFcc was propelled forward by Dr Sarah Rouse who successfully crystallised the FapFcc, and collected X-ray diffraction data at high resolution (1.8 Å) (Rouse et al., 2018b). However, the refinement statistics for FapFcc were initially poor, with ambiguity in the positions of Tyr44 sidechain density. Intrigued by a previous report that revealed the presence of a single buried and asymmetric tyrosine sidechain within the trimeric coiled-coil hSP-D, we postulated that the FapFcc Y44 side chains could also share similar asymmetric properties (Håkansson et al., 1999). The NMR experiments that were later performed strongly suggested the presence of a single buried and asymmetric Tyr44 side chain within the core of the FapFcc, and that asymmetry was specifically localised to the Tyr44 side-chain rather than an entire backbone of the FapFcc chain. This insight was provided to Dr Sarah Rouse who subsequently refined the FapFcc structure to include the single buried Tyr44 side chain, which resulted in a significant improvement in the refinement statistics. Furthermore, molecular simulations performed by Dr Rouse indicated that the single buried Tyr44 side-chain represented a stable form of the FapFcc trimer. The FapFcc structure was subsequently published (PDB ID: 6FUE) (Rouse et al., 2018b).

A search of the Coiled Coils Database (accessed 10<sup>th</sup> March 2020) reveals that the asymmetry and burial of a single tyrosine side chain in a trimeric coiled-coil is rare, with hSP-D (PDB ID: 1B08) being the only example found (Håkansson et al., 1999; Testa et al., 2009). Interestingly, the asymmetry of hSP-D was found to propagate to its carbohydrate-recognition domains, resulting in the formation of a positively charge cleft that is predicted to function as an interaction site (Håkansson et al., 1999). The asymmetry of Tyr44 may similarly result in the formation of a FapF binding site, and propagate to the FapF TMD to enable independent regulation of the helical plug position for each FapF β-barrel (Rouse et al., 2018b). Of note, substitution of the FapFcc Tyr44 to Ala did not result in any significant reductions in the stability of the FapFcc, suggesting that Tyr44 possesses a functional role within FapF (and the Fap system), which remains to be elucidated.

110
The crystal structure of FapFcc reveals asymmetric surface properties; for instance, the majority of the FapFcc surface is polar; however, a hydrophobic pocket exists in the central region of the FapFcc, which could represent a FapF-Fap substrate interaction site (Figure 40A and B) (Rouse et al., 2018b). This might allow for a sub-stoichiometric ratio of Fap substrates to be recruited to FapF, which are subsequently transported to the FapF TMD for export into the ES. It is possible that the asymmetrical surface properties of the FapFcc are present as a result of the asymmetric tyrosine, however, this was not confirmed within this study. The FapFcc core is also comprised of an extensive hydrogen bonding network; in particular, Gln23 and Gln24 form a network that was termed the “glutamine layer” or “Q-layer” by Rouse et al. (2018b) (Figure 40C and D). Buried polar residues are often reported to destabilise coiled-coils, which is consistent with the molecular simulations of the FapFcc that indicate that the side chains of Gln23-Gln24 are highly dynamic (Eckert et al., 1998; Rouse et al., 2018b). The Gln and Asn “layers” within various coiled-coils are similarly dynamic, and this dynamic nature has previously been suggested to allow for alterations of the helical packing within a coiled-coil; resultantly, this might lower the free energy required for conformational adjustments and interactions with their respective substates (Fiumara et al., 2010). Regarding Fap biogenesis, the dynamic nature of the FapFcc Q-layer may encourage transient interactions with the Fap substrates in order to direct their threading through the FapF pore, in a mechanism analogous to the postulated sugar slide function of ScrY (Dumas et al., 2000; Rouse et al., 2018b). Furthermore, the binding of FapFcc to Fap-proteins may trigger conformational changes in FapF that alters the position of the FapF helical plug. This gating mechanism would be analogous to the recent description for the channel opening mechanism of a voltage-gated sodium channel, for which channel opening is associated with coiled-coil assembly (Fujiwara et al., 2012).
Figure 40: The crystal structure of the FapF<sub>27-64</sub> peptide indicates the formation of a parallel trimeric coiled-coil. (A) The surface of FapF<sub>27-64</sub> is shown (side view) with residues coloured according to their hydrophobicity. At the centre of the coiled-coil, a large hydrophobic cleft is present and indicated by a black arrow. (B) The surface is shown (side view) and coloured according to charge, highlighting the distribution of polar residues across the structure. (C) Side-view of the FapF<sub>27-64</sub> structure (illustrated in cartoon form), displaying the localisation and orientation of Q47 and Q48 (residue and side chain shown as sticks; coloured green). (D) Top-down view of the FapF<sub>27-64</sub> structure; illustrated as described in (C). Of note, the N-terminus of the structure is truncated for the purpose of visualisation. PDB ID: 6FUE (Rouse et al., 2018b).

3.3.2. The FapFcc Trimerisation Motif

Over the past few decades, there has been a strong interest to develop an understanding of how a protein sequence relates to the structure of a coiled-coil. Short coiled-coils comprising ~15-50 amino acids are being used for a variety of applications in fields such as biotechnology and medicine, therefore, determining and detailing the factors that influence the structure of a coiled-coil is essential (Arndt et al., 2001; Bianchi et al., 2005; Kammerer et al., 2005; Sharma et al., 1998; Wang et al., 1999). The gradual accumulation of research has begun to unravel how the primary sequence of a coiled-coil relates to its structure. Recently, a report by Kammerer et al. (2005) demonstrated that a novel structural motif, the RhxxhE motif, is crucial for short parallel coiled-coil trimer assembly. Substitution of a single conserved residue of the RhxxhE motif has been shown to convert trimeric coiled-coils to alternative oligomeric states (Kammerer et al., 2005; Xu and Minor Jr, 2009). The FapFcc was found to contain one copy of this structural motif at its C-terminus.
(Rouse et al., 2018b). Furthermore, the sequence of the RhxxhE motif is conserved between various FapF homologs, suggesting that this motif serves an important structural or functional role within FapF.

This study aimed to investigate the importance of the RhxxhE motif for FapFcc trimerisation. This included the production of two independent mutants, R57A and R57E. Both mutants were designed to disrupt the R57-E62 salt bridge, and the effects of these mutants on the trimerisation of FapFcc were examined. Both R57A and R57E destabilised the FapFcc trimer, resulting in lower melting points compared to the WT. In a previously report, the substitution of Arg to Ala (corresponding to R1 of the RhxxhE motif) for the short coiled-coil trimer ccCor1 reduced its $T_m$ from 79 °C to 67 °C, for the WT and Arg to Ala mutant, respectively (Kammerer et al., 2005). The 12 °C reduction in $T_m$ is broadly consistent with the 10 °C reduction determined for FapFcc R57A, compared to FapFcc WT. Furthermore, the FapFcc R57A and R57E mutants converted the FapFcc from a trimer state to an equilibrium of monomer-dimer-trimer states; the destabilising effects were more pronounced for R57E compared to R57A. Interestingly, the ccCor1 Arg to Ala mutant converted the ccCor1 trimer to a predominantly tetrameric state (Kammerer et al., 2005). The conversion of FapFcc to a smaller species might be due to the high frequency of Leu (non-β-branched) at the heptad d positions of the FapFcc, which favours the dimeric coiled-coil state over the tetrameric state (Woolfson, 2005). The destabilising effects of these mutants on the coiled-coil trimer state is reportedly due to the destabilisation of the conserved electrostatic and hydrogen-bond networks of the RhxxhE motif, which results in the sub-optimal packing of hydrophobic core residues, thus, disrupting the trimer state (Kammerer et al., 2005). In addition, the lower stability of the R57E substitution compared to R57A is likely due to the increase in interhelical electrostatic repulsion between R57E and E62 of the FapFcc chains (Kohn et al., 1995). These results highlight the importance of the RhxxhE motif for FapFcc trimerisation.

The RhxxhE motif (structurally speaking), to date, is not taken into account by coiled-coil oligomeric state predictors, including LOGICOIL (reviewed by Li et al. (2016)) (Vincent et al., 2013). Therefore, future predictive programs should incorporate this structural motif to improve the accuracy of their algorithms (Kammerer et al., 2005; Li et al., 2016). Despite the apparent importance of the RhxxhE motif for coiled-coil trimerisation, it should be noted that the sole presence of a RhxxhE motif within a short coiled-coil is not sufficient to guarantee trimer
formation, as this motif is also reported for a small proportion of dimeric and tetrameric coiled-coils (Kammerer et al., 2005). Furthermore, its capacity to drive trimerisation can also be readily overridden by hydrophobic core components. For instance, the Kv7.1 A-domain tail consists of a parallel trimeric coiled-coil that contains the RhxxhE motif; however, the introduction of an additional heptad repeat that does not contain the RhxxhE motif, and that does not disrupt its existing heptad repeats, converted the trimer into a stable tetramer (Xu and Minor Jr, 2009). Therefore, future experiments are required to elucidate the exact contributions of the RhxxhE motif for the oligomeric state of coiled-coils. On a separate note regarding Fap biogenesis, Rouse et al. (2017) previously used the IVSA and SCCA to compare the effects of various FapF mutants on the pore state of FapF and FapFβ (Rouse et al., 2017). These techniques could be employed to investigate the effect (if any) of coiled-coil mutants that disrupt the FapF coiled-coil trimer state, such as R57A and R57E, on the FapF pore state. This would enable us to determine whether the gating of FapF is linked to, or dependent on, the coiled-coil conformation, and these results would provide insight into the pore gating mechanism of FapF. These studies may also enable us to elucidate the importance (if any) of FapF trimerisation on Fap substrate translocation, as the functional importance of the FapF trimer is currently unknown.

3.3.3. The FapF Linker Domain

The biophysical properties of the predicted flexible linker domain of FapF (FapFL) were investigated in this chapter. Previous experimental evidence indicates that the FapFL is extended, solvent exposed, and accessible to cleavage by trypsin (Rouse et al., 2017). In this study, the N-terminus of FapFL was determined to be well conserved amongst the Pseudomonas species, and it may harbour a conserved PPII helix, suggesting that it is important for the structure and/or function of FapF (Adzhubei et al., 2013). In addition, the SEC-MALS results for the CCExt reveal that the trimerisation of the FapFcc is not affected by the addition of FapF65-81 to the FapFcc C-terminus. The CCExt was subsequently analysed by CD spectroscopy, which indicated an increase in disordered content compared to the FapFcc. These results are in agreement with the secondary structure predictions, as well as the published experimental data for the FapFL (Rouse et al., 2017). In an attempt to further clarify the presence of disorder within FapF65-81, the CCExt was analysed by
NMR spectroscopy. The $^{1}$H-$^{15}$N HSQC spectrum of $^{15}$N-CCExt was near identical to that of $^{15}$N-FapFcc, indicating that CCExt maintains the backbone symmetry of FapFcc. Intriguingly, the lack of any obvious additional peaks, in comparison to the FapFcc, suggests that FapF$_{65-81}$ is undergoing intermediate exchange. However, it is also remains possible that FapF$_{65-81}$ is partially or completely degraded. Future studies should employ techniques such as mass spectrometry in order to clarify the stability and presence of FapF$_{65-81}$ in CCExt. Furthermore, the dynamics and flexibility of the CCExt can be analysed by utilising a number of NMR-based approaches that are outlined within Kleckner and Foster (2011).

In this study, it was not possible to unambiguously discern the presence of a PPII helix from the far-UV CD spectrum or $^{1}$H-$^{15}$N HSQC spectrum of CCExt; vibrational circular dichroism and Raman optical activity are optical techniques capable of unambiguously informing the presence of PPII helices (Adzhubei et al., 2013). Of note, the accuracy of the BeStSel predictions for the various CD spectra presented in this chapter is limited. For instance, the secondary structure composition of the FapFcc CD spectrum, which was deconvoluted by the BeStSel webserver server, indicates the presence of β-strands (predicted to occupy 9% of the total secondary structure composition). However, the FapFcc crystal structure reveals the sole presence of α-helical and unstructured content (Rouse et al., 2018b). This predictive error might be explained by the influence of the tyrosine that is present within FapFcc, with a report indicating that contributions of up to $\pm5000$ deg cm$^2$ dmol$^{-1}$ (mean residue ellipticity) at 220 nm can be induced, and that overlaps closely with absorptive profiles of β-strands at 218 nm (Bhattacharjee et al., 2003; Greenfield, 2006b). Aromatic contributions are not accounted for by the BeStSel server (Micsonai et al., 2018).

The function of the FapFL is yet to be elucidated. Membrane protein linker domains are commonly reported to be functionally important for channel gating. For instance, the bacterial flagellar motor activity is dependent on the conformational changes of motility protein A and motility protein B, for which the linker between these proteins plays an important role in their assembly and activation (O’Neill et al., 2011). Furthermore, a linker domain connects the voltage-gated Hv1 membrane channel to a cytoplasmic coiled-coil domain; the linker domain is reported to be essential for cooperative gating within this channel (Fujiwara et al., 2012). Also, flexible protein regions commonly host sites for interactions to various proteins, and
possess the capacity to transmit conformational changes via allostERIC mechanisms to alternative protein domains (reviewed in Ma et al. (2011) and Papaleo et al. (2016)). The FapFL may function similarly, as preliminary evidence indicates that the N-terminus of FapF is essential for pore gating (Rouse et al., 2017). Future studies should aim to investigate the structure and dynamics of the FapFL, and whether it interacts with other Fap proteins. This insight has the potential to drastically increase our understanding of Fap biogenesis.

The protocols established for the purification of the FapFcc and CCExt in this study provide a foundation for future studies to investigate whether the FapF N-terminus is able to interact with other Fap proteins. In previous reports, FapF is stabilised in buffer containing high concentrations of detergent; however, detergents have a strong capacity to prevent or disrupt native protein-protein interactions (Lee et al., 2018; Rouse et al., 2017). Therefore, the periplasmic constructs presented here, which are stable in the absence of detergents, may be suitable for enabling at least the preliminary characterisation of FapF-Fap protein interactions. Furthermore, the smaller size of the FapF N-terminal constructs, compared to full length FapF, may enable the interactions between the FapF N-terminus and Fap proteins to be studied by conventional $^{15}N/^{13}C$ NMR spectroscopic methods (Clore and Gronenborn, 1994; Kwan et al., 2011). However, it should be noted that native interactions between truncated FapF and Fap substrates might be compromised with the use of a FapF truncation.
Chapter 4 – Structure and Function Studies of FapD

4.1. Introduction

The structure of FapD and its function within the Fap system are yet to be solved. Previous evidence hints at a chaperoning and proteolytic role for FapD, a predicted C39-peptidase, and as previously described, FapD and FapF may function co-dependently (Chorev et al., 2020; Dueholm et al., 2010; Rouse et al., 2017, 2018a). In this chapter, interactions between the FapF N-terminus (CCExt) and FapD are probed, and a transient interaction between these proteins is demonstrated.

4.2. Results

4.2.1. FapD Bioinformatics Analysis

FapD is predicted to be a member of the C39-peptidase family of cysteine proteases (Dueholm et al., 2010; Rouse et al., 2017). The MSA for the FapD UK4 gene and its homologs from various Pseudomonas species indicates a high degree of sequence conservation, which includes the C39-peptidase family active site residues Q70, C76, H160, and D176, suggesting that FapD possesses a conserved structure and function (Figure 41) (Bobeica et al., 2019). Of note, FapD1-38 corresponds to the FapD signal sequence; all descriptions of FapD in this chapter exclude this sequence. The FapD N-terminus is predominantly hydrophobic, contains a mix of variable and conserved residues, and the secondary structure predictions indicate the presence of a short α-helix. The MSA also includes LahT147, which is the C39 peptidase domain of a PCAT named LahT from Lachnospiraceae bacterium C6A11 (Bobeica et al., 2019). Interestingly, the FapD N-terminus does not share sequence homology with LahT147, suggesting that it functions specifically within the Fap system; it may represent a FapF-targeting sequence akin to the N22 of CsgA, or a binding site for Fap proteins (Bobeica et al., 2019; Yan et al., 2020). The amphiphatic C-terminal α-helix (FapD227-246) of FapD contains a relatively high degree of sequence-based conservation, which is also absent in LahT147; thus, as described above, it may also possess a functionally important role that is specific to the Fap system.
Figure 41: FapD multiple sequence alignment. The alignment of the FapD UK4 amino acid sequence with its homologs (amongst some of the members of the *Pseudomonas* species), as well as LahT147, are shown. The residues are numbered according to the FapD UK4 sequence. The sequence similarity is coloured according to a percentage equivalence, which accounts for the physico-chemical properties for each residue; residues highlighted red are completely conserved, and residues coloured red are highly conserved. The predicted signal sequence of FapF UK4 is highlighted in blue. The canonical C39-peptidase active site residues within FapD (Q70, C76, H160, D176) are indicated by an asterisk (Bobeica et al., 2019). The secondary structure for FapD UK4 is also predicted and presented above the alignment; the coil, arrow, and gaps correspond to an α-helix, β-sheet, or no secondary structure, respectively.

In addition, a FapD homology model was generated by utilising the I- TASSER server, incorporating the crystal structure of LahT147 for use as a template (Bobeica et al., 2019). The multiple sequence alignment shown in Figure 41 (and Figure 9) indicates the degree of sequence conservation between FapD (UK4) and LahT147, which highlights the conservation of C39 peptidase active site residues. As the FapD N-terminus (FapD<sub>39-50</sub>) and C-terminus (FapD<sub>207-246</sub>) share no sequence homology with LahT147, these regions were excluded from the I-TASSER input sequence in order to maximise the accuracy of the I-TASSER model. The resulting sequence
conservation between FapD_{51-206} and LahT147 is 24%. The homology model of FapD was generated and shown in Figure 42A. It possesses a C-score of 0.98, which suggests that the model is a reliable prediction of the structure. The TM-score of the model is 0.85±0.08, indicating a high degree of structural similarity between the I-TASSER model of FapD and the structure of LahT147 (Yang and Zhang, 2015). The Consurf server was subsequently employed to calculate the evolutionary conservation of each residue across the FapD amino acid sequence (without its signal sequence) (Ashkenazy et al., 2016). The average conservation scores were plotted onto the FapD homology model, for which each residue is colour coded according to its conservation score (Figure 42A). The homology model generated was superimposed with that of LahT147, which is in complex with its bound substrate, to highlight the structural conservation between the PEP domains and FapD (Figure 42B and 42C) (Bobeica et al., 2019). The homology model and Consurf predictions indicate that the PEP active site residues are conserved in FapD, in both sequence and in structure. The core of the FapD homology model (FapD_{70-198}) is comprised of five α-helices that surround a central region comprised of a six-β-strand antiparallel β-sheet, and that shares strong structural homology with the folded core of LahT147 (Bobeica et al., 2019). Furthermore, the FapD homology model contains a structurally conserved hydrophobic active site and substrate-binding region (Figure 42D). The high sequence and structural homology between FapD and LahT147 suggests that FapD may bind and process substrates using a similar binding pocket and mode of action as the C39-peptidase family (Bobeica et al., 2019).
Figure 42: A FapD homology model indicates the presence of a structurally conserved C39-peptidase domain. (A) The FapD homology model generated by I-TASSER is illustrated in cartoon form (template PDB ID: 6MPZ; the PEP-domain is depicted) (Lin et al., 2015). The evolutionary conservation of each residue across the FapD amino acid sequence was determined by Consurf, and each residue on the FapD homology model was subsequently coloured according to its level of conservation. (B) Superimposition of the FapD homology model to the crystal structure of LahT147 (PDB ID: 6MPZ) (blue) in complex with its substrate (green). The active site residues are shown as ball-and-stick. (C) Close-up of the active site region from (B) indicates that the FapD homology model active site residues are conserved. The residue numbers indicated correspond to those of FapD. (D) The FapD model contains a hydrophobic active site akin to LahT147. Hydrophobic residues are coloured orange; all FapD, LahT147, and LahT147 substrate non-hydrophobic residues are coloured grey, blue, and green, respectively; PDB ID: 6MPZ (Bobeica et al., 2019).

4.2.2. FapD Expression and Purification

The expression and purification of various FapD constructs in P. sp. UK4, and homologs, was previously explored by Dr Hawthorne (Hawthorne, 2016). Notably, the expression of FapD was reported to be markedly improved when expressed into the cytoplasm, rather than the periplasm, and that is expected due to various cellular limitations associated with periplasmic-based expression (Hawthorne, 2016; Terpe, 2006). To express FapD into the cytoplasm, the gene sequence encoding FapD\textsubscript{39-246} (FapDWT) with an N-terminal his\textsubscript{6}tag (MGHHHHHHG) was synthesised to
exclude the N-terminal periplasmic secretion signal (Noinaj et al., 2016). The synthesised gene was cloned into a pET-28a(-TEV) vector (a T7-based expression vector). Of note, the cloning was performed by Genscript using a restriction enzyme-based approach. The plasmid was subsequently transformed into *E. coli* BL21 (DE3) for protein overexpression as previously described. Schematics of the constructs used in this chapter are shown within Figure 43.

