Structural biology of bacterial functional amyloid formation

William Hawthorne

Imperial College London

Department of Life Sciences

Thesis submitted for the degree of Doctor of Philosophy

Abstract

Amyloids are proteinaceous aggregates best known for their role in degenerative diseases involving protein misfolding. Research into amyloid has intensified in recent times due to its prominence in many debilitating human diseases and limited understanding of the causes. The discovery of functional amyloids in a broad range of species has enhanced our understanding of amyloid, of these the curli system of *E. coli* has been extensively studied, in this system CsgC was identified as a potent inhibitor of amyloid. An additional protein was discovered in some curli operons in other species termed CsgH and warrants further study. A morphologically similar but genetically distinct bacterial functional amyloid system was identified in *Pseudomonas* encoded by the *fap*ABCDEF operon and termed amyloid-like fibres (Alf). The study of functional amyloid has the potential to provide insights into how amyloid can be controlled.

The aims of this thesis were to investigate the novel functional amyloid system of *Pseudomonas* with a view to structural and functional characterisation of the individual components. The structure and function of the CsgH protein were also studied by nuclear magnetic resonance (NMR) and the ThioflavinT (ThT) amyloid fibrillation assay. Constructs were produced for all the Alf proteins and the more structured components, FapD and FapF, were optimised to produce constructs for structural study. The structure of CsgH was solved successfully using NMR and showed that the protein shared a similar tertiary structure to CsgC. The function of the CsgH was shown to be similar to CsgC inhibiting amyloid formation by CsgA at substoichiometric concentrations. Mutagenesis, ThT assay and NMR were used to show that CsgH and CsgA interact and that several charged residues have an important role in function. It was also interesting to note that CsgH was capable of inhibiting amyloid formation by the FapC amyloid protein of *Pseudomonas*.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Professor Steve Matthews for supervising my PhD, encouraging me and guiding my research.

Dr Jonathan Taylor and Dr James Garnett for their guidance, advice and patience over the course of my time in the lab, where they taught me most of what I know.

Dr Sarah Rouse for her work on the crystallisation of FapF, support for my coffee drinking habits and her ability to create a fun social atmosphere for the group.

Dr Sebastian Lambert for his work on crystallisation of FapF and for keeping things interesting in the lab during his PhD.

Dr Yingqi Xu who provided inestimable help with collecting, processing and analysing my NMR data.

Our previous CFNMR facility managers: Dr Jan Marchant and Dr Pete Simpson who both provided advice and assistance during the course of my experiments.

My students: Lea Sefer, Melissa Cheung, Yue Zhang and Rosemary Wenman; all of whom showed impressive determination and seem destined to future success.

To my fellow PhD students Rhian Jones and Xeni Miliara who provided great companionship to me.

Grace Harrison Ronge, for her long suffering endurance over the course of my studies over the last seven years.

My father Captain Michael Hawthorne for his support throughout my life.

My teachers, lecturers, family and friends who all helped nurture my interest in science and who have all contributed in some way to my modest success.

Abbreviations

Aβ: Amyloid-β

AD: Alzheimer's Disease

Alf : Amyloid-like fibres

APS: Ammonium persulphate

ATP: adenosine triphosphate

βOG: β-Octyl-Glucoside

BSA: Bovine Serum Albumin

C8E4: Tetraethylene Glycol Monooctyl Ether

CD: Circular Dichroism

CHAPSO: 3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxyl-1-propanesulfonate

CJD: Creutzfeldt-Jakob disease

CMC: critical micelle concentration

CTH: C-terminal poly-histidine tag

dCTP: deoxycytidine triphosphate

dGTP: deoxyguanosine triphosphate

DM: Decyl Maltoside

DMPC: 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine

DMSO: Dimethyl sulfoxide

DSF: Differential Scanning Fluorimetry

DTT: Dithioreitol

EDTA: Ethylenediaminetetraacetic acid

gDNA: genomic DNA

GST: Glutathione S-Transferase

HMQC: Heteronuclear multiple quantum coherence

HSQC: Heteronuclear single quantum coherence

IMAC: Immobilised Metal Affinity Chromatography

IG: Immunoglobulin

IPTG: Isopropyl β-D-1-thiogalactopyranoside

LB: Lysogeny Broth

LCP: Lipidic Cubic Phase

LDAO: n-dodecyl-N,N-Dimethylamine-N-Oxide

MBP: Maltose Binding Protein

MWCO: Molecular Weight Cut-off

NMR: Nuclear Magnetic Resonance

NOESY: Nuclear Overhauser Effect Spectroscopy

NOESY-HSQC: Nuclear Overhauser Effect Spectroscopy Heteronuclear Single Quantum Coherence