Expression trials for FapDWT were conducted by examining its expression into the soluble and insoluble fractions under four conditions: 18 °C for 14 hours, 25 °C for 14 hours, 30 °C for 14 hours, and 37 °C for 4 hours. Analysis by SDS-PAGE indicates that FapDWT (theoretical MW is 24.5 kDa) did not display any obvious overexpression into the soluble or insoluble fractions over all conditions tested (Figure 44A). A western blot was subsequently employed to clarify the SDS-PAGE results and discern the optimal expression conditions for FapD. Analysis by western blot indicates that FapDWT is expressed in both the soluble and insoluble fractions over all conditions tested (Figure 44B). The highest level of soluble expression resulted from a 14 hour incubation at 18 and 25 °C; however, the ratio of soluble to insoluble FapDWT was lower at higher temperatures. Thus, the former condition was taken forward for expression in large volumes.
Figure 44: Expression trials of FapDWT. (A) SDS-PAGE gel analysis of soluble “S” and insoluble “I” fractions from expression trials across a range of temperature conditions (18 °C, 25 °C, 30 °C, 37 °C) indicate the relative quantity of FapDWT expressed for each condition (corresponding gel band is indicated by the red arrow) in comparison to uninduced samples (PrS and PrI correspond to uninduced soluble and insoluble fractions, respectively). However, due to low levels of FapDWT expression, the optimal expression conditions could not be deduced solely by SDS-PAGE analysis. (B) Western blot analysis clarifies the optimal expression conditions for FapDWT. The western blot fractions correspond to those indicated for the SDS-PAGE analysis in (A), and are aligned accordingly, reveal that soluble expression for FapDWT was greatest at 25 °C.

The FapD WT protein was purified in a two-step procedure comprising Ni-IMAC and SEC. All the buffers utilised between the cell lysis and final SEC purification incorporated 5 % (v/v) glycerol. The buffer was supplemented with glycerol in order to maximise FapDWT stability; glycerol concentrations between 5 to 10 % have previous been used to purify LahT147, and it was previously demonstrated to improve the stability of a FapD PA7 homolog (Bobeica et al., 2019; Hawthorne, 2016). Furthermore, 2 mM DTT was also added to the elution fraction of FapDWT after Ni-IMAC, and that was also added to the SEC buffer, for the reasons described above (Hawthorne, 2016). In order to maximise sample purity, three wash steps were incorporated into the Ni-IMAC protocol, including wash buffers that contain imidazole at concentrations of 20 mM, 30 mM, and 40 mM. The final eluted product of the first Ni-IMAC purification step produced a somewhat heterogenous mixture that contains FapD, as identified by SDS-PAGE analysis (Figure 45A). To maximise the sample purity and remove any higher oligomeric species that are present, the Ni-IMAC elution fraction was subjected to SEC. The SEC UV trace is displayed in Figure 45B. SDS-PAGE analysis indicates that peaks between 45 ml and 70 ml of the SEC UV trace contain impurities, whereas, the peak eluting between 70 and 80 ml is comprised of a ~25 kDa protein that corresponds to FapDWT (Figure 45C) (later
confirmed by western blot analysis (Figure 46B). By pooling the fractions that eluted between 74 and 78 ml, approximately 0.5 mg of purified FapDWT is obtained from 1 L of culture grown in LB media.

4.2.3. FapDWT is Proteolytically Unstable

The C39 family of cysteine peptidases recognise and cleave leader peptides at GG-motifs (Dirix et al., 2004; Lin et al., 2015). Though FapD WT does not contain a canonical GG-motif, it possesses multiple GG-like motifs. Therefore, FapD may recognise and cleave at one or more of these sites to undergo autoproteolysis (Rouse et al., 2017; Turk et al., 2012). If so, this could lead to sample heterogeneity that can be problematic for downstream biophysical analysis (Acton et al., 2011).

To monitor the stability of FapDWT, a small-scale stability assay was performed in which purified FapDWT (at 5 μM and 25 μM) was incubated at 20 °C, and soluble fractions were collected at days 0, 1, 3, and 5 for SDS-PAGE and western blot (anti-his6tag) analysis. SDS-PAGE analysis indicates that at day 0, a single band is evident at approximately 24 kDa that corresponds to FapDWT (Figure 46A). Over the five-day incubation period, both the 5 μM sample and the 25 μM sample undergo near complete degradation into a single species that is approximately 2-3 kDa smaller (~22-23 kDa) than the starting MW of the FapDWT (24.5 kDa). SDS-PAGE analysis also indicates that the rate of degradation is similar between both samples. The western blot was performed in order to determine whether the N-terminus of
FapDWT was degraded during the five-day incubation period. Of note, the western blot was performed and imaged by Grace Wu. The image was modified and shown with permission in Figure 46B (unpublished, image provided by Grace Wu, Imperial College London). Western blot analysis indicates the presence of an intense band at day 0, corresponding to FapDWT. By day 5, the band intensity had significantly decreased with regards to the 5 µM sample, and is completely lost for the 25 µM sample. Western blot analysis indicates that the rate of sample degradation is slightly higher for the 25 µM sample than the 5 µM sample, with a significant decrease in band intensity between days 1 and 3, unlike the 5 µM sample that exhibited a subtle decrease in band intensity for this incubation period. This suggests that the rate of degradation is concentration dependent. The postulated cleavage site of FapD, which corresponds to a GG-like motif at the FapDWT N-terminus, is shown in Figure 46C.

![Figure 46: FapDWT undergoes a specific cleavage event. (A) SDS-PAGE analysis indicates the degradation of FapDWT at concentrations of 5 µM and 25 µM (24.5 kDa) to a lower molecular weight species (21-22 kDa) over a five-day period. (B) Western blot analysis (performed using an anti-his6tag antibody) of the fractions from (A), which are aligned accordingly. The results indicate that the FapDWT N-terminus is degraded over a five-day period. Grace Wu performed and imaged the western blot. The permission of Grace Wu was granted to display the resulting (modified) image in this thesis (unpublished). (C) A schematic of FapD indicates the postulated cleavage site (shown as a red dashed line); the cleavage site is predicted between FapD_{39} (Gly) and FapD_{50} (Ala). The mature protein chain of FapD is shown in purple, the N-terminal fusion tag is labelled “NTT” and coloured white, and the sequence of FapD_{39,53} is shown below (coloured purple) for reference.](image)
After 6 days of incubation, the 5 µM sample was concentrated to ~80 µM to determine whether the FapDWT degradation product is stable at high sample concentrations. SDS-PAGE analysis of soluble fractions collected over a 60 minute incubation period indicates that the intensity of a single band at ~22kDa decreased over time (Figure 47, fractions 2-4), highlighting the precipitation event; the precipitate corresponds to the FapDWT degradation product, as well as sample impurities or additional FapDWT breakdown products (Figure 47, fraction 6).

![Figure 47: FapDWT stability in the presence and absence of CCExt.](image) SDS-PAGE gel analysis indicates the stability of FapDWT (alone) and FapDWT in complex with CCExt. Lane 1 corresponds to the co-lysis assay SEC fraction at 60 ml (later presented in Figure 50C). Lanes 2, 3, and 4 correspond to the purified FapDWT sample (presented in Figure 46A) that was concentrated to ~80 µM (after 6 days of incubation at 20 °C) and incubated at 20 °C for 0, 15, and 60 minutes, respectively. The soluble fractions are presented. Lane 5 corresponds to co-lysis assay SEC fractions at 60 and 62 ml (later performed and presented in Figure 50C) that were pooled and incubated at 20 °C for 3 days. The resulting soluble fraction is shown. Lane 6 corresponds to the precipitate formed after incubation of FapDWT at 80 µM for 60 minutes (as described within the text).

These results indicate that the FapDWT specifically degrades into a single species approximately 2-3 kDa smaller than the FapDWT over a five-day incubation period at 20 °C. Furthermore, the FapDWT N-terminus undergoes near complete degradation over this incubation period. Whether the C-terminus is also degraded remains to be determined. The MW of the FapDWT degradation product approximates to the MW of FapD_{50-246}, assuming that only the FapDWT N-terminus is degraded. This suggests that the specific cleavage of FapD may be due to autoproteolysis of the GG-like motif at the FapDWT N-terminus, as described above. The methods that can be used to accurately determine the proteolytic activity of FapD, and the FapD cleavage site, are later discussed.
4.2.4. FapD Possesses a Folded Domain

To determine whether FapDWT (incubation day 0) possesses a folded domain, 1D $^1$H-NMR spectroscopy was employed. The 1D $^1$H-NMR spectrum for FapDWT reveals the presence of peaks downfield of 8.5 ppm, as well as peaks upfield of 0 ppm, indicating the presence of a folded domain (Figure 48A) (Kwan et al., 2011). In addition, broad line widths are observed that could be a combination of the high MW of FapDWT (24.5 kDa), as well as the high sample viscosity that results from the presence of 5 % (v/v) glycerol in the sample buffer. Both factors contribute to poor relaxational properties (Pfuhl and Driscoll, 2000). Of note, SDS-PAGE analysis was used to determine the stability of FapDWT before and after the NMR experiment, indicating that FapDWT did not degrade during the recording of this experiment (Figure 48B).

![Figure 48: 1D $^1$H-NMR spectrum of FapDWT. (A) The 1D $^1$H-NMR spectrum for FapDWT is shown. The peaks downfield of 8.5 ppm and upfield of 0 ppm indicate the presence of a tertiary structured domain. Sample conditions: 44 μM, 298 K, 512 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe. (B) SDS-PAGE gel analysis of FapDWT before (Pre) and after (Post) the 1D $^1$H-NMR experiment indicates that FapDWT remained stable throughout the entirety of the experiment.](image-url)
To determine the secondary structure composition of FapDWT, far-UV CD spectroscopy was employed. Qualitative analysis of the CD profile indicates the presence of a strong negative CD absorbance at 208 nm, and a weaker negative CD absorbance at 222 nm, which indicates the presence of α-helical content within the FapDWT (Figure 49A). In addition, the CD absorbance increases between 215 and 225 nm, hinting to the presence of β-sheet structure and/or disorder (Greenfield, 2006b). The BeStSel webserver was used to deconvolute and quantify the secondary structure composition of the CD spectrum (Micsonai et al., 2018). The secondary structure of FapDWT is comprised of 39.7 % “others”, 30.9 % α-helix, 15.1 % β-strand, and 14.2 % turn (Figure 49B). This is in broad agreement with the secondary structure prediction by PSI-PRED, for which 45 % of residues predicted disordered, 32 % are predicted α-helical, and 23 % are predicted β-stranded (Jones, 1999). The secondary structure composition of LahT147 (according to PDB ID: 6MPZ) is 29 % disorder, 39 % α-helix, and 32 % β-strand; thus, the secondary structure composition of FapDWT broadly encompasses that of the mixed secondary structure composition of a C39-peptidase domain (Bobeica et al., 2019).

**Figure 49**: Far-UV CD spectrum of FapDWT. (A) The CD spectrum of FapDWT (0.12 mg/ml) indicates strong (negative) CD absorbances at 208 and 222 nm (alpha-helical content), along with hints towards an increased (positive) CD absorbance at 218 nm (beta-sheet). (B) The secondary structure composition for FapDWT was predicted based on the CD profile in (A) by the BeStSel server, and the results are depicted as a pie chart.
4.2.5. FapDWT and CCExt Associate

A recent study utilised native MS to demonstrate that FapD and FapF associate (Chorev et al., 2020). In this study, FapD was probed for interactions with CCExt to determine whether the FapF N-terminus, alone, is sufficient for the interaction between FapD and FapF to occur. If an interaction is discovered, the next steps of this study would be aimed towards solving the structure of the complex using X-ray crystallography or NMR spectroscopy, in order to elucidate its function.

In this experiment, termed the “co-lysis assay” or “CLA”, FapD and CCExt were expressed independently in LB under optimised conditions that were previously described. Two litres of cell biomass containing overexpressed CCExt, and 4 L of cell biomass containing overexpressed FapD, were each resuspended in buffer containing 5% (v/v) glycerol. After resuspension, both resuspensions were combined by mixing (vortexed), and were subsequently lysed and purified by Ni-IMAC, as previously described for the FapDWT. The protocol and buffers utilised for this experiment are identical to those used for purifying FapDWT. SDS-PAGE analysis indicates that the Ni-IMAC eluent contains both a ~24 kDa species corresponding to FapDWT, and a ~10 kDa species corresponding to the CCExt (Figure 50A). The elution fraction was loaded onto SEC to probe whether FapDWT and CCExt co-elute (associate). Of note, CCExt was independently purified and subjected to SEC, using a protocol that incorporated identical buffers and an identical experimental set-up to that used for the FapDWT; the CCExt SEC profile and its corresponding SDS-PAGE analysis can be found in Figure 50B and 50E, respectively. In effect, this enables for a fair comparison of the SEC UV traces of CCExt, FapDWT, and the CLA, which are shown overlaid in Figure 50B. SDS-PAGE analysis indicates that the peaks between 40 and 60 ml of the CLA correspond to impurities carried from the Ni-IMAC step (Figure 50C). Analysis of the fractions corresponding to the asymmetric peak indicates the presence of FapDWT and CCExt (as well as low abundance impurities) between 60 and 70 ml, demonstrating their co-elution; the presence of FapDWT and CCExt in these fractions was confirmed by western blotting (Figure 50D). SDS-PAGE analysis of the shoulder peak at 72 ml indicates the presence of FapDWT and impurities.
Figure 50: Co-lysis of FapD and CCExt. (A) SDS-PAGE gel analysis of the Ni-IMAC after the co-lysis and co-Ni-IMAC of CCExt and FapDWT. The Ni-IMAC flow through (FT), wash steps (30; 40; the number corresponds to wash buffer imidazole concentration (mM)), and elution (“E”), are shown (the red and green arrow indicate the bands corresponding to FapDWT and CCExt, respectively). The FapDWT and CCExt are successfully eluted, albeit, at a low purity. (B) SEC (UV 280 nm) trace of FapDWT and CCExt (green line) after application of the Ni-IMAC elution fraction in (A) to a SEC column, and that is overlaid with the SEC profiles for CCExt (purple line; refer to (E) for the corresponding SDS-PAGE analysis) and FapDWT (black line) that were purified independently. (C) SDS-PAGE gel analysis of the specific SEC fractions collected for co-purified FapDWT and CCExt, indicating that both proteins co-elute between 60 to 70 ml. The red arrow identifies the band corresponding to FapDWT. (D) Western blot analysis (performed using an anti-his6tag antibody) clarifies the presence and position of FapDWT on the SDS-PAGE gel in (C). The marker (M) units are in kDa. (E) SDS-PAGE gel analysis of the specific SEC fractions collected for CCExt (SEC UV profile illustrated in (B)), indicating that CCExt is eluted in high purity between 70 to 84 ml.

The SEC peak that corresponds to the co-elution of FapDWT and CCExt elutes approximately 10 ml earlier than the peaks corresponding to the independent purifications for each protein. The co-elution and the earlier elution for both the FapDWT and the CCExt indicates that both proteins have associated to form a larger complex (possessing a higher Stokes radius) (Hong et al., 2012). The lack of a second peak on the SEC UV trace for CLA, which corresponds to the SEC retention volumes for independently purified FapDWT or CCExt, suggests that a transient interaction is formed between FapDWT and CCExt (Hong et al., 2012; Hu et al., 2012). Lastly, the stability of the FapDWT-CCExt complex was examined as previously described for FapDWT (alone). The complex was concentrated to 25 µM and incubated at 20 °C for
3 days; a soluble fraction was subsequently collected for SDS-PAGE analysis. SDS-PAGE analysis reveals that after 3 days of incubation, FapDWT degrades into a second species that is consistent in size with that previously described for the degradation product of FapDWT; this suggests that the binding of CCExt to FapDWT does not stabilise (proteolytically) FapDWT (Figure 47, fractions 1 and 5).

4.2.6. Expression and Purification of an N-terminally Truncated FapDWT (FapD$_{50-246}$)

The results of the FapDWT stability assay demonstrates that the FapDWT N-terminus is proteolytically unstable, with evidence to indicate the formation of a breakdown product 2 to 3 kDa smaller than the FapDWT over a five-day sample incubation period at 20 °C. The MW of the breakdown product corresponds to the loss of the N-terminal his6tag and potentially ~FapD$_{39-49}$ that terminates after a conserved GG-like motif (FapD$_{39-50}$), which would theoretically produce a 22.3 kDa product (2.2 kDa smaller than FapDWT). Therefore, the breakdown of FapDWT may correspond to an autoproteolytic event. This results in sample heterogeneity that is problematic for downstream biophysical analysis (Acton et al., 2011). Furthermore, the function of the FapD N-terminus, which shows no homology to C39-peptidase domain LahT147, is currently unknown (Bobeica et al., 2019). Therefore, its influence on the folding and stability of FapD, as well as its role in the FapD-CCExt interaction, can be investigated by studying the N-terminally truncated form of FapD. In order to recombinantly express the N-terminally truncated FapDWT that lacks the FapD$_{39-50}$ GG-like motif, FapD$_{50-246}$ was sub-cloned from the FapDWT construct and into the pET-28a(-TEV) vector using the In-Fusion design protocol.

The FapD$_{50-246}$ construct, referred to as “FapDT”, was expressed and purified according to the protocol described for FapDWT. SDS-PAGE analysis of the Ni-IMAC step indicates the presence of FapDT (MW is 23.5 kDa) as well as an abundance of impurities in the eluent (Figure 51A). The eluent was subjected to SEC as previously described. The SEC UV absorbance trace for FapDT (Figure 51B), together with SDS-PAGE analysis (Figure 51C), indicates the presence of peaks between 40 and 70 ml that contain impurities of various MWs. The peak of relatively high absorbance between 70 and 80 ml contains FapDT. SEC elution volumes between 74 to 80 ml contain FapDT at high purity (the presence of FapDT is later
confirmed by western blotting (Figure 54B)). Approximately 2 mg of purified FapDT is obtained from 1 L of culture grown in LB media, which is four-fold greater than FapDWT. This suggests that removal of FapD39-49 from FapDWT increases the stability of FapD in vitro.

**Figure 51: The purification of FapDT.** (A) SDS-PAGE gel analysis of an example Ni-IMAC purification for FapDT. The Ni-IMAC flow through (FT), wash steps (20; 30; 40; the number corresponds to wash buffer imidazole concentration (mM)), and elution (“E”), are shown (the red arrow indicates the band corresponding to FapDT). Little or no FapDT is present in the FT fraction, and little or no FapDT is lost during the wash steps. The FapDT is successfully eluted, albeit, at a low purity. (B) SEC (UV 280 nm) trace of FapDT (blue line) after application of the FapDT Ni-IMAC elution fraction to a SEC column. The UV trace for FapDWT is also overlaid (black line). Of note, the SEC profiles for FapDWT and FapDT stem from the purification of 3 L and 4 L of cell biomass (cultured in LB), respectively. (C) SDS-PAGE gel analysis of the specific SEC fractions collected for FapDT, indicating that FapDT is eluted in high purity between 74 to 80 ml.

4.2.7. The Folding of FapD is Not Affected by Truncation of FapD39-49

To determine the effect(s) of truncating the FapD N-terminus (FapD39-49) on the folding and secondary structure composition of FapD, FapDT was subjected to analysis by 1D 1H-NMR spectroscopy and CD spectroscopy. The resulting spectra are compared to those of FapDWT. Interestingly, the 1D 1H-NMR spectra of both FapDWT and FapDT display near perfect overlap in their respective amide and
methyl regions, suggesting that the folded core of FapD is not disrupted by the N-terminal FapD truncation (Figure 52).

![1D 1H-NMR spectra of FapDT and FapDWT](image)

**Figure 52: 1D 1H-NMR spectra of FapDT and FapDWT.** The 1D 1H-NMR spectra for FapDWT (blue) and FapDT (red) are overlaid. FapDT sample conditions: 60 µM, 298 K, 512 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe.

The overlaid far-UV CD spectra indicate that the CD profiles for FapDT and FapDWT are near identical throughout the entire 200 to 260 nm wavelength range (Figure 53). Of note, the CD intensity of FapDT indicates a minor increase between 205 and 225 nm, and a more noticeable increase at 208 nm, compared to the WT. This indicates that the α-helical content is reduced for FapDT compared to the WT, suggesting that FapD_{39-49} possesses α-helical content, which is in agreement with the secondary structure prediction for this region (Greenfield, 2006b). Together with the 1D 1H-NMR spectrum, these results suggest that the global folding of FapDWT is not affected by the truncation of FapD_{39-49}.
4.2.8. FapDT is Proteolytically Stable

To monitor the proteolytic stability of FapDT, a small-scale stability assay was performed in which purified FapDT (at 25 µM) was incubated at 20 °C, and soluble fractions were collected at days 0 and 5 for SDS-PAGE and western blot (using an anti-his6tag antibody) analysis. Both SDS-PAGE and western blot analysis indicate that at day 0, a single band is evident at approximately 23 kDa that corresponds to FapDT (Figure 54A). Over a five-day incubation period, the MW and intensity of this band remains consistent. (Figure 54B). These results indicate that the truncation of the FapDWT N-terminus (FapF_{39-49}) increases the proteolytic stability of FapD.