NTH: N-terminal poly-histidine tag

OD_{600}: Optical Density at 600 nm

PBS: Phosphate Buffered Saline

PCR: Polymerase chain reaction

PMSF: phenylmethanesufonylfluoride

PrP: Prion Protein

PVDF: Polyvinylidene fluoride

SDS: Sodium dodecyl sulfate

SDS-PAGE: Sodium dodecyl sulfate- Poylacrylamide Gel Electrophoresis

SEC-MALS: Size Exclusion Chromatography Multiple Anomalous Light Scattering

TAE: Tris-acetate-EDTA

TB: Terrific Broth

TEM: Transmission Electron Microscopy

TEMED: Tetramethylethylenediamine

Tm: melting temperature

TOCSY: Total Correlation Spectroscopy

ThT: ThioflavinT

TRCM: Thermophilic Rice Compost Metagenome

TROSY: Transverse Relaxation Optimised Spectroscopy

TRX: Thioredoxin

Declaration of Originality

I hereby declare that the research for, and authorship of this thesis are my own work unless otherwise stated; and that the sources used for this work have been appropriately referenced.

To Grace

1 Introduction

1.1 Amyloid

1.1.1 History and Identification of Amyloid

The name amyloid comes from the Latin word for starch amylum, this was the result of the early mistaken identification of amyloid as a starch deposit based on iodine staining [**1, 2, 3**]. This misapprehension continued until 1859 when it was shown that amyloid deposits were actually proteinaceous aggregates [**4**]. Traditionally amyloids had been viewed as amorphous aggregates composed primarily of beta-strand and they have been identified by their histological properties such as binding to the dyes ThioflavinT (ThT) or Congo Red [**2, 3**] (**Figure 1.1**). ThT exhibits a change in fluorescence in the presence of amyloid fibres as the dye binds the β-strand structure which can be detected as an increase in intensity around 500 nm [**5, 6**]. Congo Red also binds to amyloid where it has been described as producing apple-green birefringence under polarised light [**7**], however other colours can often be seen and the spectra cannot be explained simply be birefringence, this has led to the description that Congo Red staining produces anomalous colours [**8**]. Congo Red has also been shown to bind various other factors including sugars [**9**]. More recently a broader definition has been used based on the, now better understood, biophysical properties of amyloid [**10**]; this biophysical definition of amyloid extends the range to include proteinaceous deposits which don't show traditional staining but do possess the characteristic fibrous cross-β structure [**10**].

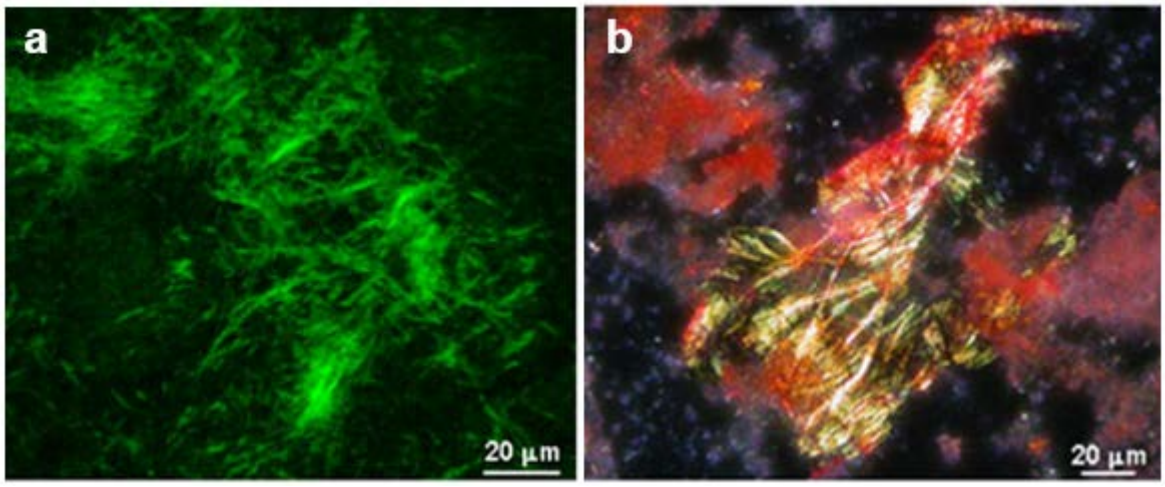


Figure 1.1: Histological Staining of Tissues for Amyloid. a, Amyloid deposits stained with Congo Red and viewed under polarised light. **b**, Amyloid deposits stained with ThioflavinT and visualised by fluorescence with excitation 488 nm, and emission in the 510-600 nm range. Edited version of figure from del Mercato et al. [**11**] which showed staining of amyloid-like poly(ValGlyGlyLeuGly) fibrils.

1.1.2 The Cross-β Structure of Amyloid

Structural studies have characterised amyloids as structurally ordered, non-branching fibrils which are rich in β-sheet [**10**]. Early work by Cohen and Calkin using electron microscopy showed that all the types of amyloid formed fibrous non-branching structures [**12**]. The structure of amyloid fibrils was further probed by Eanes and Glenner using synchrotron x-ray diffraction, which produced a characteristic cross-β diffraction pattern leading to the conclusion that amyloid fibres had an ordered cross-β structure, where the β-strands lie perpendicular to the long axis of the fibril [**13, 14**] **(Figure 1.2)**. A wide range of proteins are now known to form amyloid *in vivo* and *in vitro*, because of the broad range of proteins now known to be capable of forming amyloid and the fact the amyloid structure, as observed by Professor Christopher Dobson, "does not depend primarily on highly evolved side-chain interactions, but rather on universal physical and chemical characteristics that are inherent in the nature of all polypeptide molecules" [**15**], it has been suggested that the structure of amyloid represents a basic low-energy protein structure available to many different polypeptides **[15, 16, 17].** Although no common sequence motifs have been identified between amyloidogenic proteins, intrinsic disorder or unstable structure in amyloidogenic domains is common **[18]**. Despite the low sequence identity and lack of clear motifs amyloids can be loosely grouped based on the residues which appear to be important for aggregation for example glutamine-rich proteins such as Huntingtin and hydrophobic such as amylin and Alzheimer's β-protein **[18, 19, 20, 21]**. Study of the structural nature of amyloids has been advanced by the discovery of bacterial functional amyloids, particularly, the Curli system of *Escherichia coli* (*E. coli*), which has been used as a model for functional amyloid formation **[22, 23]**.