Figure 53: Far-UV CD spectra of FapDT and FapDWT. The CD profiles for FapDWT (blue) and FapDT (0.18 mg/ml) (red) are overlaid.
Figure 54: FapDT is proteolytically stable. (A) SDS-PAGE gel analysis of soluble fractions from the FapDT stability assay indicates that FapDT does not undergo degradation or precipitation over a five-day period of incubation at 20 °C. (B) Western blot analysis (performed using an anti-his6tag antibody) of the soluble fractions collected at day 0 and 5 from the stability assay are shown, indicating that the FapDT N-terminus remains intact throughout the incubation period. Lane (“C”) corresponds to a control lane, containing the Ni-IMAC elution fraction for FapDT (presented in Figure 51A). (C) SDS-PAGE gel analysis of the precipitate obtained after concentrating FapDT to approximately 120 µM. The corresponding soluble (Sol) and insoluble (Ins) fractions are shown. (D) SDS-PAGE gel analysis indicates that purified FapDT (post Ni-IMAC) rapidly precipitates after the addition of buffer containing 0.5 % C8E4, compared to the same sample for which C8E4 was not added. The fractions displayed for each sample correspond to the soluble fractions collected over 10 minutes of incubation at 20 °C. The lane marked (I) corresponds to the precipitate that had formed after a 10 minute incubation at 20 °C. The red arrow indicates the band corresponding to FapDT.

To determine whether FapDT is stable at high concentrations, a sample of FapDT that was not subjected to the stability assay (i.e. incubation day 0) was purified and concentrated. At approximately 100 µM, the sample exhibited signs of rapid precipitation. This indicates that FapDT, like FapDWT, is prone to aggregation at high sample concentrations (soluble and insoluble fraction shown in Figure 54C). In an attempt to reduce the aggregation propensity of FapDT, 0.5 % (v/v) C8E4 (tetraethylene glycol monoocetyl ether) was added to the purified FapDT sample (post Ni-IMAC) with aim to shield any hydrophobic patches that may be present on the surface of FapDT, which may be driving the aggregation of FapD (le Maire et al., 2000; Otzen, 2011). However, as confirmed by SDS-PAGE analysis, FapDT instantly showed signs of visible precipitation at 20 °C (monitored over a 10 minute period) (Figure 54D). These results are in agreement with a previous study which demonstrated that FapD (Pseudomonas PA7) rapidly precipitates in the presence of buffer containing 0.5 % (v/v) LDAO (N,N-dimethyldodecylamine N-oxide) (Hawthorne, 2016).
4.2.9. FapDT and CCExt Associate

A transient interaction between FapDWT and CCExt during SEC was previously indicated in this study. To determine whether the removal of the FapD N-terminus interferes with the FapD-CCExt interaction, SEC was used to probe for an interaction between FapDT and CCExt (SEC binding assay). First, FapDT and CCExt were independently purified as described previously. Purified FapDT (1 ml, 60 µM) was mixed with purified CCExt (1 ml, 40 µM (trimer)) and subjected to SEC. Analysis of the SEC UV trace indicates the presence of two overlapping peaks (Figure 55A); the first higher absorbance peak is present between 62 to 72 ml, and the second lower absorbance peak elutes at 72 to 88 ml. SDS-PAGE analysis indicates the co-elution of FapDT and CCExt between 64 and 74 ml (Figure 55B) (their presence is confirmed by western blotting (Figure 55C)). Both FapDT and CCExt co-elute approximately 8 to 10 ml earlier than when independently subjected to SEC. This indicates that FapDT and CCExt associate (Hong et al., 2012; Hu et al., 2012). Interestingly, SDS-PAGE analysis indicates that the peak at 76 to 82 ml corresponds to FapDT. The absence of CCExt in this fraction suggests that the molar ratio of FapD:CCExt is in excess. Therefore, this implies that the molar ratio of binding for FapDT to CCExt is below 1.5:1, however, further evidence is required to accurately determine the binding stoichiometry.
Figure 55: Co-purification of FapDT and CCExt. (A) Purified FapDT was combined with purified CCExt at a molar ratio of 1.5:1 and applied to a SEC column. The resulting SEC profile (UV 280 nm) is indicated (green line) and overlaid with the SEC profiles for CCExt and FapDT that were independently purified. (B) SDS-PAGE gel analysis of the specific SEC fractions collected for co-purified FapDT and CCExt indicates that both proteins co-elute between 62 to 74 ml. The purified FapDT (“D”) and CCExt (“CCE”) samples that were independently purified prior to co-SEC are indicated. (C) Western blot analysis confirms the presence of FapDT and CCExt on the SDS-PAGE gel in (B). Lane (“C”) corresponds to a control lane, containing the Ni-IMAC elution fraction for CCExt (shown in Figure 37A).

To monitor the stability of the DTCE complex, a small-scale stability assay was performed in which the SEC co-eluent (SEC fraction at 64 ml) was concentrated to 25 µM and incubated at 20 °C for three days. Soluble fractions were collected at days 0 and 3 for SDS-PAGE analysis. SDS-PAGE analysis indicates that at day 0, a ~23 kDa band is evident that corresponds to FapDT, and a ~10 kDa band is present that corresponds to CCExt (Figure 56). After three days of sample incubation, there were no changes in band intensity, and no new bands had formed, indicating that the DTCE complex is stable under these experimental conditions. Furthermore, to determine the stability of the DTCE complex at higher concentrations, the SEC co-eluent (SEC fraction at 64 ml; sample was not subjected to incubation at 20 °C) was concentrated to 120 µM using a 10 kDa MWCO concentrator. The sample instantly showed signs of visible precipitation. Interestingly, SDS-PAGE analysis of the insoluble fraction...
indicates the presence of a higher proportion of FapDT compared to CCExt in the sample (dissimilar to the soluble fractions), suggesting that FapDT, rather than CCExt, is the main species undergoing aggregation (Figure 56). The presence of CCExt in this fraction suggests that a smaller proportion of the DTCE is also undergoing aggregation.

**Figure 56: The FapDT-CCExt complex is proteolytically stable.** SDS-PAGE gel analysis indicates that FapDT and CCExt (~25 µM) do not undergo degradation or precipitation over a three-day period of incubation at 20 °C. The FapDT-CCExt complex was concentrated to 120 µM, and the sample underwent rapid precipitation, indicated in lane (I) (insoluble fraction shown). Of note, the intensity of the bands on this gel is sub-maximum due to insufficient staining.

### 4.2.10. Thermodynamic Binding Parameters of the FapDT-CCExt Interaction

Isothermal titration calorimetry (ITC) was used to quantify the binding parameters of the interaction between FapDT and CCExt. Biomolecular interactions release or absorb heat energy; this heat change is directly measured by ITC during the titration of a ligand (protein A) into a cell containing a sample of interest (protein B), which enables the determination of the binding parameters for a protein-protein interaction (Leavitt and Freire, 2001; Velazquez-Campoy et al., 2004).

In this experiment, purified FapDT was concentrated to 25 µM and loaded into the ITC cell (300 µl, equilibrated at 298 K), and the ITC syringe (40 µl) was loaded with 500 µM of purified CCExt (trimer), both in identical buffer containing 200 mM NaCl, 50 mM HEPES pH 7.5, 1 mM TCEP, and 5 % (v/v) glycerol. The resulting binding isotherm is shown in Figure 57 (top panel), demonstrating a direct interaction of CCExt with FapDT. Each titration step displays a negative heat change that indicates an exothermic process. The heat change decreased as the injections progressed, which indicates a decrease in the number of available FapDT binding sites as the concentration of CCExt (in the cell) increases; this becomes increasingly evident after a ~1:1 molar ratio is achieved. Of note, the presence of a positive heat change between titration 10 and 11 is likely due to the introduction of an air bubble into the
Additionally, the baseline noise intensity increased as the experiment progressed, which is likely due to the formation (and/or injection) of additional air bubbles into the ITC cell (Salim and Feig, 2009; Velazquez-Campoy et al., 2004).

Figure 57: Isothermal titration calorimetry (ITC) on FapDT and CCExt. A 500 µM sample of purified CCExt was titrated into a 25 µM sample of purified FapDT. The top panel displays the resulting raw data (plotted as heat rate versus time) from injecting CCExt into FapDT over forty successive injections. The bottom panel displays the binding curve obtained from integration of the heat rate over time. Data for the thermodynamic parameters was obtained through use of a one-site binding model. Of note, injection number 1, 21, 35, and 40 were discarded from calculations due to their anomalous nature.

Integration of the peaks corresponding to each injection in the top panel of Figure 57 are plotted against the FapDT:CCExt molar ratio, and the binding isotherm resulting is shown in the bottom panel of Figure 57. The isotherm was fitted using a nonlinear least-squares algorithm with a one-site binding model. Peak 1 (corresponds to a smaller volume trial injection), 21, 35, and 40 were excluded to improve the fitting accuracy (Mizoue and Tellinghuisen, 2004). The sigmoidal titration curve yields a number of parameters, which includes the binding enthalpy ($\Delta H$), the
stoichiometry (N), and the association constant (\( K_a \)) as described in section 2.9.4. (Callies and Daranas, 2016; Velazquez-Campoy et al., 2004). The thermodynamic parameters are displayed in the bottom panel of Figure 57; of note, a complete sigmoid titration curve was not attained by use of the current experimental set-up, therefore, the thermodynamic parameters obtained here are limited in their accuracy and provide only a broad indication of the nature of the binding interaction. Suggestions to improve the quality of ITC data are later detailed (section 4.3.2). The dissociation constant of the interaction of FapDT and CCExt is in the low-micromolar range (~3 \( \mu \)M), which corresponds to a transient interaction; this result is in agreement with the previous results of the SEC binding assay for the DTCE complex (Acuner Ozbabacan et al., 2011). The enthalpy of binding was calculated to be -15.01 kcal mol\(^{-1}\), which indicates a favourable enthalpic (exothermic) process. The \(-T\Delta S\) is -7.39 kcal mol\(^{-1}\), which indicates an unfavourable entropic contribution to the binding process. The resulting Gibbs free energy is -7.62 kcal mol\(^{-1}\). Together, these results indicate that the binding process is favourable at 298 K, and that is driven by enthalpy-entropy compensation, with the negative enthalpy providing the main contributions to the binding energy (Chodera and Mobley, 2013; Gallicchio et al., 1998). Albeit further evidence is required, this hints that the binding process is contributed mainly by the formation of non-covalent bonds such as polar bonds, van der Waals forces, and electrostatic interactions, rather than hydrophobic forces (Du et al., 2016; Leavitt and Freire, 2001). The number of binding sites (N or N-value) calculated is approximately 0.5; however, the accuracy of this figure is unlikely to be accurate, as later discussed (Tellinghuisen and Chodera, 2011). These results also suggest that the FapD N-terminus (FapD\(_{39-49}\)) is not solely essential for binding to CCExt, however, further investigation is required to confirm this.

### 4.2.11. FapD C76S is Unstable

In an attempt to determine whether the predicted C39-active site of FapD is essential for the interaction with CCExt, a FapDT C76S mutant was constructed. Site-directed mutagenesis was performed to substitute Cys76 to Ser within the FapDT construct. Serine was selected as the replacement for cysteine as the size and hydrophilicity of Ser is similar to Cys (Taylor, 1986). Furthermore, the substitution of
the C39-peptidase active site Cys to Ser has previously been demonstrated to abolish its proteolytic activity (Furgerson et al., 2008).

FapDT C76S (referred to as “C76S”), was expressed and purified according to the protocol described for FapDWT. SDS-PAGE analysis of the Ni-IMAC step indicates the presence of C76S (MW is 23.5 kDa) as well as an abundance of impurities in the eluent (Figure 58A). The eluent was subjected to SEC as previously described. The SEC UV absorbance trace for C76S, together with SDS-PAGE analysis, indicates the presence of peaks between 40 and 70 ml that contain impurities of various MWs (Figure 58B and C). The peak region 74 to 78 ml corresponds to C76S, as well as higher and lower MW impurities. The presence of C76S is confirmed by western blotting (Figure 58D). The SEC UV traces of FapDT and FapDWT are overlaid with that of C76S, indicating that the yield of C76S is drastically reduced compared to FapDT. This indicates that the C76S mutant destabilises FapDT, which is likely due to the disruption of its folded core.

Figure 58: Purification of FapDT C76S. (A) SDS-PAGE gel analysis of a Ni-IMAC purification for FapDT C76S. The Ni-IMAC flow through (FT), wash steps (30; 40; the number corresponds to wash buffer imidazole concentration (mM)), and elution (E), are shown (the red arrow indicates the band corresponding to FapDT C76S). Little or no FapDT C76S is present in the FT fraction, and little or no FapDT C76S is lost during the wash steps. The FapDT C76S is successfully eluted, albeit, at a low purity. (B) SEC (UV 280 nm) trace of FapDT C76S (green) after application of the FapDT C76S Ni-IMAC elution fraction to a SEC column. The UV trace for FapDT (blue) and FapDWT is also shown overlaid (black). Of note, the SEC profiles for FapD C76S, FapDT and FapDWT stem from the purification of 4 L, 4 L, and 3 L of cell biomass (cultured in LB), respectively. (C) SDS-PAGE gel analysis of specific fractions collected after FapDT C76S SEC indicates that the peak at 74 ml contains FapD C76S. FapDT C76S is eluted in low purity between 74 to 78 ml. (D) Western blot analysis (performed using an anti-his6tag antibody) clarifies the presence and position of FapDT C76S on the SDS-PAGE gel in (C). Lane (“C”) corresponds to a control lane that contains the Ni-IMAC elution fraction for FapDT (presented in Figure 51A).
4.3. Summary and Discussion

4.3.1. Is FapD a Proteolytically Active C39 Peptidase?

In this study, FapDWT (UK4) was expressed, purified, and confirmed to be folded. Prior to further biophysical analysis, the stability of FapWT was examined. As a postulated C39-peptidase that possesses GG-like motifs, FapDWT autoproteolysis was considered a possibility (Khan and James, 1993). Interestingly, the FapDWT stability assay indicated the specific degradation of FapDWT from a 24.5 kDa species to a ~22-23 kDa species over a five-day incubation period at 20 °C. The FapDWT N-terminus was degraded, however, it is unclear whether the C-terminus remained intact. The conversion from a single higher MW species to a single lower MW species suggests that a specific cleavage event occurs, and the 2-3 kDa MW difference between the two species approximately corresponds to the degradation of FapDWT up to the predicted cleavage site of the FapD GG-like motif (FapD39-49). The greater proteolytic stability of FapD50-246 compared to FapDWT agrees with this hypothesis. Of note, the association of CCExt to FapD did not improve the proteolytic stability of FapD. These results suggest that FapDWT is able to undergo autoproteolysis, which is potentially occurring at the predicted cleavage site corresponding to the GG-like motif described above. Autoproteolysis is not an unusual event for the cysteine proteases of the papain superfamily. For example, cathepsin B and cathepsin L are two members of the papain superfamily that are synthesised as inactive zymogens. The cathepsin B and L zymogens are maintained under neutral pH during their transportation to endosomes and lysosomes. Upon entering the acidic environment of these cellular compartments, the inhibitory prosegments located at their N-termini are autocatalytically processed to result in their activation. Therefore, FapD39-49 may also represent an inhibitory prosegment, and under specific cellular conditions, FapDWT may undergo autocatalytic cleavage at this site in order to activate its proteolytic activity (Khan and James, 1993; Turk et al., 2012; Verma et al., 2016).

Further research is required to elucidate whether FapD possesses specific catalytic activity akin to the C39-peptidase family, as well as determine whether it is able to recognise and cleave the GG-motif or the GG-like motifs of the Fap system. As a previous study has demonstrated, a high performance liquid chromatography (HPLC)-MS configuration can be employed to study the proteolytic activity of FapD towards peptides containing the GG-motif or the GG-like motifs that are scattered across the
Fap proteins (Bobeica et al., 2019). These methods would enable the specific substrate cleavage sites (if any) to be determined, which would also elucidate whether FapD undergoes autoproteolysis (Bobeica et al., 2019). In addition, the proteolytic activity of FapD towards the reported substrates for LahT147 and PCAT1 can also be investigated to determine whether the substrate binding site and cleavage activity of FapD is similar to those of the canonical C39-peptidase domains (Bobeica et al., 2019; Kieuvongngam et al., 2020). If the peptidase activity and substrate specificity are found to be similar between FapD and the PEP domains, the studies performed for FapD may provide insight that is directly translatable to the studies of the C39-peptidase family and the PCATs (and vice-versa), as the C39-peptidase family share a similar structure and function (Bobeica et al., 2019; Ishii et al., 2010; Kieuvongngam et al., 2020; Lecher et al., 2012). Furthermore, various FapD mutants can also be constructed to evaluate the importance, if any, of the predicted C39-peptidase active site residues; this can include substitution of C76 and H160 to alternative residues (Barrett and Rawlings, 1996; Furgerson et al., 2008). Importantly, the proteolytic activity of FapD should also be studied in complex with FapF, as the presence of a transporter in the case of PCAT1-PEP is essential for the catalytic activity of its respective PEP domains; thus, the catalytic activity of FapD may be dependent on its association to FapF (Lin et al., 2015). These assays can be expanded to query the activity of FapD under combinations of various temperatures, buffer pHs, and reducing agents (Lee et al., 2012; Verma et al., 2016). For instance, FapD activity may be sensitive to temperature, as nearly 40% of Pseudomonas fluorescens genes are reported to be thermoregulated (Hoštacká et al., 2010; Regeard et al., 2000). On the other hand, if the proteolytic activity of FapD to substrates containing the GG-motifs or GG-like motifs is shown to be nonexistent, the binding of the above substrates to FapD should be probed, using methods such as ITC, to elucidate whether FapD is able to transiently interact with Fap substrates (Kotake et al., 2008).

Solving the structure of FapD has the potential to significantly advance our understanding of its function, as this information would provide insight into whether it possesses the canonical C39-peptidase fold, as well as indicate the presence of potential binding sites, or other unknown functions. However, attempts to solve the structure of FapD by NMR spectroscopy and X-ray crystallography is hindered by the poor solubility of FapD at high sample concentrations (>80 µM). The aggregation propensity of FapD at relatively high concentrations does not appear to be improved
by removal of the FapD N-terminus (FapD$_{39-49}$), or the formation of a complex with the CCExt. Various buffer conditions have previously been used to achieve C39-peptidase domain concentrations of 5-10 mg/ml (Bobeica et al., 2019; Ishii et al., 2010). Optimal buffer conditions differ from protein to protein, thus, various buffer additives should be screened to investigate their effect on the stability of FapD. Of note, this can be performed in high-throughput using differential scanning fluorimetry (DSF) (Senisterra and Finerty, 2009; Vedadi et al., 2006). A list of buffer additives that are reported to promote the solubility of proteins, and their recommended concentration ranges, is summarised in Bondos and Bicknell (2003). Also, the removal of the his6tag may improve the stability of FapD; a TEV-cleavable site can be introduced between FapD and the his6tag for TEV protease digestion (Booth et al., 2018). Alternatively, it may be possible to solve the structure of FapD in complex with FapF using cryo-EM; this strategy is detailed within the next section (4.3.2).

### 4.3.2. Interactions Between FapD and FapF

Within this study, both FapDWT and FapDT were demonstrated to interact with CCExt in vitro. The co-elution of both FapDWT/FapDT and CCExt during SEC first demonstrated their ability to associate, and the formation of a transient interaction was confirmed by measurement of the binding parameters using ITC. The results from this study indicate that the FapD N-terminus (FapD$_{39-49}$) is not essential for binding to FapF, however, its exact contributions to the binding process (in vitro) were not determined. The function of the FapD N-terminus remains to be elucidated.

The ITC results suggest that the FapDT-CCExt interaction is characterised mainly by the formation of non-covalent bonds such as polar contacts, electrostatic interactions, and van der Waals forces, however, the contributions from hydrophobic forces for this interaction might be under represented due to the presence of 5 % (v/v) glycerol within the ITC buffer (Du et al., 2016; Leavitt and Freire, 2001; Vagenende et al., 2009). Glycerol has been reported to interact with the hydrophobic surface regions of proteins, and therefore, it may reduce the strength of hydrophobic forces that are involved in protein-protein interactions (Vagenende et al., 2009). Provided the stability of FapD is not compromised, future ITC experiments should aim to reduce or abolish the glycerol content within the ITC buffer to more accurately quantify the binding parameters for this interaction. Nonetheless, the interaction
between FapDT and CCExt occurs in the presence of 5 % (v/v) glycerol, implying that the presence of non-covalent interactions are sufficient for this interaction to occur. This suggests that CCExt may not bind to the predicted FapD substrate binding site/active site that is predominantly hydrophobic; therefore, the predicted FapD active site may remain accessible for binding to substrates via their GG- or GG-like motifs, akin to the PEP domain of PCAT1 (Kieuvongngam et al., 2020). Furthermore, the results of the SEC co-elution assay for the DTCE complex indicate that the interaction stoichiometry (FapDT:CCExt) is <1.5:1. A recently published report that indicates a 1:1 stoichiometry between FapF and FapD (under native conditions), which is unusual considering the presence of three, potentially open, β-barrels within full length FapF (Chorev et al., 2020; Rouse et al., 2017). Despite the knowledge that FapF is trimeric, the question still remains as to why FapF is trimeric (what is the functional importance of a FapF trimer?), and whether FapD, or other Fap proteins, possess a role in modulating the pore-open versus pore-closed configuration of FapF. Whether the FapF plugs are dynamically independent of each other also remains to be elucidated (Rouse et al., 2017, 2018b).