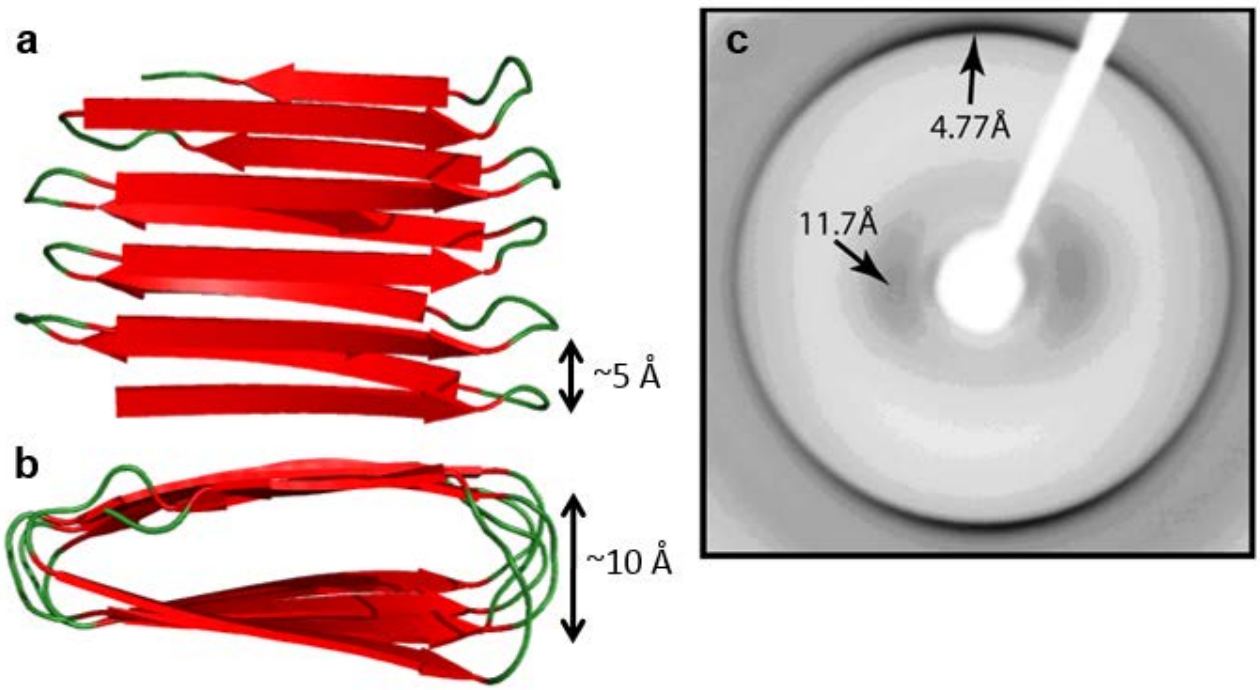


Figure 1.2: General Cross- β Structure of Amyloid. Viewed **a**, The Side and **b**, The Top Cartoon Diagram, β-strands are shown in red, loops in green. Based on model structure provided by Dr Jonathan Taylor (Unpublished) **c**) X-ray diffraction pattern of B2-microglobulin amyloid fibres (Figure From Magdalena I. Ivanova et al. 2004 [**212**]), showing characteristic diffraction pattern of amyloid with ~5 Å meridional and ~10 Å equatorial diffraction, the related distances in the model structure are indicated in A and B with arrows. [**13, 14**]

1.2 Amyloids and Proteopathic Human Diseases

1.2.1 Overview

Amyloid formation has been observed in a vast range of human diseases (**Table 1.1**) including several important neurodegenerative diseases such as Alzheimer's, Huntington's and Parkinson's. These diseases involve the misfolding of human proteins and are therefore proteopathies [**18, 24**]. Generally these involve a process where specific soluble, folded, functional protein monomers are converted into insoluble, non-functional, ordered, fibrillar aggregates. The discovery that amyloid diseases involved the aggregation of proteins into fibrous deposits was an important step in the study of amyloid diseases. Although the fibrous amyloid deposits both intracellular and extracellular certainly can contribute to cell death and dysfunction evidence suggests that soluble oligomeric assemblies of the amyloid forming proteins are sufficient to cause cellular dysfunction before the appearance of mature amyloid deposits [**25, 26**]. This is supported by the observations that the amyloid load does not correlate strongly with neurological dysfunction in Alzheimer's disease, that amyloid deposits often localise separately to sites of neuron loss and that deposits can develop in individuals who have no evidence of neuron damage [**26**].

Table 1.1: Human Diseases Putatively Involving Amyloid. A non-exhaustive list of diseases which have been associated with amyloid formation in humans. **[27, 28, 29, 30, 31, 32, 33, 34]**

Disease	Amyloid-forming Protein
Alzheimer's Disease	Beta Amyloid
Huntington's Disease	Huntingtin
Parkinson's Disease	Alpha-synuclein, Beta Amyloid
Atherosclerosis	Apolipoprotein AI
Familial Amyloid Polyneuropathy	Transthyretin
Finnish Amyloidosis	Gelsolin
Cerebral amyloid angiophathy	Beta Amyloid
Diabetes mellitus type 2	Islet Amyloid Polypeptide (Amylin)
Rheumatoid Arthritis	Serum amyloid A
Hereditary non-neuropathic systemic amyloidosis	Lysozyme
Systemic AL amyloidosis	Immunoglobulin light chain AL
Medullary carcinoma of the thyroid	Calcitonin
Prolactinomas	Prolactin
Dialysis related amyloidosis	Beta 2 microglobulin
Creutzfeldt-Jakob Disease	PrP

1.2.2 Alzheimer's Disease

Alzheimer's Disease (AD) is a major cause of dementia, accounting for almost 70 % of cases [**35**] and is the 6th most common cause of death in the United States **[36]**. AD is a chronic neurodegenerative disease which generally starts in old age (>65 years), the disease progresses gradually generally from minor short term memory loss to loss of bodily functions and finally to death. AD is characterised by the accumulation of both amyloid-β (Aβ) into senile plaques (**Figure 1.3**) and hyperphosphorylated tau into neurofibrillary tangles. The molecular mechanism underlying AD is still unclear with several theories, including protein misfolding and amyloid deposition of Aβ protein [**37, 38**], hyperphosphorylation of tau protein **[39]**, problems with acetylcholine production **[40],** a viral cause **[41]** or a combination of some or all of these factors. Genetic evidence from familial forms of AD strongly supports a role for Aβ in the disease [**42**]; however neither the load nor the localisation of amyloid plaques correlates with cognitive impairments in patient [**43, 44, 45**] and an experimental vaccine which could clear the amyloid plaques in early human trials had no significant effect on dementia [**46**]. Neurofibrillary tangles of Tau protein do correlate with cognitive decline and neuron death [**44, 45, 47, 48**], however genetic evidence indicates that tau causes frontotemporal dementia rather than AD [**49, 50**]. It has been suggested that both Tau and Aβ act together at the neuronal synapses where, when dysfunctional they disrupt the synaptic structure and function [**51, 52, 53**]. This suggested molecular basis for AD is supported by the suggestion that toxic intermediates of Aβ protein are responsible for AD rather than the mature amyloid fibres [**54, 55**], interestingly putative targets of these intermediates include the receptor Prion Protein implicated as the cause of Creutzfeldt-Jakob disease (CJD) [**56**].

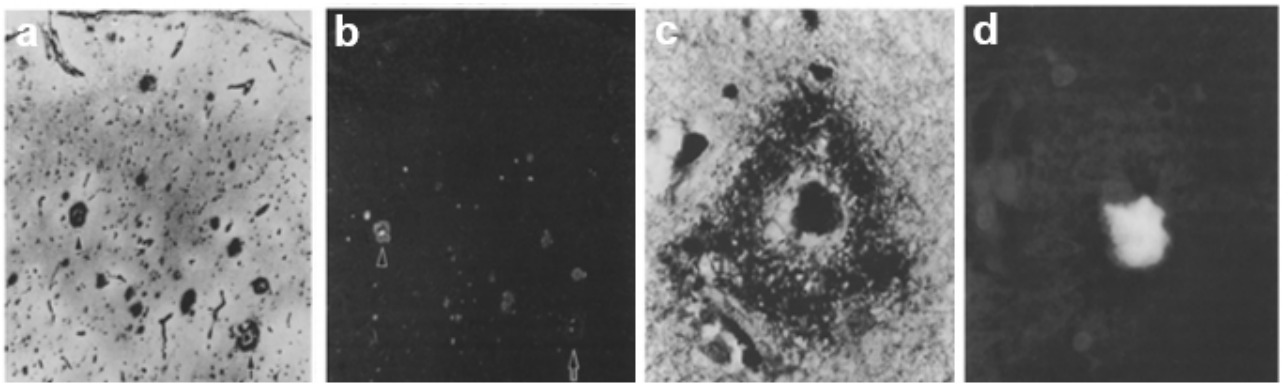


Figure 1.3: Histological samples of Alzheimer's plaques. a, Bielschowsky stained section of the frontal cortex from an Alzheimer's disease brain showing the presence of numerous plaques. **b**, Thioflavin S stained section of the same brain showing plaques (indicated with arrows) but detecting fewer of them. **c**, Bielschowsky stained classical plaque showing dense core and peripheral halo. **d**, ThioflavinS stained classical plaque showing staining only of the dense core. Images taken from Wisniewski et al., 1989 [**57**].

1.2.3 Creutzfeldt-Jakob Disease

Creutzfeldt-Jakob disease (CJD) is an incurable and highly unusual transmissible human neurodegenerative disease which can be directly linked with a group of animal diseases known as transmissible spongiform encephalopathies. The spongiform description comes because as the disease causes rapid neurodegeneration it causes neurons in the brain to die forming holes in the brain tissue resulting in a more sponge-like texture. These diseases can occur in genetic, spontaneous or acquired forms; in the acquired forms the transmissible agent, according to the prion hypothesis, is an epigenetic pathogen: the Prion Protein (PrP), with the diseases being, in a sense, forms of transmissible amyloidosis [**58, 59**]. In this suggested model for transmissible spongiform encephalopathies misfolded copies of the PrP replicate by converting correctly folded molecules in their host to the misfolded phenotype, the number of these molecules increases exponentially disrupting cell function and causing cell death **(Figure 1.4)** [**59**]. Studies have shown that mutations in the gene for PrP, which cause a change in secondary structure propensity from alpha helical to beta pleated sheets, appear to contribute to the protease resistant, amyloid forming structure of the disease causing protein [**60**]. CJD causes a rapidly progressing form of dementia often resulting in death within 6 months although various strains have been reported with a diverse range of incubation periods and life expectancies [**61**]. Although the prion hypothesis for CJD has been generally accepted and is broadly supported experimentally there is also evidence from some studies of a potential viral involvement or cause for the disease [**62**].