Of note, the accuracy of the ITC-derived binding parameters in this study is limited due to the lack of a complete sigmoidal titration curve, as well as the inaccuracies associated with protein concentration measurements (Tellinghuisen and Chodera, 2011; Velazquez-Campoy et al., 2004). Regarding the former, the high molar ratio of CCExt:FapDT (20:1) used in this experiment resulted in the early saturation of FapDT binding sites, thus, future ITC studies for FapDT-CCExt should aim to reduce the difference in molar ratio (Velazquez-Campoy et al., 2004). Additionally, one study suggests that experimental precision can be improved by using fewer titrations (injections) (Tellinghuisen and Chodera, 2011). Lastly, future ITC studies for this interaction should increase the injection volume in order to increase the experimental signal-to-noise ratio, and therefore, improve the accuracy of the binding parameters (Velazquez-Campoy et al., 2004).

Solving the structure of full length FapF in complex with FapD would provide substantial insight into the FapD-FapF interaction, as well as Fap biogenesis. However, the structural study of full length FapF by NMR or X-ray crystallography has previously proven difficult due to its high MW and its disordered linker domain (Hawthorne, 2016; Rouse et al., 2016, 2017). On the other hand, attempts to study the structure of FapF by cryo-EM have, to date, not been reported. The recent
advancements in cryo-EM methodology have enabled various structures of high MW membrane proteins to be solved, such as PCAT1 in complex with its substrate, as previously described (Kieuvongngam et al., 2020; Nogales, 2016). However, established protocols for the solubilisation of FapF have so far only utilised detergents, which have been shown to compromise the stability of FapD (Hawthorne, 2016; Rouse et al., 2017). To overcome this problem, FapF can be incorporated into lipid nanodiscs (as discussed in the next chapter), which avoids the requirement for detergent, and that are compatible for study by cryo-EM (Denisov and Sligar, 2016; Efremov et al., 2017). Furthermore, unlike NMR spectroscopy or X-ray crystallography, cryo-EM does not require high sample concentrations, thus, avoiding the problem of FapD aggregation at high sample concentrations (Grassucci et al., 2007). This strategy presents a suitable approach for solving the structure of FapD in complex with FapF. It may also be possible to solve the structure of the FapD-FapF complex bound to a substrate, akin to the PCAT1-Cta structure (Kieuvongngam et al., 2020). ITC can be used to supplement this work by investigating the importance of various FapD and FapF residues that are discovered (or postulated) to be important for the FapD-FapF binding interaction (Leavitt and Freire, 2001). Such examples are included in Figure 9, which indicates the conserved residues between FapD and LahT147 that are proximal (<4 Å) to the LahT147 substrate binding site, and that might be involved in FapD-substrate interactions (Bobeica et al., 2019).

4.3.3. The Role of FapD Within the Fap System

Our molecular level understanding of the Fap system is limited by our lack of insight into the structures, interactions, conformational changes, and dynamics of the Fap proteins. The transient association between FapD and the FapF N-terminus, which was demonstrated in this work (in vitro), opens up a number of ideas for Fap substrate (FapB, FapC, and FapE) recruitment and substrate translocation mechanisms. Taking concepts from the chaperoning role for CsgE within the curli system, FapD may similarly chaperone unfolded Fap substrates in the periplasm and subsequently guide them to FapF for translocation (Goyal et al., 2014). The GG-like motifs of the Fap substrates may transiently bind to FapD, potentially via the predicted hydrophobic substrate binding site (active site) of FapD (Kieuvongngam et al., 2020; Kotake et al., 2008). The interaction between FapD and a Fap substrate may
occur in the absence of a cleavage event, and that retains the Fap substrate within a stable unfolded conformation to prevent its aggregation within the periplasm (Kotake et al., 2008; Lecher et al., 2012). Next, FapD and its bound cargo (Fap substrate) may bind to the FapF N-terminus at a specific location. FapD may subsequently undergo a conformational change to release its cargo towards the FapF TMD (Kieuvongngam et al., 2020). The FapF linker domain may then facilitate the transportation of the unfolded Fap substrates (via transient interactions) towards the FapF TMD for translocation into the ES (Perkins et al., 2010; Rouse et al., 2018b). Alternatively, FapD (without bound cargo) may bind to the FapF NTDs and remain transiently associated, akin to the PEP domain of PCAT1 (Kieuvongngam et al., 2020). In this model, FapD may form interactions with the Fap substrates (via their GG-like motifs) and guide them towards the FapF TMD for translocation into the ES (illustrated in Figure 59). This suggested translocation mechanism is more likely than the former due to the transient interaction that was identified between FapDT and CCExt in this study. Once the concentration of unfolded periplasmic Fap substrates decreases (presumably after their translocation has occurred), FapD may recognise and cleave the GG-like motif (FapF_{84-95}) present at the FapF linker domain, as under these circumstances, the FapF linker domain is no longer shielded by the unfolded Fap substrates (Rouse et al., 2017). The cleavage of FapF may result in the repositioning of the FapF plug, akin to FapFβ, resulting in the closure of the FapF pore (Rouse et al., 2017). This mechanism would prevent the translocation of non-specific substrates into or out of the periplasm (Goyal et al., 2014). The cleavage of the FapF linker domain is an idea supported by Rouse et al. (2017); within their report, whole-cell trypsin digestion and tandem MS were used to analyse the whole-cell lysates of P. sp. UK4, which expressed the fap operon, for signs of Fap protein degradation. With the exception of the FapE N-terminus and the FapF linker domain, all other Fap proteins were found to be intact (Rouse et al., 2017). Also, the postulated capacity for FapD to bind to some specific Fap GG-like motifs for cleavage (high interaction affinity), and bind to others in the absence of cleavage (low interaction affinity), could occur due to the substrate specificity of the PEP domains, as demonstrated by Kotake et al. (2008). Of note, within these models, the role of FapF trimerisation for Fap substrate secretion is unclear. Also, the role of FapA remains to be elucidated; previous studies indicate that FapA influences the Fap fibre composition, therefore, it may function to
chaperone specific Fap substrates (Dueholm et al., 2013b; Rouse et al., 2017). Further research is required to elucidate the Fap system translocation mechanism(s).

**Figure 59: The predicted function of FapD within the Fap system.** Within this model, FapD binds to FapF where it interacts with incoming and unfolded Fap substrates via their GG-like motifs, guiding them towards the FapF barrel lumen for extracellular export. At high Fap substrate concentrations, Fap substrates may shield the GG-like site in the FapF linker domain as they translocate up the FapF linker towards the TMD. Once Fap substrates are depleted from the periplasm, the GG-like site within the periplasmic linker domain of FapF may become accessible to recognition and cleavage by FapD (Rouse et al., 2017). Cleavage of the linker domain may result in displacement of the FapF helical plug into the FapF lumen, akin to the conformation of FapFβ (Rouse et al., 2017). The FapF pore closes, preventing non-specific substrate translocation into and/or out of the cell.
4.3.4. The Role of FapD Outside the Fap System

The study of FapD will no doubt yield further insight into Fap biogenesis. Further structural and functional studies may eventually enable and facilitate the design of small-molecules that bind to FapD, or other Fap proteins, to modulate the secretion activity of the extracellular Fap components, FapB, FapC, and FapE. Therefore, these studies have the potential to contribute to the dismantling of Fap-biofilm and its associated pathologies (Worthington et al., 2012). In addition, the high-degree of structural homology between the C39-proteases suggests that insight gathered for one C39-peptidase domain may effectively translate into useful insight for the entire family of PEP domains (Bobeica et al., 2019; Ishii et al., 2010; Lecher et al., 2012). Thus, if the structure and function of FapD is confirmed to be highly similar to those of the PEP domains, the insight generated for FapD may be translated to other members of the C39-peptidase family. Beyond the potential relationship between FapD and bacterial biofilms, further insight into the PEP domains and FapD may lead to a greater understanding of how to modulate the production of class II lantibiotics that possess antimicrobial activity (Furgerson et al., 2008; Majchrzykiewicz et al., 2010). This insight may prove useful for combatting AMR, and it may be of commercial use; for example, class II lantibiotics have a potential use in food preservation (Field et al., 2015; Mathur et al., 2018; Nes and Holo, 2000). Lastly, the C39-peptidase domains possess a high fidelity, which may have use within the realm of biotechnology; for instance, PEP domains could be used in the removal of protein tags akin to TEV protease (Furgerson et al., 2008; Raran-Kurussi et al., 2017).
Chapter 5 – Enabling the Study of FapFβ by Solution-State NMR Spectroscopy

5.1. Introduction

The structure of the FapF truncation, FapFβ, was solved by X-ray crystallography to reveal a conformation that is postulated to correspond to the closed state of FapF (Rouse et al., 2017). However, insight into the residue specific dynamics- and exchange-based molecular processes of FapFβ could not be obtained from the static snapshot of its crystal structure; instead, solution-state NMR spectroscopy represents a technique capable of detailing these processes, as well as protein-protein interactions, in near-native conditions (Markwick et al., 2008; Wiesner and Sprangers, 2015). To gain insight into the postulated gating mechanism of FapF, structure, function, and dynamics studies of both its open (FapF) and closed state (FapFβ) are required. This includes the study of the FapF helical plug, the conserved “PTG” motifs, the interactions between FapF and other Fap proteins, and the effects of various mutants (such as those postulated to stabilise the FapFβ plug; Figure 7D) (Chorev et al., 2020; Rouse et al., 2017). Additionally, environmental factors such as lipids and pH have previously been demonstrated to influence the pore gating of OmpF, an OMP (Liko et al., 2018). Thus, these factors can be examined for their effects, if any, on the structure, dynamics, and interactions of FapF and FapFβ (Rouse et al., 2017). In this chapter, our attention is focused towards enabling the study of FapFβ, a 106 kDa protein (FapFβ trimer), by solution-state NMR spectroscopy. The eventual aim of this study is to determine the factors that regulate the pore-open versus pore-closed state of FapF. As well as the various studies described above, the NMR spectra obtained for FapFβ will serve to facilitate the resonance assignment of the FapF TMD within equivalent NMR spectra collected for full length FapF in future investigations (Pfuhl and Driscoll, 2000). Lastly, beyond Fap biogenesis, the NMR studies of FapFβ may provide insight into the structure, function, and dynamics of the T5a ATs, such as the closed state of EspP, due to their structural homology (Barnard et al., 2007; Rouse et al., 2017).

Despite the unmatched utility of solution NMR for these studies, the conventional use of $^{15}$N and $^{13}$C isotopic labelling schemes for multidimensional triple-resonance heteronuclear NMR experiments, which are commonly employed for resonance
assignment, are limited at protein molecular weights above 25-30 kDa (Clore and Gronenborn, 1994; Puthenveetil and Vinogradova, 2019). Beyond these MWs, the likelihood of chemical shift overlap increases due to the presence of additional amino acids and their associated spectral resonances (Kwan et al., 2011; Puthenveetil and Vinogradova, 2019). Furthermore, spectral peak linewidths are correlated with molecular rotational correlation times \( (\tau_c) \) the time it takes for an approximately spherical protein to rotate one radian) (Pfuhl and Driscoll, 2000; Pandya et al., 2018). The Stokes-Einstein equation \( (\tau_c \approx 4\pi\eta r^3/3kT) \) relates the approximate protein correlation time \( (\tau_c) \) to its hydrodynamic radius \( (r) \), viscosity \( (\eta) \), temperature \( (T) \), and the Boltzmann constant \( (k) \) (Pandya et al., 2018). As \( \tau_c \) increases, the rate of spin-spin relaxation \( (R_2) \) increases, and the \( T_2 \) relaxation time decreases \( (R_2 = 1/T_2) \). The peak line width is inversely proportional to \( T_2 \), and therefore, peak line width increases in proportion to the correlation time \( (\text{MW}) \) of a protein molecule (reviewed in Pandya et al. (2018) and Pfuhl and Driscoll (2000)). The main factors contributing to \( T_2 \) relaxation (and therefore, peak broadening) are briefly described below; for a review of this topic, the reader is guided to Pfuhl and Driscoll (2000). Simply put, the \( T_2 \) relaxation mechanisms for high MW proteins are primarily the result of dipole-dipole (DD) interactions and chemical shift anisotropy (CSA). The DD interactions occur as a result of the magnetic field interactions between a pair of NMR-active nuclei; the strength of the dipolar interaction correlates with the rate of \( T_2 \) relaxation \( (R_2) \) (Pandya et al., 2018; Pfuhl and Driscoll, 2000). The strength of DD interactions are inversely proportional to sixth power of the distance between two nuclei, thus, proximal NMR active nuclei experience large increases in their \( R_2 \) (Pandya et al., 2018; Pfuhl and Driscoll, 2000). In addition, the strength of DD interactions is proportional to the second power of the gyromagnetic ratio (Breukels et al., 2011; Pfuhl and Driscoll, 2000). The second source of relaxation, CSA, is largely dependent on the covalent structure of a protein (for further descriptions of the CSA, see Pfuhl and Driscoll (2000)); as a result, the majority of methods employed to attenuate spin-spin relaxation for high MW proteins are directed towards the DD interactions (Pfuhl and Driscoll, 2000).

In one of the most commonly employed methods to suppress the DD interactions, the protons of a protein are commonly substituted for deuterons (in a process referred to as protein deuteration) that possess smaller gyromagnetic ratios \( (\gamma_H/\gamma_D = \sim 6.5) \) (Pfuhl and Driscoll, 2000; Sattler and Fesik, 1996). Secondly, the transverse
relaxation-optimized spectroscopy (TROSY)-based NMR experiments are commonly employed in combination with deuteration to increase the signal-to-noise and resolution of the NMR spectral peaks. The TROSY type experiments achieve this by exploiting the destructive interference between the DD and CSA interactions, and maximal interference occurs at field strengths of approximately 1 GHz, with respect to the amide groups (Pfuhl and Driscoll, 2000; Takeuchi et al., 2016). These methods have enabled larger macromolecules up to approximately 35 kDa to be studied by solution NMR spectroscopy (Puthenveetil and Vinogradova, 2019).

The selective protonation and $^{13}$C labelling of methyl groups in an otherwise perdeuterated sample represents a relatively novel method (referred to as “methyl labelling”) to overcome the problems of peak overlap and low sensitivity exhibited by high MW proteins (Pfuhl and Driscoll, 2000; Kerfah et al., 2015). Methyl groups exhibit high spectral sensitivity due to their proton multiplicity. In addition, methyl groups have the capacity to bypass the relaxation problems associated with high MW proteins due to their rapid internal rotation (Pfuhl and Driscoll, 2000; Kerfah et al., 2015). The establishment of cost-effective and efficient biosynthetic methods have enabled the selective isotopic labelling of one or more methyl-bearing amino acids in a protein by using the E. coli expression system. Resultantly, this has allowed methyl groups to serve as probes for insight into the structure, dynamics, and interactions of high MW proteins (Goto et al., 1999; Kerfah et al., 2015; Schütz and Sprangers, 2020).

In a study by Hajduk et al. (2000), the high sensitivity of the methyl labelling strategy was demonstrated by measuring the signal-to-noise (S/N) ratio of $^1$H-$^{15}$N HSQC and $^1$H-$^{13}$C HSQC spectra obtained for identical [U-$^2$H,$^{15}$N,$^{13}$CH$_3$]-protein samples (MWs of 42 kDa and 110 kDa). The results of this study indicated a gain in S/N of approximately 4-7-fold for the $^1$H-$^{13}$C HSQC compared to the $^1$H-$^{15}$N HSQC (using a 500 MHz spectrometer) (Hajduk et al., 2000). Furthermore, the S/N and resolution of a $^1$H-$^{13}$C NMR spectrum for a [U-$^2$H,$^{15}$N,$^{13}$CH$_3$]-protein sample can be further enhanced by utilising the $^1$H-$^{13}$C heteronuclear multiple quantum correlation (HMQC)-based NMR experiment that employs the methyl-TROSY effect (Tugarinov et al., 2003). Unlike the TROSY effect described for amide groups, for which the interference mechanism is derived from CSA and DD sources of relaxation, the HMQC is field independent, and it relies solely on the DD sources of relaxation (Hajduk et al., 2000; Pfuhl and Driscoll, 2000; Sheppard et al., 2010; Tugarinov et al.,
The gain in sensitivity for using the HMQC over the HSQC is reported to be up to 2.6-fold (for methyl-labelled samples) (Tugarinov and Kay, 2004). Therefore, the combination of protein methyl labelling, and the exploitation of the methyl-TROSY effect, has the potential to significantly increase the spectral S/N of high MW proteins and their complexes. The success of these methods is exemplified by the successful study of the dynamics and interactions of a 1.1 MDa-complex using solution-state NMR spectroscopy, as well as many other high MW protein systems (Hsu et al., 2009; Solt et al., 2017; Sprangers and Kay, 2007).

In order to extract dynamics and interaction based information from solution state NMR studies of high MW membrane proteins, such as FapFβ, advanced (costly) isotope labelling strategies are required (O’Brien et al., 2018). Therefore, the protocols utilised for the recombinant expression and purification of isotopically labelled protein must be efficient in order to minimise sample production costs. Previous reports have detailed the expression of FapFβ using an in vivo method to incorporate FapFβ into the E. coli outer-membrane (OM) for extraction, however, the yield of FapFβ resulting from this method is insufficient for complex NMR studies (personal communication with Dr Sarah Rouse) (Bannwarth and Schulz, 2003; Noinaj et al., 2016; Rouse et al., 2016). In contrast, Hawthorne (2016) successfully refolded FapF and FapF truncations (of the P. aeruginosa PAO1 strain) from inclusion bodies (IBs), resulting in high yields of purified FapF, which is sufficient for NMR studies (Bannwarth and Schulz, 2003; Hawthorne, 2016; Schwarzer et al., 2017). In this chapter, we apply and optimise the previously established FapF refolding protocol to extract and recover high-yields of soluble FapFβ from IBs. Additionally, a protocol for the transfer of FapFβ from a micellular environment to nanodiscs, a near-native lipid based environment, is established (Ritchie et al., 2009). The challenges associated with studying FapFβ by NMR spectroscopy, which are due to its poor relaxational properties (associated with its high MW), are also demonstrated, and the impacts of various isotopic labelling strategies to overcome this problem are examined (Puthenveetil and Vinogradova, 2019). This chapter aims to pave the way for future research into the structure, dynamics, and interactions of FapF and its mutants.
5.2. Strategies for the Expression of FapFβ

Membrane protein expression is typically conducted either via an *in vivo* (native) approach, in which the outer membrane protein (OMP) is expressed and transported by cellular machinery into the OM, and that is subsequently extracted for purification; or the *in vitro* refolding approach, in which the OMP is expressed into the cytoplasm to form insoluble IBs, and the OMP is subsequently solubilised and refolded *in vitro* (Bannwarth and Schulz, 2003; Noinaj et al., 2016). Each of these strategies possess their own unique advantages and disadvantages. The native approach is more likely to result in correctly folded protein, as the OMP is natively inserted into the OM and subsequently extracted by use of buffers containing detergent (Bannwarth and Schulz, 2003; Noinaj et al., 2016). However, the yield of OMP that can be extracted is often relatively low, as the over-expression of OMPs in the native approach results in the congestion of the cellular machinery that is responsible for translocating OMPs into the periplasm (Bannwarth and Schulz, 2003; Wagner et al., 2006). Furthermore, the limited space of the periplasmic OM restricts the quantity of OMPs that can be inserted into the OM (Wagner et al., 2006). On the other hand, membrane proteins are frequently expressed without Sec signal sequences in order to direct their expression into the cytoplasm for the formation of IBs (Noinaj et al., 2016). The main advantage of this method is the production of high yields of target OMP, as this method does not require the cellular machinery or membrane space described above (Bannwarth and Schulz, 2003). However, IBs must be solubilised and refolded, that is often time-consuming, inefficient, and the associated protocols usually require extensive fine-tuning (Noinaj et al., 2016; Popot, 2014; Schwarzer et al., 2017). Despite these problems, various reports have demonstrated the successful refolding of β-barrel membrane proteins from IBs for structural studies (Bannwarth and Schulz, 2003; Buchanan, 1999).