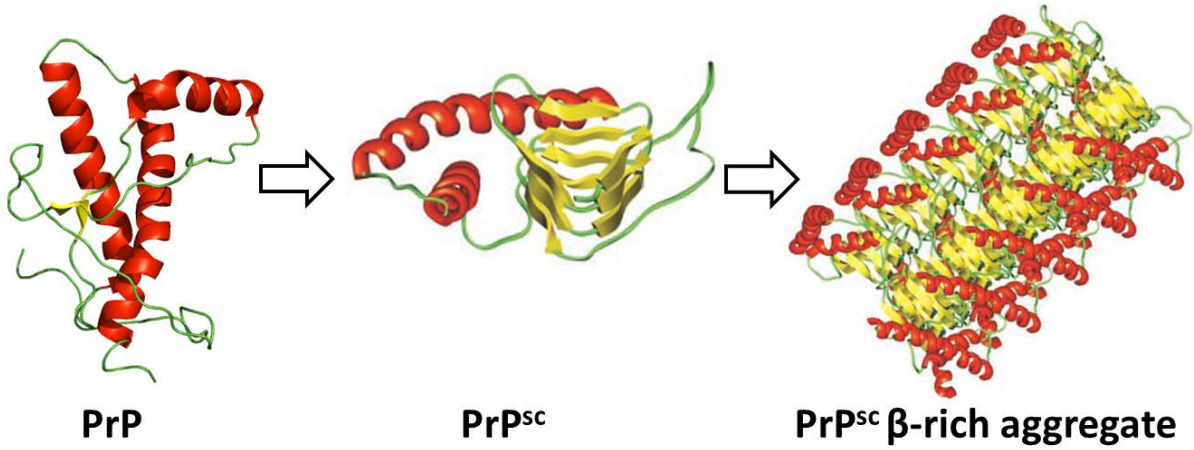


Figure 1.4: Conversion of PrP from normal form to scrapie form in CJD and related diseases. Prion Protein has two main structural forms for which structural data has been elucidated. A largely α-helical structure, with a long unstructured N-terminus, labelled here as PrP, discovered by Biljan et al., 2012 [**2LSB**] [**63**], and a β-strand rich structure described, and labelled here, as PrP^{sc} a model of which has been suggested by Govaerts et al., 2004 [**64**]. The prion hypothesis suggests that PrP^{sc} encounters native PrP and converts it to the PrP^{sc} form, these molecules propagate and can aggregate into PrP^{sc} β-rich aggregates. The PrP^{sc} and PrP^{sc} β-rich aggregate model structures shown here are taken from Govaerts et al., 2004 [**64**].

1.3 Functional Amyloids In Nature

1.3.1 Overview

Interestingly in addition to disease causing amyloids composed of malfunctioning proteins, amyloids have also been shown to be functionally employed by a broad range of species including bacteria, insects and mammals (**Table 1.2**). In these systems specific proteins are produced that natively form filamentous aggregates with many of the biophysical properties of amyloids, sometimes with their own devoted export machinery (Type VIII Secretion System) [**65**]. These amyloids include the curli system of *E. coli,* **[22, 66]** the chaplin proteins of the bacterium *Streptomyces* **[67]**, and the TapA/TasA spore coat component of *Bacillus subtilis* **[68]**. The presence of amyloid in the biofilm of many different species of bacteria suggests that it has an important role in biofilm structure **[69]**, some amyloids have also been implicated in pathogenesis **[70, 71, 72, 73]**. The physical properties of amyloid, insoluble, aggregative fibres, resistant to thermal and chemical denaturation capable of forming under a broad range of conditions, are likely to make them especially useful when their formation can be controlled, their potential for biotechnological purposes is also interesting and is currently being explored [**74, 75**].

Table 1.2: Functional Amyloid Systems in Nature. Table illustrating both the broad range of species possessing some form of functional amyloid and some of the functional roles of amyloid.

Functional Amyloid	Species	Role
Curli/Tafi	*E. coli/Salmonella*	Biofilm/Pathology
GvpA	*Anabaena flos-aquae*	Gas Vesicles
HpaG	*Xanthomonas*	Virulence Factor
Fap System	*Pseudomonas*	Biofilm
Chaplins	*Streptomyces coelicolor*	Aerial mycelium formation
Swi1p	*Saccharomyces cerevisae*	Chromatin Remodelling
Sup35	*Saccharomyces cerevisae*	Translation termination
Microcin E492	*Klebsiella pneumonia*	Cytotoxin
Merozoite Surface Protein 2	*Plasmodium falciparum*	Attachment to Host
MTP	*Mycobacterium tuberculosis*	Adhesion
Spider Silk Proteins	Some Spider Species	Silk Production
Pmel17	*Homo sapiens*	Melanin Production
TasA/TapA system	*B. subtilus*	Spore Coat Component

1.3.2 Biofilm & Amyloid

The importance of biofilm has only relatively recently been made clear, with most scientific research focussing on the planktonic forms of bacteria. In nature many micro-organisms form natural aggregates which have been described as biofilms, in this environment bacteria can behave very differently from their planktonic forms, possessing differentially altered gene expression, developing complex architecture, engaging in co-operative behaviour and frequently co-existing with a broad range of other microbes in a symbiotic relationship [**76, 77, 78, 79**]. With their complex structure and behaviour bacterial biofilm is similar to tissues formed of eukaryotic cells in multicellular organisms [**80**]. Biofilms serve several important roles for bacteria including increasing their resistance to desiccation, antibiotics, radiation, predation, host immune defences as well as both physical and chemical stress [**81, 82]**A variety of factors have been suggested to provide bacterial biofilm with increased resistance: the matrix provides a physical barrier which even retards the diffusion of small molecules such as antibiotics [**83**]; the heterogeneity of bacteria in the biofilm also confers further protection, with a sub-population of bacteria likely to be resistant to any particular metabolically targeted attack, additionally slow-growing or senescent cellular populations will be more resistant to many common antimicrobial agents; the physical properties of the biofilm itself also help protect the bacteria from physical stresses such as host clearance mechanisms or the flow of water. Biofilm formation and composition are of considerable interest in many fields, as they can block and corrode pipes, are involved in food spoilage and they are intricately involved in hospital acquired infections and disease **[81, 84]**. The majority of a biofilm by dry mass is composed of the extracellular matrix within which the cells are buried [**85, 86**]. The extracellular matrix is a complex mixture of different biopolymers: various sugars, nucleic acids and proteins including extracellular amyloid fibres **[76, 84, 87, 88, 89, 90]**. *Pseudomonas* biofilms develop in a series of stages (**Figure 1.5**) initiated by attachment of planktonic cells to a surface in response to favourable nutrient conditions, formation of microcolonies, maturation of these colonies and then dispersal, where planktonic cells are released again **[91]**. The presence of amyloid in *Pseudomonas* biofilm is particularly interesting as, although *Pseudomonas* is a model organism for the study of biofilm formation, the role of amyloid is uncertain [**87, 92**].