Regarding FapFβ, a previously published protocol for its native extraction and purification from the OM was detailed, and up to 0.1 mg of purified FapFβ could be recovered from 1 L of expression in LB media (personal communication with Dr Sarah Rouse) (Rouse et al., 2016, 2017). The low yield of FapFβ obtained from this type of preparation is cost inefficient for NMR studies (O’Brien et al., 2018). Therefore, to significantly increase the yield of FapFβ, a protocol for the refolding of FapFβ is necessary. To obtain soluble, folded FapFβ from the IBs, the Sec-secretion
signal is removed from the FapF gene to enable expression into the cytoplasm, and a multi-step denaturation and refolding protocol is employed (Bannwarth and Schulz, 2003; Hawthorne, 2016). In principle, the procedure for OMP refolding involves three key steps (reviewed in Junge et al. (2008)). First, IBs (containing the overexpressed OMP of interest) are solubilised in buffer containing either a chaotrope, a strong detergent, or an organic solvent. Second, the solubilised IBs are transferred into a buffer containing a strong detergent (to mimic the lipid bilayer of the bacterial OM), which transfers the protein from an unfolded soluble state to a soluble folded state (Junge et al., 2008). Third, the sample is exchanged into a buffer containing a mild detergent (or lower detergent concentration) that is more suitable for downstream functional or biophysical analysis (Junge et al., 2008; Linke, 2009).

The OMP refolding procedure involves a complex intertwining of several variables that influence the balance between misfolding/aggregation and the successful renaturation of the native OMP conformation (Schwarzer et al., 2017). This includes the refolding concentration of the OMP, the composition of the refolding buffer (the detergent concentration), and other refolding parameters such as the temperature of refolding, and the refolding method (Schwarzer et al., 2017). The high variability of the conditions used for refolding, even amongst the OM β-barrel proteins, suggests that there is no universal set of condition for the successful refolding of membrane proteins (Arachea et al., 2012; Buchanan, 1999). Favourably, a protocol for the refolding of full length FapF, and FapF truncations, has been reported, and that was used as a guide in this study for the refolding of FapFβ (Hawthorne, 2016). Within this protocol, FapF inclusion bodies are solubilised in denaturing buffer (containing 8 M urea) that is subsequently purified and concentrated by Ni-IMAC. The eluent, containing unfolded FapF, was refolded by pulse refolding. At this step, full length FapF, which is solubilised in a volume of denaturing buffer (containing 8 M urea), is slowly titrated (drop-wise) into a 20-fold higher volume of stirred refolding buffer (containing 5 % (v/v) LDAO) at room temperature. The refolding solution is subsequently purified and concentrated by Ni-IMAC, and lastly, subjected to SEC for clean-up (Hawthorne, 2016). However, in the report by Hawthorne (2016), the impact of the FapF refolding concentration and the concentration of LDAO within the refolding buffer on the refolding efficiency of FapF were not detailed. Here, we investigate and optimise the independent effects of these two important refolding parameters on refolding efficiency of FapFβ (Schwarzer et al., 2017). This work
aimed to increase the efficiency of the FapF refolding protocol, in order to reduce the overall costs of sample production for NMR studies (O’Brien et al., 2018).

5.3. Results
5.3.1. FapFβ Expression Trials

FapFβ (FapF_{107-430}) was sub-cloned (using the LIC cloning method) from a pMMB190 vector containing the *P. sp. UK4 fapABCDEF* operon to the pNIC28-Bsa4 expression vector (described previously), for the purpose of cytoplasmic expression (inclusion body formation). The FapFβ construct (schematic shown in Figure 60) was subsequently transformed into *E. coli* BL21 (DE3) for overexpression.

Figure 60: Schematics of the domain organisation of FapF and the FapFβ construct utilised. The domain boundaries of FapF, as approximated by Rouse et al. (2017), are illustrated as a schematic. Each domain is labelled according to its position in the amino acid chain of FapF, and is individually coloured as in Figure 7A. The FapF deletion mutant utilised throughout this chapter, FapF_{107-430}, is also illustrated, and is colour coded according to the descriptions for Figure 7A. The N-terminal fusion tag of the construct is abbreviated to “NFT”; its corresponding amino acid sequence is “MHHHHHHSSGTE NLYFQSM” (written N- to C-terminus; detailed within section 2.1.5.).

Expression trials for FapFβ were conducted by examining its insoluble expression under four conditions: 18 °C, 25 °C, and 30 °C for 14 hours, as well as 37 °C for 4 hours. Analysis by SDS-PAGE indicates that FapFβ (theoretical MW is 37.5 kDa) is expressed in all the insoluble fractions over all expression conditions tested (Figure 61). The highest level of insoluble expression resulted from a 14 hour incubation at 30 °C. This condition was taken forward for expression of FapFβ in large volumes.
5.3.2. FapFβ Inclusion Body Solubilisation

Following the solubilisation and refolding protocol for FapF that is previously reported by Hawthorne (2016), the cell pellets collected after the overexpression of FapFβ were lysed, and the insoluble matter (IBs) was collected by high-speed centrifugation (Hawthorne, 2016). The insoluble matter (~6 g) was dissolved into 120 ml of buffer containing 8 M urea, which is a denaturing agent, and subsequently centrifugated to collect the soluble fraction (Popot, 2014). To determine the whether a 20 ml:1 g ratio of denaturing buffer (DB; 20 mM Tris-HCl, 300 mM NaCl, 8 M urea, pH 8.0) to IBs (mass) is sufficient to solubilise all or most of FapFβ, a fraction of the supernatant was collected for SDS-PAGE analysis. The insoluble matter was subsequently redissolved in 20 ml of DB, and the soluble fraction was again collected for analysis. SDS-PAGE analysis indicates that the majority of FapFβ present within the IBs is dissolved in 120 ml of DB, and a 20 ml:1 g ratio of DB (ml) to IB (g) was taken forward for all FapFβ preparations (Figure 62A).
Figure 62: Solubilisation of FapFβ from inclusion bodies. (A) SDS-PAGE gel analysis indicates that FapFβ is extracted from inclusion bodies using 120 ml of denaturing buffer (DB) (lane 1), and re-solubilisation of the insoluble matter into 20 ml of DB after centrifugation (lane 2) yields a small quantity of FapFβ. (B) SDS-PAGE gel analysis indicates the sample fraction corresponding to the inclusion bodies (Ins), and the fraction corresponding to the denaturing Ni-IMAC elution fraction (DE).

5.3.3. Optimising the Refolding of FapFβ

After solubilisation of the IBs, the sample was subjected to Ni-IMAC for purification and concentration. SDS-PAGE analysis indicates that the purity of FapFβ in the Ni-IMAC eluent is slightly improved compared to the IBs, and it is suitable for pulse refolding (Figure 62B). In the refolding step of the protocol previously established by Hawthorne (2016), the FapF sample (in DB) is diluted into a 20-fold greater volume of non-denaturing buffer (NDB; methods 2.4.10.) containing 5 % (v/v) LDAO (Hawthorne, 2016). This was performed drop-wise into a stirring solution of NDB at room temperature. The 20-fold dilution event reduces the concentration of urea in the sample from 8 M to 0.4 M, which is a sub-denaturing concentration (Biggar et al., 2012; Lambrechts et al., 2002). In addition, the high concentration of LDAO (~100-fold greater than the LDAO critical micelle concentration (CMC)) promotes the formation of LDAO micelles that stabilise FapF (Hawthorne, 2016; le Maire et al., 2000). However, the refolding concentration of FapF (final concentration of FapF in the NDB) was not described by Hawthorne (2016). The optimal refolding concentration of membrane proteins (MPs) is typically 10-200 µg/ml (final concentration); lower MP concentrations reduce the probability of MP aggregation during the refolding step (Schwarzer et al., 2017; Vallejo and Rinas, 2004). Additionally, the concentration of detergent (LDAO in this study) within the NDB has also been shown to affect the refolding efficiency of MPs, and that was not reportedly optimised by Hawthorne (2016). Detergent concentrations above their respective CMCs enable the formation of detergent micelles that stabilise membrane
proteins (Gall et al., 2003; Schwarzer et al., 2017). However, high detergent concentrations (excess) may destabilise MPs (Otzen et al., 2009; Palazzo et al., 2010; Yang et al., 2014). This study aims to investigate the effect of the FapFβ refolding concentration, as well as the NDB LDAO concentration, on the refolding efficiency of FapFβ.

The effect of the FapFβ refolding concentration on its relative refolding efficiency was investigated by diluting FapFβ samples at 0.25, 0.50, and 1.00 mg/ml 20-fold into NDB containing 5 % (v/v) LDAO; the concentration of FapFβ within the refolding buffer after dilution (termed the “refolding concentration”) is 12.5, 25.0, and 50.0 µg/ml, respectively; this experiment is referred to as the “FapFβ variable concentration” (FVC). A total of 5.5 mg of FapFβ was refolded per experiment. In a second batch of experiments, the effect of the LDAO concentration (that is present in the NDB (refolding buffer)) on the refolding efficiency of FapFβ was investigated by diluting 0.50 mg/ml (25.0 µg/ml refolding concentration) samples of FapFβ into NDB containing LDAO at concentrations of 1.0 %, 2.5 %, 5.0 %, and 10.0 % (v/v); this experiment is referred to as the “LDAO variable concentration” (LVC). A total of 6.0 mg of FapFβ was refolded per experiment. In both sets of FVC and LVC experiments, the 20-fold dilution of FapFβ was performed drop-to-drop at a rate of approximately 0.25 ml/min by use of a peristaltic pump. The LVC and FVC refolding mixtures were gently stirred at 20 °C for 12 and 18 hours, respectively. The refolding mixtures for the LVC and FVC were subsequently dialysed against NDB containing 0.5 % (v/v) LDAO at 4 °C, as performed by Hawthorne (2016). The dialysis step reduces the sample LDAO concentration, as well as the concentration of residual imidazole and urea. After six hours of dialysis, each sample was collected and centrifuged to check for the presence of precipitate. With the exception of a small quantity of precipitate observed for the FapFβ sample that was refolded in NDB containing 10 % LDAO, no precipitate was observed in the other fractions. SDS-PAGE analysis was used to confirm the presence of FapFβ within the soluble fraction and precipitate of the 10 % LDAO LVC sample (Figure 63). This suggests that the presence of 10 % LDAO (v/v) within the NDB may have destabilised FapFβ, which resulted in the misfolding and aggregation of FapFβ (Otzen et al., 2009; Palazzo et al., 2010; Yang et al., 2014). After dialysis, the LVC and FVC samples were individually subjected to Ni-IMAC for the purpose of sample purification and concentration.
Figure 63: FapFβ precipitate. SDS-PAGE analysis indicates an example of the soluble (Sol) and insoluble (Ins) fractions collected after the refolding (after dialysis) of FapFβ.

To evaluate the relative refolding efficiencies for each sample condition, the Ni-IMAC eluents were individually analysed by SEC. The SEC UV traces corresponding to samples from the LVC and FVC are shown overlaid in Figure 64A and B, respectively. The SEC results for all sample conditions indicate a single broad symmetrical peak with a retention volume of ~85 ml. SDS-PAGE analysis indicates that the elution fraction corresponding to 84-86 ml (in all samples) is comprised of FapFβ at high purity (Figure 64C, D, and E). The late elution of this peak for all samples suggests the presence of a folded FapFβ species, and that possesses a similar or identical global conformation across all samples (due to their similar or identical hydrodynamic radii) (Hong et al., 2012). In the SEC profiles corresponding to the FVC samples, the low intensity peak at ~50 ml may correspond to misfolded FapFβ or FapFβ aggregates. Since the SEC profiles corresponding to the LVC samples do not indicate the presence of a similar peak at ~50 ml, it suggests that the extended refolding time (6 additional hours) for the FVC samples may have contributed to the destabilisation (aggregation) of FapFβ. Of note, the off-baseline SEC UV absorbance between 50 ml and 70 ml for the LVC samples may correspond to higher FapFβ oligomeric states, and/or FapFβ aggregates (Hong et al., 2012).
The SEC UV traces that correspond to the LVC samples indicate that the intensity of the UV absorbance for the main peak at ~85 ml increases as the LDAO concentration within the refolding buffer is decreased from 10.0 % to 1.0 %. Increasing the concentration of LDAO within the refolding buffer from 1.0 % to 2.5 % or 5.0 % results in a minor decrease in the SEC peak UV intensity; however, the peak UV intensity is significantly decreased for the FapFβ sample refolded in 10.0 % LDAO. This is consistent with the precipitation for this sample that was observed after dialysis. Together, these results indicate that higher concentrations of LDAO in the refolding buffer, within the range examined, may decrease the stability of FapFβ that results in sample aggregation. To determine the final yield for each refolding condition, the fractions corresponding to the main SEC peak (82 to 88 ml) for each sample were pooled and concentrated to 60 µM (to improve the accuracy of sample concentration measurements) using a 50 kDa MWCO concentrator. Of note, a 50 kDa MWCO concentrator was utilised to prevent the build up of LDAO micelles (le Maire
et al., 2000). For LVC samples (starting quantity of 6.0 mg of FapFβ), the recovery (and percentage recovery) obtained from the refolding of FapFβ in NDB containing 1.0 %, 2.5 %, 5.0 %, and 10.0 % LDAO yielded 2.6 mg (43 %), 2.4 mg (40 %), 2.3 mg (38 %), and 1.4 mg (23 %), respectively. Of note, the values for yield that are indicated do not correspond to the true refolding efficiency for each sample refolding condition, as multiple purification and concentration steps are incorporated prior to sample yield determination.

The SEC UV traces for the FVC samples indicate that the intensity of the UV absorbance for the main peak at ~85 ml increases as the FapFβ refolding concentration is increased from 12.5 to 50.0 µg/ml. Increasing the concentration of refolded FapFβ from 12.5 to 25.0 to 50.0 µg/ml results in a minor and progressive increase in the peak UV intensity. Together, these results indicate that lower concentrations of refolded FapFβ, within the range examined, may decrease the stability of FapFβ during refolding, which results in sample aggregation. To determine the final yield for each refolding condition, each sample was prepared for concentration measurement as described above. For FVC samples (starting quantity of 5.5 mg of FapFβ), the recovery (and percentage recovery) obtained from the FapFβ refolding concentrations for 12.5, 25.0, and 50.0 µg/ml yielded 1.8 mg (33 %), 2.1 mg (38 %), and 2.4 mg (44 %), respectively.

One dimensional 1H-NMR spectroscopy was employed to crudely determine whether FapFβ, from the fractions corresponding to the main SEC peak for each of the FVC and LVC samples, possesses comparable folded domains. The fractions corresponding to the main SEC peak (82 to 88 ml) for each sample were pooled and concentrated to approximately 60 µM as previously described. The 1D 1H-NMR spectra for the FVC samples are overlaid in Figure 65A, and the 1D 1H-NMR spectra for the LVC samples are overlaid in Figure 65B. The 1H-NMR spectrum for each sample demonstrates the presence of peaks downfield of 8.5 ppm and upfield of 0 ppm, which indicates the presence of a tertiary folded domain (Kwan et al., 2011). Furthermore, each set of spectra that correspond to the FapFβ FVC and LVC samples superimpose with no obvious differences, suggesting that the conformation of FapFβ is identical within each set of samples. Both sets of spectra also indicate large line-widths, which are due to the high MW of FapFβ (monomer MW is 37.5 kDa) in a LDAO micelle (approximately 17 kDa), and that results in poor relaxation properties (Le Maire et al., 2000; Pfuhl and Driscoll, 2000).
Figure 65: Refolded FapFβ was obtained under all FapFβ refolding conditions examined. (A) The 1D 1H-NMR spectra corresponding to purified samples of FapFβ that were refolded at various FapFβ concentrations (refolding concentration) are overlaid. (B) The 1D 1H-NMR spectra corresponding to purified samples of FapFβ that were refolded at various concentrations of LDAO in NDB are overlaid. All FapFβ sample conditions: 60 µM, 298 K, 512 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe.
The refolding efficiency of FapFβ will likely be improved by further adjustment of the refolding parameters, however, no further optimisation of these parameters was performed in this study (Schwarzer et al., 2017). To produce samples of FapFβ, which are later described in this chapter, FapFβ was refolded at a refolding concentration of 50.0 µg/ml in NDB containing 1.0 % LDAO. This protocol yields approximately 11 mg of refolded and purified FapFβ per 1 L of culture grown in LB media, with a final FapFβ recovery of 46 %.

5.3.4. Probing the Oligomeric State of Refolded FapFβ

To probe the oligomeric state of FapFβ, FapFβ was expressed, refolded, and purified according to the methods previously described; fractions corresponding to the main peak from SEC were pooled and concentrated to 3.3 mg/ml, and the average MW of FapFβ was determined by SEC-MALS. SEC-MALS indicates that the average MW of FapFβ is 98.0 kDa ± 3.2 kDa (Figure 66). Furthermore, SEC-MALS reveals that FapFβ is somewhat polydisperse, with the average MW ranging from ~103 kDa to ~93 kDa. The theoretical MW of FapFβ is 37.5 kDa, and the MW of the LDAO micelle is reported to be ~17 kDa (le Maire et al., 2000). Together, the refolded FapFβ monomer within the LDAO micelle is expected to form a 54.5 kDa complex; a dimer and trimer would approximate to 109.0 kDa and 163.5 kDa, respectively. By this calculation, the average MW of FapFβ corresponds to ~1.8 FapFβ-LDAO micelles. Together, this data suggests that FapFβ exists in an equilibrium of multiple oligomeric states, which likely includes FapFβ monomers, dimers, and trimers.

![Figure 66: SEC-MALS analysis of FapFβ in buffer containing 0.1 % LDAO.](image)

The SEC-MALS data indicates that FapFβ possesses an average MW of ~98.0 kDa (the average MW value is derived from the solid red line corresponding to the molar mass). The MW of the sample varies across the RIU peak (solid black line) indicating that the sample is polydisperse. The sample concentration was 3.3 mg/ml, and 0.5 ml was loaded into the S200A column.
5.3.5. $^1$H-$^{15}$N HSQC Spectrum of FapFβ

To determine whether a $^{15}$N-FapFβ sample would be amenable to conventional $^{15}$N/$^{13}$C 3D NMR experiments, a $^{15}$N-FapFβ sample was prepared and used to collect a 2D $^1$H-$^{15}$N HSQC spectrum (Kwan et al., 2011). To prepare a uniformly $^{15}$N-labelled FapFβ sample, FapFβ was expressed in $^{15}$N minimal media, refolded, and purified (as previously described). The $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-FapFβ indicates significant peak dispersion downfield of 8.5 ppm that is indicative of a folded protein that possesses β-sheet structure (Figure 67). Additionally, a high proportion of peaks were clustered between 8.0 and 8.5 ppm that is characteristic of solvent-exposed unstructured protein regions. These peaks may correspond to the flexible loops between the FapFβ β-strands, as well as the α-helical plug (Kwan et al., 2011). As expected for a high MW protein, very few peaks could be clearly defined due to peak broadening (poor relaxation properties of FapFβ) (Pfuhl and Driscoll, 2000). Furthermore, 359 amide peaks are expected to be present on the HSQC, and their abundance contributes to the extensive peak overlap observed. Together, these results highlight the need for alternative isotopic labelling schemes in order to improve spectral resolution and sensitivity for samples of FapFβ.

Figure 67: $^1$H-$^{15}$N HSQC spectrum of FapFβ. The $^1$H-$^{15}$N HSQC spectrum for $^{15}$N-labelled FapFβ is displayed. Sample conditions: 200 μM, 303 K, 256 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe.
5.3.6. Detergent Exchange

In previous reports, FapFβ was initially solubilised in buffer containing LDAO before being exchanged into buffer containing 0.5 % (v/v) C8E4 for various studies, including X-ray crystallography (Rouse et al., 2016, 2017). The properties of LDAO and C8E4 differ; LDAO is a zwitterionic detergent that is classified as “harsh”, whereas, C8E4 is a non-ionic detergent that is classified as “mild” (le Maire et al., 2000; Linke, 2009). Describing these detergents as harsh or mild refers to their capacity to destabilise membrane protein structures (harsh being more likely to result in MP instability) (le Maire et al., 2000; Linke, 2009; Seddon et al., 2004). A list of detergents and their corresponding properties are presented by le Maire et al. (2000).

The NMR spectral quality of MPs is reported to vary depending on the choice of membrane mimetic (Krueger-Koplin et al., 2004). In this study, refolded and purified FapFβ is transferred from buffer containing 0.1 % (v/v) LDAO to buffer containing 0.5 % (v/v) C8E4 in order to examine the effects of each detergent on the NMR spectral quality of FapFβ. NMR studies of FapFβ in buffers containing each detergent will also allow us to determine whether the conformation of FapFβ is consistent, or altered, between the membrane mimetics.

A protocol for the slow exchange of a purified FapFβ sample from NDB containing 0.1 % (v/v) LDAO to NDB containing 0.5 % (v/v) C8E4, using a Ni-IMAC step, was previously established and detailed by Rouse et al. (2016) (described in 2.4.11.). Of note, SDS-PAGE analysis indicates that a small quantity of FapFβ was lost during the Ni-IMAC buffer exchange step (Figure 68A). The Ni-IMAC eluent (FapFβ in NDB containing 0.5 % C8E4) was applied to a SEC column to remove aggregates and imidazole. The SEC UV traces that correspond to FapFβ in NDB containing 0.1 % (v/v) LDAO (FapFβL) and FapFβ in NDB containing 0.5 % (v/v) C8E4 (FapFβC) are overlaid (Figure 68B), indicating that both SEC traces are comprised of a highly symmetrical peak that elute at near identical volumes; of note, SDS-PAGE analysis confirms the presence of highly pure FapFβC (Figure 68C). This suggests that the Stokes radii of FapFβL and FapFβC are near-identical, and suggests that the oligomeric state(s) and conformation(s) of FapFβ for each buffer condition is similar (Hong et al., 2012).
5.3.7. Comparing the Secondary Structure Composition of Native FapFβ and Refolded FapFβ

To determine whether the secondary structure compositions of refolded FapFβC and native FapFβC (nFapFβC) are consistent, far-UV CD spectroscopy was employed. A purified nFapFβC sample (detailed in Rouse et al. (2016)) was kindly provided by Dr Sarah Rouse (expressed and purified in accordance with protocols in Rouse et al. (2016)). Both samples were dialysed against buffer containing 20 mM PO₄, 150 mM NaCl, 0.5 % C8E4, pH 6.8, and diluted to 0.10 mg/ml. The overlaid CD spectra indicate that the CD profiles for FapFβC and nFapFβC are near identical throughout the entire 200 to 260 nm wavelength range (Figure 69). The presence of a single CD minimum at ~218 nm for both spectra is consistent with the presence of β-sheet rich secondary structure (Greenfield, 2006b). The slight differences in CD intensity between 200 and 210 nm may be attributed to the presence of a TEV cleavage site present for FapFβC, which is not present for nFapFβC. Further structural validation is required to confirm whether their tertiary structures, which includes the localisation of the plug domain, are conserved (Kelly and Price, 2000). The CD data corresponding to nFapFβC was published in Rouse et al. (2017).
5.3.8. NMR Studies of Partially Deuterated FapFβ

One of the main contributors to broad spectral line widths for high MW proteins in NMR are the $^1$H-$^1$H dipolar interactions (Pfuhl and Driscoll, 2000). In order to improve the relaxational properties of FapFβ (improve NMR spectral line widths), a [85 %-$^2$H,$^1$U-$^{15}$N]-FapFβL sample was prepared and used to collect a $^1$H-$^{15}$N TROSY-HSQC spectrum (Markus et al., 1994; Pfuhl and Driscoll, 2000). In addition, the spectral differences between [85 %-$^2$H,$^1$U-$^{15}$N]-FapFβL and [85 %-$^2$H,$^{15}$N]-FapFβC were investigated by recording a 1D $^1$H-NMR and a $^1$H-$^{15}$N TROSY HSQC spectrum for each sample using an 800 MHz NMR spectrometer. These results will be used to examine whether a combination of the partial deuteration of FapFβ, and the TROSY-methodology, will enable the downstream study and spectral assignment of FapFβ using heteronuclear NMR experiments (Clore and Gronenborn, 1994).

To prepare a uniformly $^{15}$N-labelled FapFβ sample with overall deuteration of ~85 %, *E. coli* were cultured in $^{15}$N minimal media that is formulated in D$_2$O rather than H$_2$O (Leiting et al., 1998; Pfuhl and Driscoll, 2000). This isotopic labelling strategy has been suggested to result in ~99 % deuterium enrichment of the Hα site (Pfuhl and Driscoll, 2000). Of note, the proportion (%) of FapFβ deuteration was undetermined for either FapFβ sample, however, the samples produced in this study are referred to as 85 % deuterated based on the descriptions of Pfuhl and Driscoll (2000). Once expressed, FapFβ was refolded and purified as previously described (using H$_2$O formulated buffers); this strategy results in the protonation of the amide groups (due
to exchange with protons from the solvent), thus, enabling the recording of $^1$H-$^{15}$N-based spectra (Pfuhl and Driscoll, 2000).

The 1D $^1$H-NMR spectrum of $[85 \%$-$^2$H,$^1$U-$^{15}$N]-FapFβL and $[85 \%$-$^2$H,$^1$U-$^{15}$N]-FapFβC (referred to as $^2$H-FapFβL and $^2$H-FapFβC, respectively) are overlaid. The spectra indicate a significant increase in the S/N and resolution of the peaks downfield of 8 ppm (amides) and upfield of 0 ppm (upfield shifted methyl groups) compared to the non-deuterated FapFβL ($^1$H-FapFβL) 1D $^1$H-NMR spectra (Figure 70). The overlaid 1D $^1$H-NMR spectra of $^2$H-FapFβL and $^2$H-FapFβC display near perfect overlap in their respective amide and methyl regions, suggesting that the tertiary folding of FapFβ is consistent between the LDAO and C8E4 micelles.

![Figure 70: 1D $^1$H-NMR spectrum of partially deuterated FapFβ.](image)

The 2D $^1$H-$^{15}$N TROSY HSQC (TROSY) spectrum of $^2$H-FapFβL and $^2$H-FapFβC are overlaid, similarly indicating that the S/N and peak resolution of amide peaks are significantly improved compared to the non-deuterated $^{15}$N-FapFβL $^1$H-$^{15}$N HSQC spectrum (Figure 71). Both TROSY spectra indicate a number of peaks downfield of 8.5 ppm that correspond to the presence of a tertiary structured domain, and the large
clustering of peaks between 8.0 and 8.5 ppm may correspond to the solvent exposed unstructured regions of FapFβ, as previously described (Kwan et al., 2011). The majority of peaks between both TROSY spectra are somewhat overlapping, however, a number of peaks display minor differences in their chemical shifts. This suggests that FapFβL and FapFβC adopt a similar tertiary folding, and the minor chemical shift differences between each TROSY spectrum may occur as a result of minor perturbations in the FapFβ structure. Structural perturbations may occur due to mismatches between the dimensions of each micelle and the dimensions of the hydrophobic surface of FapFβ (Columbus et al., 2009). The TROSY spectra indicate a substantial reduction in the amide peak line widths, and that has significantly increased the number of clearly defined peaks compared to the previous HSQC spectrum corresponding to the non-deuterated 15N-FapFβL sample. A total of 359 peaks are expected on the TROSY spectrum for FapFβ (assuming a symmetrical FapFβ oligomer, or monomer); approximately 260 and 267 peaks could be identified on the TROSY spectrum corresponding to 2H-FapFβL and 2H-FapFβC, respectively. However, there is substantial peak overlap in the centre of each TROSY spectrum, making it difficult to count peaks within this spectral region. Interestingly, the peak intensity across both TROSY spectra is heterogeneous. This is highlighted by the peaks downfield of 9.8 ppm (1H) in both TROSY spectra that typically correspond to the tryptophan side chains (Larion et al., 2010). Four tryptophan side chain peaks are expected, however, five peaks are observed in both the TROSY spectra. One of these five peaks is of significantly lower intensity compared to the remaining four peaks. Together, these observations suggest the presence of slow exchange motions on the NMR timescale, such as those potentially occurring between FapFβ conformers within the LDAO or C8E4 micelle, the flexible loops between β-strands, and the exposed amides that may be in proximity to the detergent micelle (Columbus et al., 2009; Kleckner and Foster, 2011; Liang and Tamm, 2007; Rouse et al., 2017).
Figure 71: $^{1}$H-$^{15}$N TROSY HSQC spectrum of partially deuterated FapFβ. The $^{1}$H-$^{15}$N TROSY HSQC spectra for [85 %-$^{2}$H,U-$^{15}$N]-FapFβL (blue) and [85 %-$^{2}$H,U-$^{15}$N]-FapFβC (red) are overlaid. The spectra display improved line widths and sensitivity compared to non-deuterated FapFβ samples. $^{2}$H-FapFβL-$^{2}$H-FapFβC sample conditions: 200/150 µM, 303 K, 512 scans, performed on the Avance II HD 800 MHz spectrometer with cryoprobe.

Together, these results indicate that deuteration of FapFβ, in combination with TROSY-methodology and an 800 MHz NMR spectrometer, significantly improves the NMR spectral quality compared to the $^{15}$N-FapFβL HSQC recorded on a 600 MHz NMR spectrometer. However, the high proportion of missing peaks, as well as the abundance of overlapping peaks and low intensity peaks, suggests that $^{2}$H-FapFβ is not suitable for the canonical $^{13}$C/$^{15}$N-based triple resonance NMR experiments that are used for resonance assignment purposes (Kwan et al., 2011). Alternative isotopic labelling strategies are required to reduce the extent of peak overlap within the NMR spectrum, as well as increase the spectral sensitivity. Therefore, methyl-labelling strategies are subsequently explored (Kerfah et al., 2015; Schütz and Sprangers, 2020).

5.3.9. Methyl Labelling Strategy

In theory, the methyl-labelling of a high MW protein in an otherwise perdeuterated background of nuclei, in combination with HMQC methodology, represents a suitable
strategy for increasing NMR spectral sensitivity and minimising spectral peak overlap (Kerfah et al., 2015; Rosenzweig and Kay, 2014; Schütz and Sprangers, 2020). These methods were explored to determine their compatibility for the downstream study of FapFβ by complex heteronuclear NMR experiments (Clore and Gronenborn, 1994; Schütz and Sprangers, 2020).

FapFβ is comprised of 112 methyl containing residues, constituting 33% of all the FapFβ residues. Methyl-group resonances reside within a narrow chemical shift range (~20 ppm and ~2 ppm for $^{13}$C and $^1$H, respectively) (Schütz and Sprangers, 2020). Therefore, isotopically labelling all FapFβ methyl groups simultaneously would likely lead to substantial peak overlap. Additionally, the abundance of protons may hinder the experimental sensitivity due to their strong dipole-dipole interactions, as that could lead to to increased spin-spin relaxation rates; thus, methyl-labelling both the intra-residue methyl groups of Ile, Leu, and Val should be avoided (Pfuhl and Driscoll, 2000). On the other hand, compromising the number of observable NMR active nuclei will limit the amount of information that can be extracted from each NMR experiment (Kerfah et al., 2015). Therefore, an optimal methyl-labelling strategy is required that maximises the number of isotopically labelled methyl groups across the entire FapFβ structure, whilst minimising the quantity of overlapping peaks and strong dipole-dipole interactions (Pfuhl and Driscoll, 2000; Kerfah et al., 2015). Additionally, one of the aims of this study includes the characterisation of the FapFβ helical plug (dynamics, exchange, and interactions) by NMR spectroscopy. Within the helical plug, two alanines, one valine, and one leucine are present, thus, the methyl-labelling strategy should include at least one of these residues (Rouse et al., 2017).

To determine which of the methyl group containing amino acids of FapFβ should be isotopically labelled, CH3Shift was employed in order to predict the $^1$H-$^{13}$C methyl group chemical shifts for all methyl containing groups of FapFβ (with the exception of methionine) on the basis of the FapFβ structure (PDB ID: 5O65) (Sahakyan et al., 2011). The results of the CH3Shift prediction were used to approximate the distribution of FapFβ methyl group chemical shifts, without over-analysing the predicted chemical shifts for each residue, due to the limited accuracy of the CH3Shift method that is later discussed (Nerli et al., 2018; Sahakyan et al., 2011). The predicted chemical shifts are plotted onto a $^1$H-$^{13}$C HSQC-like correlation graph, and the peaks corresponding to the predicted (and distinguishable) methyl group chemical shifts of the helical-plug are indicated (Figure 72A). The CH3Shift
prediction indicates that the Ala-β methyl groups possess the largest range of chemical shift dispersion, with $^{13}$C and $^1$H resonances being predicted over a \( \sim 7 \) ppm and \( \sim 1 \) ppm range, respectively. Additionally, the predicted chemical shifts of Ala113 and Ala123, which reside within the FapFβ helical plug, are not overlapping with other alanine resonances. Furthermore, alanine residues are well distributed across the FapFβ β-barrel; this includes Ala256 that is proximal to a conserved “PTG” motif (FapF$_{257-259}$) (Figure 72B) (Rouse et al., 2017). In contrast, the predicted methyl group chemical shifts of Leu-δ2, Val-γ2, Thr-γ2, and Ile-γ2 demonstrate an abundance of inter- and/or intra-residue overlap that would complicate their downstream assignment and analysis. Additionally, these methyl groups, with the exception of Leu-δ2, are predicted to overlap with those of Ala-β and that may overcomplicate the downstream analysis of Ala113 and Ala123. On the other hand, the predicted chemical shifts for the Ile-δ1 methyl groups are found in an upfield region of the spectrum, with respect to the $^{13}$C dimension, and they do not overlap with Ala113 or Ala123. Furthermore, the predicted chemical shifts for Ile-δ1 methyl groups are not extensively overlapped, and they are well distributed across the structure of FapFβ (Figure 72B). For the above reasons, Ala-β and Ile-δ1 were methyl-labelled in this study.
5.3.10. NMR Studies of Methyl-Labelled FapFβ

The expression of \{U-[\(^2\)H\(^{15}\)N], (Ala-β, Ile-δ1)-[\(^{13}\)CH\(_3\)]\}-FapFβ (referred to as MFapFβ) was performed according to the protocols described in section 2.4.5.; the metabolic pathways of \textit{E. coli} that are associated with the various methyl-labelling strategies are summarised by Schütz and Sprangers (2020). Once expressed, MFapFβ was refolded and purified as previously described (using H\(_2\)O formulated buffers) and transferred into buffer comprised of 20 mM PO\(_4\), 150 mM NaCl, 0.1 % LDAO (protonated), pH 6.8, formulated in 100 % D\(_2\)O, prior to recording a \(^1\)H-\(^{13}\)C HMQC (methyl-TROSY) spectrum; this sample is referred to as “MFapFβL”. After, the sample was exchanged into NDB containing 0.5 % C8E4 as previously described, and subsequently transferred into buffer comprised of 20 mM PO\(_4\), 150 mM NaCl, 0.5 % C8E4 (protonated), pH 6.8, formulated in 100 % D\(_2\)O. Next, a methyl-TROSY spectrum was recorded; this sample is referred to as “MFapFβC.” The recording of a \(^1\)H-\(^{13}\)C HMQC spectrum for MFapFβL and MFapFβC enables us to compare the spectral qualities and differences between each MFapFβ sample. These results will be
used to determine whether the methyl labelling of FapFβ, and the methyl TROSY-strategy, improve the NMR spectral quality of FapFβ.

As well as displaying peaks corresponding to the FapFβ methyl resonances, the \(^1\text{H}-\text{\textsuperscript{13}}\text{C}\) HMQC (HMQC) spectra for MFapFβL and MFapFβC will also present peaks corresponding to the methyl groups of their respective detergents (detergent peaks) (Hiruma-Shimizu et al., 2015). In order to identify the detergent peaks on each HMQC spectrum, unlabelled FapFβL and FapFβC were obtained as previously described, and a natural abundance HMQC was recorded for both samples. The HMQCs corresponding to MFapFβL and unlabelled FapFβL are overlaid (Figure 73), and the HMQCs corresponding to MFapFβC and unlabelled FapFβC are overlaid (Figure 74); this method allows the detergent peaks on the HMQC, which is both free and FapFβ-bound, to be identified.

Figure 73: \(^1\text{H}-\text{\textsuperscript{13}}\text{C}\) HMQC spectrum of \{U-\textsuperscript{2}H\textsuperscript{14}N\}, (Ala-β, Ile-δ1)-\textsuperscript{13}CH\textsubscript{3}\}-FapFβ in 0.1 % LDAO. The \(^1\text{H}-\text{\textsuperscript{13}}\text{C}\) HMQC spectra for MFapFβL (blue) and unlabelled FapFβL (red) are overlaid in order to determine the spectral peaks that correspond to FapFβ or LDAO. Regions of the spectrum containing peaks that are postulated to correspond to Ala-β and Ile-δ1 are encircled green and purple, respectively. MFapFβL sample conditions: 175 µM, 303 K, 256 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe. Unlabelled FapFβL sample conditions: As described for MFapFβL; sample concentration: 100 µM.
Figure 74: $^1$H-$^{13}$C HMQC spectrum of {U-$^1$H,$^{15}$N}, (Ala-β, Ile-δ1)-[$^{13}$CH$_3$]-FapFβ in 0.5 % C8E4. The $^1$H-$^{13}$C HMQC spectra for MFapFβC (blue) and unlabelled FapFβC (red) are overlaid in order to determine the spectral peaks that correspond to FapFβ or C8E4. Regions of the spectrum containing peaks that are postulated to correspond to Ala-β and Ile-δ1 are encircled green and purple, respectively. MFapFβC sample conditions: 125 µM, 303 K, 256 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe. Unlabelled FapFβC sample conditions: As described for MFapFβC; sample concentration: 85 µM.

The HMQC spectrum of MFapFβL and MFapFβC are overlaid (Figure 75), indicating that the S/N and peak resolution for both HMQC spectra are improved compared to the non-deuterated $^{15}$N-FapFβL $^1$H-$^{15}$N HSQC spectrum, as well as the $^2$H-FapFβL and $^2$H-FapFβC TROSY spectra. Two regions of clustered peaks within each HMQC spectrum broadly correspond to the $^1$H-$^{13}$C resonances of Ala-β and Ile-δ1 methyl groups that are predicted by CH3Shift; their corresponding peaks are indicated in Figure 73 and Figure 74. A total of 20 FapFβ Ala-β peaks are expected on each HMQC spectrum corresponding to MFapFβ (assuming a symmetrical FapFβ oligomer; or the presence of monomers); approximately 19 Ala-β peaks could be identified on the HMQC spectra corresponding to MFapFβL and MFapFβC. Of note, the identification of peaks in both HMQC spectra at ~1.4 ppm ($^1$H), ~19 ppm ($^{13}$C) is stifled due to the presence of peak overlap in this spectral region. Nonetheless, the remaining Ala-β peaks are well dispersed. A total of 14 Iso-δ1 peaks are expected on the HMQC spectra for FapFβ (assuming symmetry or a monomer); approximately 14
Iso-δ₁ peaks could be identified on each of the HMQC spectra corresponding to MFapFβL and MFapFβC. The Iso-δ₁ peaks in both spectra are somewhat dispersed.

Figure 75: Comparison of the $^1$H-$^{13}$C HMQC spectra of MFapFβL and MFapFβC. The $^1$H-$^{13}$C HMQC spectra for MFapFβL (blue) and MFapFβC are overlaid (red). The arrow indicates a potential chemical shift perturbation between the two spectra, as described within the text. The sample details are described in Figure 73 and Figure 74, respectively.

Similar to the $^2$H-FapFβ TROSY spectra previously described, a number of peaks across both HMQC spectra are heterogeneous with regards to their broadness (intensity); the potential causes of this are previously described for the TROSY spectra recorded for $^2$H-FapFβ samples (section 5.3.8). For example, in the HMQC spectra corresponding to MFapFβL and MFapFβC, a number of relatively broad peaks are present, including one peak (at ~0.8 ppm ($^1$H), 12.5 ppm ($^{13}$C)) that appears to have broadened into two or three sub-peaks on both the HMQC spectra. This suggests that multiple residues are undergoing slow exchange on the NMR timescale (Kleckner and Foster, 2011). Resultantly, the total number of peaks corresponding to individual Ala-β or Ile-δ₁ methyl group resonances on each MFapFβ HMQC spectrum cannot be accurately quantified.

The majority of peaks between the MFapFβL and MFapFβC HMQC spectra (corresponding to MFapFβ) are somewhat overlapping, however, the majority of
peaks display minor differences in their chemical shifts. These differences could be due to the reasons previously described for the chemical shift differences observed between the $^2$H-FapFβ TROSY spectra (section 5.3.8). Of note, it is unclear whether the chemical shift (and peak intensity) of one postulated Ala-β peak is significantly altered between the MFapFβL and MFapFβC HMQC spectra, or whether one peak disappears and a new peak forms (the postulated peak shift is indicated by an arrow on the spectra in Figure 75). Both of the indicated peaks may correspond to an Ala-β or Ile-δ1 residue within a relatively flexible region of FapFβ, such as a β-strand loop, or potentially, the α-helical plug, due to the capacity for these regions to undergo relatively large conformational changes (Columbus et al., 2009).

Lastly, the MFapFβC sample sensitivity and stability were investigated. The MFapFβC sample was incubated at 20 °C for three days and a HMQC spectrum was subsequently obtained with 32 scans, instead of the 256 that were previous performed. The HMQC spectrum of MFapFβC obtained with 256 scans (HMQC-256) was overlaid with the HMQC spectrum of MFapFβC obtained with 32 scans (HMQC-32) after three days of sample incubation (Figure 76A). Additionally, samples of the MFapFβC soluble fraction were collected before and after the three-day incubation period in order to check for sample degradation by SDS-PAGE analysis. Overlaying the HMQC-256 and the HMQC-32 demonstrates that all of the peaks corresponding to MFapFβC are present and overlapping between both spectra. Firstly, this suggests that FapFβ remains stable over this incubation period, and that is confirmed by SDS-PAGE analysis (Figure 76B). Second, this indicates that the methyl-labelling and methyl-TROSY strategy applied in this study is of high sensitivity. Resultantly, the improved spectral line widths, S/N, and peak dispersion, enables the NMR study of FapFβ. This may include the use of a number of complex and multidimensional NMR experiments, such as those used to obtain NMR dynamics measurements, and that can be obtained within a reasonable time period (2-3 days) (Schütz and Sprangers, 2020; Törner et al., 2020).
Figure 76: Methyl-labelled FapFβ possesses a high spectral sensitivity. (A) The $^1$H-$^{13}$C HMQC spectra for MFapFβ obtained with 256 scans (blue) and 32 scans (after 3 days of incubation at 20 °C) (red) are overlaid. Each spectrum contains the same number of peaks (that correspond to FapFβ). MFapFβC sample conditions (32 scans): 125 µM, 303 K, 32 scans, performed on the Avance III 600 MHz spectrometer with cryoprobe. (B) SDS-PAGE gel analysis indicates that FapFβ (150 µM) in buffer containing 0.5 % C8E4 does not undergo degradation or precipitation over a three day period of incubation at 20 °C. The soluble fractions collected after day 0 and day 3 of sample incubation are presented.

5.3.11. Reconstitution of Refolded FapFβ into Nanodiscs

Membrane proteins are typically incorporated into detergent micelles for the study of MP structure and dynamics due to their capacity to extract, refold, and aid protein crystallisation (Buchanan, 1999; Loll, 2014). Therefore, detergents have the capacity to maintain the structural integrity of MPs, however, there are reports to indicate that detergents reduce or abolish MP dynamics and functional activity (Frey et al., 2017; Kofuku et al., 2014). In contrast, the transfer of a MP from a detergent environment to a lipid environment is reported to restore MP activity (Kofuku et al., 2014). Recently, the residue-specific dynamics of outer membrane protein X (OmpX; eight-stranded β-barrel protein) was characterised by solution NMR in both a DPC micelle environment and a DMPC nanodisc environment. The dynamics of OmpX were significantly increased (particularly its β-strands) within the nanodisc environment compared to the micelle environment. The authors of this study suggested that the
incorporation of MPs into micelles significantly compromises their inherent flexibility, unlike the nanodisc environment (Frey et al., 2017).

Nanodiscs are soluble membrane mimetics comprised of a phospholipid bilayer that is encircled by two rings of amphipathic membrane scaffold protein (MSP); the nanodisc structure is maintained by non-covalent interactions (Bayburt and Sligar, 2010). MSP is a genetically engineered protein that originates from apolipoprotein A-1, a protein that normally functions in cholesterol transport and lipoprotein metabolism (Bayburt and Sligar, 2010; Curtiss et al., 2006). The diameter and thickness of a nanodisc is highly customisable, allowing proteins of various dimensions to be incorporated; the length of the MSP is responsible for controlling nanodisc diameter, and the nanodisc thickness is determined by the acyl chain length of the lipid molecules incorporated (Schuler et al., 2013; Siuda and Tieleman, 2015). As well as their high solubility, nanodiscs are easily concentrated, monodisperse, and stable at high temperatures (Puthenveetil et al., 2017). In this study, a protocol for the incorporation of FapFβ into nanodiscs was established in order to enable future investigations of FapFβ to be carried out in the nanodisc environment, and that can be compared to the micelle environment.

In order to incorporate FapFβ (or another MP) into a nanodisc, FapFβ must first be solubilised and purified in buffer containing detergent, as previously performed. Subsequently, purified FapFβ is mixed with a phospholipid and MSP of choice, which are also solubilised in detergent. After mixing, the detergents are removed, resulting in the assembly of the nanodisc structure that incorporates FapFβ. The nanodisc assembly process is illustrated in Figure 7 (Ritchie et al., 2009; Yeh et al., 2018). In order to incorporate FapFβ into a nanodisc, there are various parameters that influence the success of nanodisc formation, and that require fine-tuning. This includes the length of the MSP, the choice of phospholipid, the lipid to MSP molar ratio, as well as the MSP to FapFβ molar ratio (Ritchie et al., 2009).
Figure 77: Nanodisc assembly is a multistep process. Initially, a membrane protein is solubilised in detergent and purified. Subsequently, the membrane protein is mixed with purified MSP proteins, as well as lipids (solubilised in detergent). In order to incorporate the respective membrane protein into a nanodisc, the detergent is removed. Figure modified from Yeh et al. (2018) under creative commons license CC BY 4.0.

Previous reports indicated the incorporation of OmpX and OmpA, two β-barrel OMPs that are of similar dimensions to monomeric FapFβ, into ~9.5 nm diameter (length) nanodiscs using MSP1D1 and the DMPC lipid (Hagn et al., 2013; Sušac et al., 2014). The diameter of a FapFβ trimer (length) spans up to ~7.5 nm, which is compatible with its insertion into MSP1D1-nanodiscs. The optimal DMPC:MSP1D1 ratio for OmpX and OmpA was reported at 1:80, and the ratios of OmpX or OmpA to MSP1D1 were 1:4 and 1:10, respectively; these ratios were used as a guide for the insertion of FapFβ into MSP1D1-nanodiscs (Hagn et al., 2013; Sušac et al., 2014). In an attempt to incorporate FapFβ into MSP1D1-nanodiscs, unlabelled FapFβ was expressed, refolded, purified, and transferred into NDB containing 0.5 % C8E4 as previously described. MSP1D1 was expressed and purified according to previously established protocols, with the addition of his6tag removal; the protocols are described in section 2.7. (Ritchie et al., 2009).

Before attempting to incorporate FapFβ into nanodiscs, a control study was performed in which the capacity for the MSP1D1, purified in this study, to form nanodiscs was probed. Briefly, a 1:80 ratio of purified MSP1D1 to DMPC was mixed, the detergent present in this sample was subsequently removed, and the sample was subjected to analytical SEC to analyse whether the sample is homo- or heterogeneous. The SEC UV trace indicates the presence of a single symmetrical peak at 13.3 ml (Figure 78A); SDS-PAGE analysis of the corresponding fractions indicate the presence of MSP1D1 (Figure 78B); of note, a species that is approximately 5 kDa
smaller than the main MSP1D1 species is present throughout all SEC fractions, and that likely corresponds to a degradation product of MSP1D1 that is commonly reported (Faas et al., 2018). A previous study demonstrates that the MSP1D1-nanodisc elutes off an S200 analytical column at ~13.7 ml, that is in broad agreement with the elution volume for the MSP1D1-DMPC complex obtained here, which suggests the formation of an MSP1D1-nanodisc (Hagn et al., 2013).

Figure 78: The insertion of FapFβ into MSP1D1-nanodiscs. (A) Overlaid SEC (UV 280 nm) traces of FapFβ alone (blue), empty MSP1D1-nanodiscs (green), and FapFβ-MSP1D1-nanodiscs (purple). (B) SDS-PAGE gel analysis of the specific SEC fractions collected for MSP1D1-nanodiscs. (C) SDS-PAGE gel analysis of the specific SEC fractions collected for FapFβ-MSP1D1-nanodiscs.

To provide further evidence for the formation of an MSP1D1-nanodisc, the sample was analysed by dynamic light scattering (DLS). In DLS (reviewed by Bhattacharjee 2016), a lightsource is applied to the sample and the resulting fluctuations in light scattering are detected; the light scattering fluctuations are due to particle movement, which is assumed to result from Brownian motion. The extent of light scattering correlates with the particle translational diffusion coefficient and the particle hydrodynamic radius, from which, the intensity weighted average hydrodynamic particle size (z-average) can be calculated. Additionally, a value for the polydispersity index (PDI) is calculated that quantifies sample polydispersity. The values for PDI are scaled between zero and one, with values closer to one indicating an increasingly polydisperse sample; for highly monodisperse samples, a PDI value of below 0.1 is expected (Bhattacharjee, 2016). DLS analysis of the postulated MSP1D1-nanodisc formed in this study is indicated to possess a z-average particle size of 9.5 nm (SD ± 0.1 nm) and a PDI of 0.074 (SD ± 0.010) (Figure 79). This indicates the formation of MSP1D1-nanodiscs, and that this sample is monodisperse (Hagn et al., 2013).
Next, to determine whether purified FapFβ could be incorporated into MSP1D1-nanodiscs, a 1:4:320 molar ratio (60 µM:240 µM:19.2 mM) of purified FapFβ:MSP1D1:DMPC was mixed. After sample mixing and detergent removal, the resulting sample was subjected to analytical SEC as described previously. In addition, a sample of FapFβ was refolded and purified, and exchanged into NDB containing 0.5 % C8E4, as previously described. This sample was also subjected to analytical SEC in order to compare the retention volume of FapFβ between the samples containing FapFβ (alone) and FapFβ-MSP1D1-DMPC (FMD). The SEC UV traces of the MSP1D1-nanodiscs, FMD, and FapFβ are overlaid (Figure 78A). Regarding the FMD sample, the SEC UV trace indicates the presence of a single symmetrical peak at 13.3 ml, the retention volume of which is consistent with that of the MSP1D1-nanodisc. Of note, a low intensity peak is present at ~8 ml that may correspond to an aggregation product of MSP1D1 or FapFβ. SDS-PAGE analysis indicates that MSP1D1 and FapFβ co-elute at 13.3 ml, suggesting the formation of a FapFβ-MSP1D1-nanodisc (FapFβ-ND) (Figure 78C). Furthermore, FapFβ elutes at 14.2 ml, which is 0.9 ml later than the elution volume of FMD, indicating that the FMD complex is a larger species than FapFβ (Hong et al., 2012). Subsequent DLS analysis of the postulated FapFβ-ND indicates a z-average particle size of 9.7 nm (SD ± 0.1 nm) and a PDI of 0.081 (SD ± 0.006) (Figure 79). This indicates the formation of FapFβ-NDs, and that this sample is monodisperse (Hagn et al., 2013). Although this study demonstrates the formation of FapFβ-NDs, it remains possible that empty-MSP1D1 nanodiscs were also formed. To remove empty-MSP1D1 nanodiscs from this sample, the FapFβ-ND
sample can be applied to Ni-IMAC column; FapFβ-ND will in theory bind to the column due to the presence of a his6tag at the N-terminus of FapFβ, and empty-MSP1D1 nanodiscs will flow through the column due to their lack of N-terminal his6tag, thus, FapFβ-ND and empty-MSP1D1 nanodiscs can be separated.

5.3.12. FapFβ Unfolds in the Presence of 4 M Urea

Single-channel conductance assays (SCCAs) have previously been used to distinguish between the open and closed state of FapF; both FapF and FapFβ were previously subjected to the SCCA (described within Rouse et al. (2017)). The basis of the SCCA is described by Zakharian (2013). Regarding the SCCA previously performed for full length FapF, the results indicated that the FapF pore is open, suggesting that the FapF helical plug does not occlude the FapF pore(s). In contrast, the SCCA for FapFβ did not exhibit any changes in the detected current under the applied voltages, suggesting that the FapFβ pore is closed. Interestingly, after the addition of 4 M urea into the recording chamber, the SCCA for FapFβ was repeated, and the results indicated a proportional increase in the detected current as the voltage applied was increased. These results matched those described for full length FapF. Rouse et al. (2017) suggested that the addition of 4 M urea into the SCCA recording chamber unfolds the FapFβ plug, releasing it from the FapFβ β-barrel lumen, and enables the flow of ions through the FapFβ pore; however, the effects of 4 M urea on the structure of FapFβ were not investigated (Rouse et al., 2017). Therefore, it is unclear whether the proportional increase in current detected in the presence of 4 M urea was due to the local unfolding of the FapFβ plug, or the global unfolding of FapFβ. This study sought to investigate the folding of FapFβ in the presence of 2 M and 4 M urea using solution NMR spectroscopy.

A purified 15N-FapFβC sample was prepared as previously described, and divided into three fractions; the first, second, and third fraction contain 0 M, 2 M, and 4 M urea, respectively. Of note, 8 M urea was dissolved into buffer containing 20 mM Tris-HCl, 300 mM NaCl, 0.5 % C8E4, pH 8.0, pH adjusted to pH 8.0, and the required quantity of this buffer was added to the second (2 M urea) and third fraction (4 M urea); the concentration of FapFβ for all samples is 120 µM. A 1H-15N HSQC spectrum was recorded for all three FapFβ samples in order to gain insight into the folding of FapFβ in 2 M and 4 M urea. The 1H-15N HSQC spectra corresponding to
each sample are overlaid (Figure 80). The HSQC spectra indicate that in the presence of 0 M urea, FapFβ displays peak dispersion downfield of 8.5 ppm, which is consistent with the presence of a folded domain. The addition of 2 M urea to the FapFβ sample results in a minor decrease in peak dispersion, and the addition of 4 M urea significantly reduces the peak dispersion downfield of 8.5 ppm; the loss of peak dispersion likely owes to the intermediate-exchange processes of the newly unfolded regions of FapFβ, as well as their uniform solvent exposure (de Oliveira and Silva, 2015; Kleckner and Foster, 2011; Kwan et al., 2011). This suggests that, in the presence of buffer containing 2 M urea, FapFβ remains folded, and the presence of 4 M urea results in the global unfolding of FapFβ. Of note, the intense peaks at 117 ppm (\(^{15}\)N) likely correspond to the NH\(_2\) groups of urea (de Oliveira and Silva, 2015). Therefore, the results of this study suggest that the presence of 4 M urea in the SCCA recording chamber results in the global unfolding (denaturation) of FapFβ, rather than the specific unfolding of the FapFβ helical plug.

![Figure 80: \(^1\)H-\(^{15}\)N HSQC spectra of FapFβ in the presence of 0, 2, and 4 M urea. The addition of 4 M urea destabilises the global conformation of FapFβ. The \(^1\)H-\(^{15}\)N HSQC spectra for FapFβ (in 0.5 % C8E4) recorded in the presence of 0, 2, and 4 M urea are overlaid. The addition of 4 M urea indicates a reduction in peak dispersion downfield of 8.5 ppm (\(^1\)H), suggesting that the global conformation of FapFβ is destabilised. All FapFβ sample conditions: 150 μM, 298 K, 100 scans (200 scans used for FapFβ in buffer containing 4 M urea), performed on the Avance III 600 MHz spectrometer with cryoprobe.](image)
5.4. Summary and Discussion

5.4.1. Optimising the Refolding Efficiency of FapFβ

Membrane protein refolding protocols are complex, with their efficiencies being dependent on various experimental parameters. One of these parameters relates to the concentration of the refolded MP protein (Schwarzer et al., 2017). The folding intermediates formed during the refolding process possess surface exposed hydrophobic patches that have the potential to associate and aggregate (Fink, 1998; Vallejo and Rinas, 2004). Therefore, higher MP refolding concentrations are expected to be more prone to aggregation than lower concentrations, due to the increased probability of inter-protein interactions. In addition, the concentration of detergent present in the refolding buffer (relative to its respective CMC) is also an important MP refolding parameter (Gall et al., 2003; Schwarzer et al., 2017). A sufficient detergent concentration (>CMC) in the refolding buffer is required to shield the hydrophobic surfaces of MPs; however, an excessive concentration of detergent may also destabilise MPs and therefore, reduce the refolding efficiency (Gall et al., 2003; le Maire et al., 2000; Nath and Rao, 2001; Otzen et al., 2009; Schwarzer et al., 2017; Yang et al., 2014).

This study demonstrates that the refolding efficiency of FapFβ was significantly higher at relatively low refolding buffer LDAO concentrations (1.0 %) rather than relatively high concentrations (up to 10.0 %). The binding of multiple (excess) LDAO molecules to the hydrophobic surfaces of FapFβ may destabilise its native state, or induce the formation of a less stable misfolded (or intermediate) FapFβ state, which leads to the aggregation and precipitation of FapFβ (Nath and Rao, 2001; Otzen et al., 2009; Yang et al., 2014). The refolding efficiency of FapFβ was also improved at higher refolding concentrations of FapFβ (50.0 µg/ml) compared to lower FapFβ refolding concentrations (12.5 µg/ml), which is somewhat consistent with previous OMP refolding studies (Schwarzer et al., 2017). Although lower MP refolding concentrations are generally optimal, the improved refolding efficiency at higher FapFβ concentrations is likely related to the molar ratio of FapFβ to LDAO; reducing the molar concentration of FapFβ (in this ratio) will increase the concentration of free LDAO (that is not bound to FapFβ), thus, increasing the likelihood of detergent induced MP instability as described above (Nath and Rao, 2001; Otzen et al., 2009; Yang et al., 2014).
FapFβ refolding concentrations above 50.0 µg/ml and NDB LDAO concentrations below 1.0 % were not explored in this study; further optimisation of these two parameters may further increase the refolding efficiency of FapFβ. For instance, one study demonstrated that the optimal refolding concentration of OmpW, OprG, and TodX is ~0.3 mg/ml. Additionally, the mixing rate of the refolding buffer and the temperature of the refolding buffer have previously been demonstrated to impact the refolding efficiencies of OMPs (Schwarzer et al., 2017). These variables were not investigated in this study, thus, their optimisation may further increase the refolding efficiency of FapFβ. Approximately 11 mg of refolded and purified FapFβ was obtained from 1 L of culture grown in LB media when FapFβ was refolded at 50.0 µg/ml in NDB containing 1.0 % LDAO. Considering that ~4 mg of FapFβ was required for the methyl-TROSY experiments performed in this study, the final yield of FapFβ obtained is sufficient for the cost-effective study of FapFβ using heteronuclear (methyl) NMR experiments.

5.4.2. Biophysical Characterisation of Refolded FapFβ

NMR analysis of refolded FapFβ indicates the presence of a folded domain, and its CD profile is consistent with that of native FapFβ (both the CD profiles indicating the predominant presence of β-structure), suggesting that refolded FapFβ adopts its native conformation. Additionally, SEC-MALS indicated that the average MW of refolded FapFβL is approximately 98 kDa, which approximates to the MW of of a FapFβ dimer in an LDAO micelle. However, the sample polydispersity, and the broad UV peak observed during SEC (preparative), suggests the presence of multiple FapFβ species that exist within an exchange equilibrium of FapFβ monomer-dimer-trimer states (Hong et al., 2012). These results are consistent with a previous report that demonstrates the presence of multiple FapFβC species that exist within an exchange equilibrium of FapFβ monomer-dimer-trimer states, and the relative abundance of these states is 3:2:1, respectively (Rouse et al., 2017).

5.4.3. FapFβ and the Single-Channel Conductance Assay

Although SCCAs were not performed for refolded FapFβ, which would enable us to determine whether refolded FapFβ possesses a blocked pore (occluded by the plug), the discussions of a previously reported SCCA corresponding to the native
FapFβ is questioned in this chapter. A SCCA was previously conducted for native FapFβ (by Rouse et al. (2017)), the results for which indicated that the addition of 4 M urea into the SCCA recording chamber enabled the flow of ions through the lipid bilayer (i.e. through an open state of FapFβ). The authors of this study suggested that this was due to the local unfolding of the FapFβ helical plug, and that was released from the FapFβ lumen in the presence of 4 M urea (Rouse et al., 2017). However, the results of this study suggest that FapFβ undergoes global unfolding in the presence of 4 M urea, rather than the specific unfolding of the plug. Before conclusions can be drawn as to the validity of the published SCCA, it should be noted that a previous report indicates that the stability (resistance to urea unfolding) of LeuT, a MP, is greater in a lipid environment compared to a detergent micelle (Sanders et al., 2018). Thus, further investigation is required to elucidate whether FapFβ remains stable in diphytanoylphosphatidylcholine (DPhPC) bilayers in the presence of 4 M urea.

5.4.4. The NMR Spectral Quality of FapFβ is Improved by Methyl-Labelling and Methyl-TROSY Strategies

High MW proteins are problematic for NMR studies due to their poor relaxational properties. The average MW of the refolded FapFβ-LDAO micelle is approximately 98 kDa, which exceeds the ~25 kDa size limit for conventional $^{15}$N/$^{13}$C triple resonance-based NMR experiments (Puthenveetil and Vinogradova, 2019). This was demonstrated by the poor quality (peak broadening) of the $^1$H-$^{15}$N HSQC spectrum obtained for $^{15}$N-FapFβL. In an attempt to improve the NMR spectral properties of FapFβ, $^2$H-FapFβ samples were prepared and $^1$H-$^{15}$N TROSY-HSQC experiments were performed using an 800 MHz NMR spectrometer. Together, these factors significantly increased the spectral peak resolution and S/N compared to the $^1$H-$^{15}$N HSQC spectrum corresponding to $^{15}$N-FapFβL (Puthenveetil and Vinogradova, 2019). However, approximately a third of the peaks were missing from the TROSY spectra, a number of peaks were overlapping, and the spectral sensitivity was insufficient for triple-resonance NMR experiments. In order to reduce the number of spectral peaks (to reduce peak overlap) and further improve the spectral sensitivity, FapFβ Ala-β and Iso-δ1 methyl groups were methyl-labelled in an otherwise perdeuterated sample, and $^1$H-$^{13}$C HMQC experiments were performed for the MFapFβ samples to take advantage of the methyl-TROSY effect (Rosenzweig and Kay, 2014; Schütz and
Sprangers, 2020). The methyl-TROSY spectra obtained for MFapFβ indicate the presence of most or all spectral peaks that are expected, and that are mostly dispersed. Importantly, the spectral sensitivity was significantly improved compared to the previous NMR spectra ($^1$H-$^{15}$N HSQC / $^1$H-$^{15}$N TROSY-HSQC) obtained. The spectral sensitivity of the combined methyl labelling and methyl-TROSY strategy enables the detailed study of FapFβ dynamics, exchange, and interactions, using solution-state NMR methods (Schütz and Sprangers, 2020). Also, due to the high-MW of the FapFβ-ND, which is estimated at 150 to 200 kDa, the NMR study of the FapFβ-ND will also require the methyl-labelling approach (or similar) that is described here (Hagn et al., 2013; Kerfah et al., 2015).

### 5.4.5. Methyl Resonance Assignment Strategies for FapFβ

The methyl labelling strategy utilised in this study is not compatible with the conventional methods used for performing backbone assignments in NMR, which is due to the lack of $^1$H, $^{15}$N and $^{13}$C probes (Clore and Gronenborn, 1994; Puthenveetil and Vinogradova, 2019). Various strategies can be employed to assign the methyl resonances of methyl-labelled proteins, as reviewed by Schütz and Sprangers (2020); those that are most applicable to FapFβ are described below. One method incorporates site-directed mutagenesis (SDM); in this method, the methyl-probe of interest is substituted to an alternative residue and a $^1$H-$^{13}$C HMQC is recorded (Amero et al., 2011; Schütz and Sprangers, 2020). By overlaying the spectra corresponding to the WT and mutant, the missing peak can be identified and unambiguously assigned (assuming the remaining methyl-resonances are not affected). However, SDM may induce non-native protein conformations, and it is both time-consuming and costly (despite the recent optimisation of this strategy) (Amero et al., 2011; Schütz and Sprangers, 2020). Another strategy used to assign methyl resonances involves the use of various programs that predict the chemical shifts of the protein methyl groups. Many programs, such as the MAP-XSII and FLAMEnGO, use swapping procedures to back-calculate structural restraints from an available crystal structure and compare them to experimentally derived restraints (through-space) (Chao et al., 2012; Schütz and Sprangers, 2020; Xu and Matthews, 2013). Experimentally derived restraints include those from 3D/4D NOESY spectra, and in some cases, paramagnetic relaxation enhancement (PRE) data (PRE reviewed in Clore and Iwahara (2009))
The assignment accuracy of these methods increases as more experimental restraints are provided, and the accuracy is reported to be over 90% in some cases (Chao et al., 2012). It should be noted that the SDM approach should also be incorporated into this strategy in order to experimentally confirm at least a small proportion of the automated assignments (Schütz and Sprangers, 2020). On the other hand, it is worth noting that the CH3Shift program utilised in this study does not incorporate any experimentally derived restraints for the prediction of methyl group chemical shifts (Sahakyan et al., 2011). Therefore, this program was not used to assign any FapFβ methyl resonances due to the reported inaccuracies of CH3Shift, as well as the inaccuracies of the FapFβ crystal structure from which the methyl resonances are back-calculated (Nerli et al., 2018; Rouse et al., 2017; Sahakyan et al., 2011).

5.4.6. Future Work: NMR and Beyond

The methyl labelling strategy described in this study enables FapFβ (in detergent micelles or MSP1D1 nanodiscs) to be subjected to various NMR experiments that have the potential to yield a plethora of insight into the dynamics and exchange of FapFβ, the factors that control the postulated FapF gating mechanism, as well as its potential interactions to the Fap proteins (Rosenzweig and Kay, 2014; Rouse et al., 2017; Schütz and Sprangers, 2020). For instance, a recently published study used a novel sonication of lipid vesicles for MS (SoLVe-MS) approach to demonstrate that FapF is able to interact with FapC and/or FapE, however, the exact binding site(s) of FapF were not revealed (Chorev et al., 2018, 2020). NMR experiments can not only be used to investigate the FapFβ interaction site(s) to the Fap substrates, but they can also inform on the exchange processes and dynamics of FapFβ that occur before, during, and after an interaction has taken place; these results can be compared with studies of full length FapF to provide insight into Fap biogenesis (Rosenzweig and Kay, 2014; Schütz and Sprangers, 2020; Shi and Kay, 2014; Wiesner and Sprangers, 2015). Furthermore, NMR studies can be expanded to include the use of paramagnetic moieties that have the potential to further contribute to our understanding of these processes (Brath et al., 2015; Clore and Iwahara, 2009; Schütz and Sprangers, 2020). Various FapFβ mutants can also be examined by NMR to determine their effects on the structure, dynamics, and interactions of FapFβ (Rosenzweig and Kay, 2014). This
includes the mutagenesis of residues that are reported to stabilise the FapFβ plug within the FapFβ lumen (Figure 7D), as to determine which of these residues affect the lability of the plug, and that would contribute insight into the FapF gating mechanism (Rouse et al., 2017). Additionally, mutagenesis of the conserved “PTG” motifs may yield insight into their function(s), if any, as one study postulates that these motifs may tether Fap substrates (in the ES) to FapF (Figure 7E) (Rouse et al., 2017). The effects of environmental factors such as lipids and pH on the pore gating and dynamics of FapFβ can also be investigated (Liko et al., 2018). Also, NMR studies enable the structure and dynamics of FapFβ to be compared between the detergent micelle and nanodiscs environments (Frey et al., 2017). The results of which, if shown to differ significantly, may have implications for the experimental design of future FapF experiments (Hagn et al., 2018). In addition, it might be possible to apply the methyl labelling strategy described for FapFβ in this study, or similar, to the study of full length FapF by NMR, for which similar or identical NMR experiments to those described above can be performed. Transferring assignments for FapFβ to full length FapF may also be possible, and this would facilitate the assignment of the FapF TMD in equivalent full length FapF NMR spectra (Sprangers and Kay, 2007). The combined NMR studies of FapFβ and full length FapF may enable the postulated pore gating mechanism, and function(s) of FapF, to be elucidated. As previously described, these studies may also contribute insight into the dynamics and pore gating mechanisms of the T5a ATs due to their structural homology (Rouse et al., 2017).

Beyond the NMR-based methods described above, cryo-EM represents a suitable technique for solving the structure of the >150 kDa full length FapF (in detergent micelles or nanodiscs), both in isolation, and in complex with various Fap substrates such as FapD (Kieuvongngam et al., 2020; Nogales, 2016). Furthermore, the recently established SoLVe-MS approach can be utilised to investigate near-native interactions between the Fap proteins (Chorev et al., 2018, 2020). Therefore, by studying various Fap mutants, SoLVe-MS has the potential to be used to investigate the specific regions of Fap proteins that are required for Fap protein interactions, as well as elucidate the functions of Fap proteins (Chorev et al., 2018, 2020). In order to obtain a complete understanding of the conformation(s) of full length FapF, and its interactions with Fap proteins such as FapD, a multidisciplinary approach that incorporates techniques such as NMR, cryo-EM, and native MS is required. An
impressive example of the multidisciplinary approach includes the study of protein folding on the ribosome. Cryo-EM was used to analyse the ribosome structure, and solution-state NMR was used to probe the structure and dynamics of the ribosome-bound nascent polypeptide chain (Waudby et al., 2013). Therefore, the results from such a multidisciplinary Fap study will provide further insight into the open, closed, and intermediate states of FapF, as well as the functions of various Fap proteins, and that will contribute towards the elucidation of the FapF translocation mechanism. Furthermore, the insight obtained from these experiments may be valuable for the design of FapF/OMP inhibitors, or the engineering of FapF/OMPs, which have potential use within health and biotechnology, respectively (Carter and Hussain, 2017; Moroni et al., 2014; Oxenoid and Chou, 2016; Romero et al., 2013).
Chapter 6 - Concluding Remarks

6.1. The FapF N-Terminal Domains

Prior to this research, insight into the structure of the NTDs of FapF was lacking. The biophysical data presented for FapF27-64 in this thesis reveals the presence of a stable trimeric coiled-coil, which concurs with previous evidence that indicates that FapF is natively trimeric (Rouse et al., 2017). The formation of a trimeric coiled-coil was later confirmed by the obtainment of crystallographic data for the FapFcc (obtained by Dr Rouse), which indicated that the FapFcc exists as a parallel trimeric coiled-coil. However, the refinement statistics for the initial FapFcc model were unsatisfactory, due to ambiguity in the positions of the Tyr44 sidechain density, and remodelling was required (personal communication with Dr Sarah Rouse) (Rouse et al., 2018b). Intriguingly, a previous report on the structure of hSP-D, a parallel trimeric coiled-coil, had indicated that a single tyrosine side chain was asymmetrically buried (Håkansson et al., 1999). Subsequently, this study aimed to determine whether a single Tyr44 side chain of FapFcc is also asymmetrically buried, with aim to improve the refinement statistics for the FapFcc model. The 2D ¹H-¹H NOESY NMR spectrum for FapFcc indicated that a single tyrosine sidechain is asymmetric, and that was postulated to be buried within the coiled-coil, akin to hSP-D (Håkansson et al., 1999). The NMR data was relayed to Dr Sarah Rouse, who subsequently remodelled FapFcc to include the burial of a single tyrosine side-chain, which led to improved refinement statistics for the structural model. Also, molecular simulations performed by Dr Rouse concurred with these findings. The FapFcc model was subsequently published in the PDB (PDB ID: 6FUE) (Rouse et al., 2018b). The structure of the FapFcc reveals the presence of a hydrophobic cleft, which may represent a substrate binding site, as well as a conserved glutamine layer, which is postulated to enable conformational changes within FapFcc (Rouse et al., 2018b). Both these structural features suggest that the FapF coiled-coil is involved in the regulation of a FapF gating mechanism via interaction to other Fap proteins; FapF-Fap protein interactions may modulate the pore-open versus pore-closed configuration of FapF, potentially via an allosteric mechanism (Papaleo et al., 2016; Rouse et al., 2018b). However, further evidence is required to demonstrate the regulation, if any, of the FapF gating mechanism. Also, despite the knowledge that FapF exists as a
trimer, the question still remains as to why FapF is a trimer, as its functional importance, if any, is yet to be elucidated.

The FapFcc structure also indicates the presence of a conserved RhxxhE motif that is implicated as a driver for coiled-coil trimerisation (Kammerer et al., 2005). The biophysical studies of FapFcc mutants, R57A and R57E, indicated that each mutant disrupted the trimerisation of FapFcc. This work highlights the importance of the RhxxhE motif for the trimerisation of FapF27-64 (Rouse et al., 2018b). An expansion of these studies to include additional RhxxhE motif related mutants may lead to an enhanced understanding of the factors that drive the specific coiled-coil oligomeric states, as well as improve the accuracy of coiled-coil oligomeric state prediction algorithms (Kammerer et al., 2005).

Lastly, this study presented preliminary experimental evidence to indicate that a short C-terminal extension of the FapF27-64 construct, FapF65-81, is comprised of random coil. Furthermore, FapF65-81 may undergo intermediate exchange, and/or may be subject to degradation. Further study is required to clarify the structure, flexibility, and stability of FapF65-81. Also, a PPII is predicted in this region, which could be functionally relevant to FapF (Adzhubei et al., 2013). Although the role of the FapFL for Fap substrate translocation is currently unknown, it was earlier hypothesised that the FapF linker domain could transiently bind Fap substrates in order to guide their translocation to the FapF TMD for export into the ES. Also, the binding of Fap proteins to the FapF NTDs may trigger allosteric signals that propagate via the FapFL to the FapF TMD, which could regulate the postulated pore gating mechanism of FapF (Papaleo et al., 2016; Ma et al., 2011). In addition, the accessibility of the FapFL to trypsin, as well as the presence of a GG-like motif, enables us to postulate that the FapFL may serve as a recognition site for FapD binding and cleavage, which may result in the formation of a FapFβ-like conformation that results in the pore closure of FapF (Dirix et al., 2004; Rouse et al., 2017). However, further research is required to elucidate the exact role of the FapF N-terminal domains in Fap substrate translocation, as well as the interactions and translocation mechanism(s) of FapF.

6.2. FapD

The structure and interaction partners of FapD, a predicted C39-peptidase, were unknown prior to this investigation (Rouse et al., 2017). A previous study had
indicated the co-evolution of fapD and fapF that suggested their functional co-dependence (Rouse et al., 2018a). In order to investigate this further, FapD was recombinantly expressed and purified with aim to solve its structure, and interactions between the FapF N-terminal domains and FapD were probed. However, prior to biophysical analysis, the stability of FapD was monitored due to the presence of multiple GG-like motifs within its sequence, which could lead to autoproteolysis (Dirix et al., 2004; Furgerson et al., 2008). Interestingly, the FapDWT N-terminus was found to undergo specific degradation of ~2-3 kDa over an incubation period of five days at 20 °C; the loss of 2-3 kDa at the FapDWT N-terminus results in a truncation up to the approximate location of a GG-like motif (FapD39-49), which may represent a FapD cleavage-site (Dirix et al., 2004; Dueholm et al., 2010). The proteolytic stability of FapDT concurred with this hypothesis, as FapDT did not demonstrate any detectable degradation over a five-day incubation period. Members of the clan CA of the cysteine peptidases are commonly demonstrated to undergo autoproteolysis in order to “activate” their proteolytic activity under specific environmental conditions; this mechanism likely evolved to minimise their non-specific proteolytic activity in order to protect the host cell (Khan and James, 1993; Turk et al., 2012). Therefore, FapD may also undergo autoproteolysis under certain cellular conditions in order to activate its postulated proteolytic activity. However, further research is required to elucidate whether FapD is a C39-peptidase family member. To provide insight into the function of FapD, future studies should employ methods such as HPLC/MS to investigate whether FapD is able to cleave peptides containing the GG-motif, and variants, akin to previous studies for the C39-peptidases (Bobeica et al., 2019; Furgerson et al., 2008). These studies should include peptides that contain the various GG-like motifs present within FapA-F. In addition, the binding parameters between FapD and the GG-motifs / GG-like motifs should be probed, using techniques such as ITC.

In this study, attempts to solve the structure of FapDWT and FapDT by X-ray crystallography or NMR spectroscopy were hindered by their high aggregation propensities at sample concentrations above 100 µM. Future studies hoping to achieve sample concentrations above this should screen the effects of various buffer components, at various concentrations, on the stability (aggregation propensity) of FapD (Senisterra and Finerty, 2009; Vedadi et al., 2006).
Interactions between FapD and the FapF periplasmic N-terminus were also probed in this thesis. SEC was used to demonstrate that FapDWT and FapDT independently associate to CCExt, and the results of ITC revealed a dissociation constant in the low micromolar range for the interaction between FapDT and CCExt. Based on these results, it was suggested that FapD may function akin to the C39-peptidase domain of PCAT1 (Kieuvongngam et al., 2020). Structural studies of the FapD-CCExt complex were again hindered by the high-aggregation propensity of FapD. To circumvent this problem, future studies should attempt to solve the structure of FapD in complex with full length FapF using cryo-EM (Nogales, 2016). The cryo-EM-based approach has previous been utilised to study the structure of PCAT1 in complex with a substrate (Cta) (Kieuvongngam et al., 2020). The cryo-EM approach would alleviate the problems associated with the high-aggregation propensity of FapD, as low sample concentrations are required (Grassucci et al., 2007). Further research into the interaction(s) between FapD and full length FapF will substantially increase our insight into the molecular mechanisms involved in Fap biogenesis, and that insight may translate over into the studies of the AMS/PCATs.

6.3. FapFβ

The crystal structure of the FapF transmembrane domain was previously solved (Rouse et al., 2017). In order to provide insight into the postulated FapF pore gating mechanism, atomic-level dynamics- and exchange-based insight into FapF and FapFβ is required, and NMR spectroscopy is well suited to providing this information (Markwick et al., 2008; Wiesner and Sprangers, 2015). However, the study of high-MW proteins by NMR spectroscopy is a costly endeavour, requiring high quantities of isotopically labelled protein (O’Brien et al., 2018). In order to produce MPs in large quantities, protein refolding protocols are commonly utilised (Noinaj et al., 2016). MP refolding protocols are complex, often requiring the optimisation of various parameters in order to produce successfully refolded protein (Schwarzer et al., 2017). In a previous study, a refolding protocol for FapF was outlined by Hawthorne (2016), and this was used as a template for the refolding of FapFβ. In this study, the refolding protocol was subsequently optimised by adjustment of the refolding concentration of FapFβ, as well as the concentration of LDAO in the refolding buffer. The resulting yield of FapFβ was sufficient for heteronuclear-based NMR studies.
In order to study FapFβ by solution-state NMR spectroscopy, a complex isotopic labelling strategy is required. The conventional NMR approach that employs $^{15}\text{N}$ and $^{13}\text{C}$ backbone labelling strategies are limited by the high MW of FapFβ that results in broad linewidths (Kwan et al., 2011). Nonetheless, the combined Ala-β and Ile-δ1 methyl-labelling strategy later employed in this study demonstrated high spectral sensitivity as expected, which enables downstream NMR studies on FapFβ and its various mutants (Schütz and Sprangers, 2020). Of note, the assignment of methyl-NMR spectra likely requires the use of highly accurate chemical shift prediction algorithms, for which a crystal structure (FapFβ) and experimentally derived restraints (NOESY/PRE data) are required (Chao et al., 2012; Xu and Matthews, 2013). Furthermore, SDM of specific methyl-containing residues can be performed to unambiguously assign specific residues, assuming that the mutation does not disrupt the global conformation of FapFβ (Amero et al., 2011).

Within this project, a protocol for the transfer of FapFβ from a detergent-micelle environment to MSP1D1 nanodiscs was established. Unlike the micellar environment, the nanodiscs environment is reported to enable additional, native-like membrane protein dynamics, akin to the native lipid bilayer (Frey et al., 2017; Kofuku et al., 2014). Therefore, future research into the dynamics- and exchange-based motions of FapF should preferably be studied within the nanodisc environment.

Overall, this work enables future studies to investigate the dynamics- and exchange-based motions of FapFβ under various experimental conditions. Beyond its potential contribution to the elucidation of the FapF gating mechanism, this work may also provide further insight into secretion mechanisms of other OMP transporters, such as those of the type 5a secretion systems (Rouse et al., 2017; van Ulsen et al., 2014). Furthermore, the protocols presented for the refolding and incorporation of FapFβ into MSP1D1 nanodiscs can potentially be applied to full length FapF, for which a refolding protocol already exists (FapF of P. aeruginosa PAO1) (Hawthorne, 2016). The methyl-NMR spectra presented in this study may also aid the future assignment of full length FapF Ala-β and Iso-δ1 methyl-NMR spectral peaks.

6.4. Final Remarks

Overall, the results from this thesis have advanced our structural and functional understanding of the Fap secretion system, and they provide a solid foundation for the
growth of future research in this field. To propel our insight into the Fap system, a multidisciplinary approach, which incorporates techniques such as cryo-EM, NMR spectroscopy, and native MS is required.

Ultimately, elucidating the structure and function of FuBA secretion system components will enable the modulation of bacterial amyloid formation, which will have vast implications within medicine, industry, and biotechnology (Galie et al., 2018; Knowles and Buehler, 2011; Van Gerven et al., 2018).
References


PyMOL. The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.


Appendix

A.1. Copyright Permissions I

The following figures (modified/unmodified – as stated) were licensed with a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Web link: https://creativecommons.org/licenses/by/4.0/

Figure 1 (modified). Obtained from: Sonesson et al. (2017)
Figure 3A (modified). Obtained from: Zeng et al. (2015)
Figure 5 (unmodified). Obtained from: Rouse et al. (2018a)
Figure 12 (modified). Obtained from: Kieuvongngam et al. (2020)
Figure 13 (unmodified). Obtained from: Kieuvongngam et al. (2020)
Figure 14 (modified). Obtained from: Kieuvongngam et al. (2020)
Figure 15A (modified): Obtained from: Truebestein and Leonard (2016)
Figure 77 (modified): Obtained from: Yeh et al. (2018)

A.2. Copyright Permissions II

Figure 15B (Coiled-coil formation is driven by the specific positioning and spacing of hydrophobic and polar residues within α-helices) was obtained from:
DOI: https://doi.org/10.1039/B914339B

The copyright holder of this work: © Royal Society of Chemistry 2020.
Permission Requested on 10 March 2020; permission to use this figure was granted in the form of a written permission (Figure 81).

Many thanks for sending the permissions request below. The Royal Society of Chemistry hereby grants permission for the use and adaptation of the material specified below in your thesis. Please note that if the material specified below or any part of it appears with credit or acknowledgement to a third party then you must also secure permission from that third party before reproducing that material.

Please ensure that the published article carries a credit to The Royal Society of Chemistry in the following format:

[Original citation] - Reproduced by permission of The Royal Society of Chemistry

and that any electronic version of the work includes a hyperlink to the article on the Royal Society of Chemistry website.

Best wishes,

Chloe Szobrat
Contracts and Copyright Executive
Royal Society of Chemistry
Thomas Graham House
Science Park, Milton Road
Cambridge, CB4 0WF, UK
Tel: +44 (0) 1223 488829
www.rsc.org

Figure 81: Copyright permission. Copyright permission for use of Figure 15B in this thesis by Apostolovic et al. (2010) is shown.