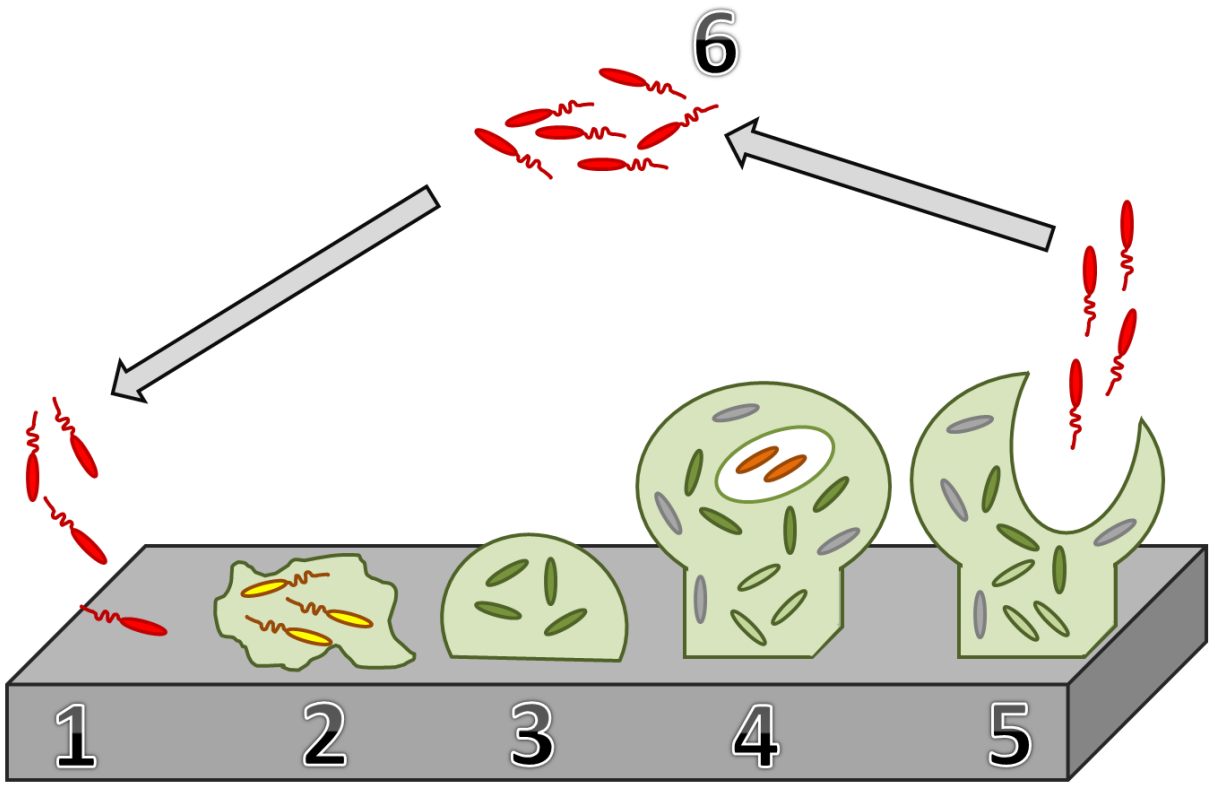


Figure 1.5: *Pseudomonas* Biofilm Maturation. Cartoon diagram illustrating the steps in the formation, maturation and propagation of *Pseudomonas* biofilm. 1) Association planktonic bacteria encounter a surface and reversibly attach themselves. 2) Adherence, after attachment bacteria can irreversibly adhere to the surface, switch their gene expression (indicated by colour change) and begin secreting biofilm polymers, the biofilm matrix is shown here in green. 3) Microcolony Development, bacteria some of their motility (indicated by loss of flagellum), biofilm develops into microcolony. 4) Maturation, the bacterial colony matures forming a mushroom like structure, bacteria show differential gene expression (indicated by the range of colours), some form the stalk (light green), some prepare to switch back to planktonic expression (orange), differentially expressing bacteria close to the surface (grey) have been suggested to express amyloid. 5) Dispersal, a subset of the bacteria are released to return to their planktonic lifestyle. 6) Planktonic bacteria migrate through the environment until conditions favour biofilm formation.

1.4 Curli

1.4.1 *Escherichia coli*

Escherichia coli (*E. coli*) is a common Gram-negative commensal bacterium and a model organism in microbiology, however there are several pathogenic strains which can cause serious diseases such as gastroenteritis and haemolytic-uremic syndrome [**93**]. *E. coli* infections are generally extracellular with bacteria colonising the lumen of the gut and adhering to the surface of epithelial cells; biofilm formation is an important element for both commensal and pathogenic *E. coli* allowing increased survival in the environment, resistance to antibiotics, the immune system, and host clearance mechanisms [**94**]. The developing world suffers disproportionately from the gastrointestinal diseases with up to 4 billion people infected each year [**95**] leading to approximately 2.2 million deaths per year [**96**].

1.4.2 Curli System

The Curli functional amyloid from *E. coli* has been studied extensively as the model for functional amyloid formation in bacteria. The curli system produces aggregative extracellular fibres (**Figure 1.6**) that have been shown to be involved in biofilm formation, comprising the major proteinaceous component of the matrix, as well as having a role in colonisation and the binding to host cell proteins including fibronectin, laminin, plasminogen and human contact phase proteins **[72, 73].** Curli expression in biofilms is restricted to a distinct subpopulation, for example within rugose colony biofilm the curli expressing cells localise to the air-colony interface [**97**]. In addition to their role in biofilms curli fibrils have been implicated in pathogenesis and infection due to their adhesive properties and since they may be involved in internalizing *E. coli* into eukaryotic cells **[70]**. Curli fibres appear to be produced by many bacterial species as homologs to the fibre component CsgA are found in many families of bacteria and even the fungus *Beauveria bassiana* (**Figure 1.7**). Many of these bacteria are widespread and several have been identified as opportunistic pathogens.

The curli system (**Figure 1.8**) is encoded by two divergently transcribed operons *csgAB* and *csgDEFG*. The csgDEFG operon is carefully regulated, reflecting Curli's specific spatiotemporal expression and the diverse range of environmental signals guiding expression **[98, 99]**. CsgD itself is a transcriptional regulator from the LuxR family involved in regulating expression of many biofilm components and repressing the flagellar genes **[100, 101]**. CsgA has been shown to be the main component of the fibre and is secreted into the extracellular space as an unstructured, soluble, monomer where its polymerisation is nucleated by CsgB on the extracellular surface **[102]**, indeed the two proteins show interbacterial complementation between colonies a few millimetres apart **[102]**. This mechanism of secretion has been described as nucleation-precipitation or as a Type VIII secretion system [**65**]. CsgE and CsgF have been shown to act as assembly factors of some kind

ensuring the polymerisation of CsgA on the cell surface and anchoring the fibres to CsgG **[103, 104]**. The structure of CsgG has been solved recently showing that the lipoprotein oligomerises to form a large B-barrel structure which presumably provides the channel through which the other extracellular curli components exit the cell **[105]**. The role of CsgC was unclear despite a high-resolution structure of the protein, initially it was suggested that it modifies CsgG at position C230 via its oxido-reductase activity and could have a role in regulating the system **[106]**. More recently functional studies of CsgC's role in amyloid fibre formation have indicated that it inhibits amyloid formation by an unknown mechanism **[107]**. Interestingly analysis of Curli protein conservation across a broad range of bacterial classes showed the existence of a seventh curli component CsgH present in some species which the authors suggested may compensate for the absence of CsgE and CsgF in these bacteria **[108]**.

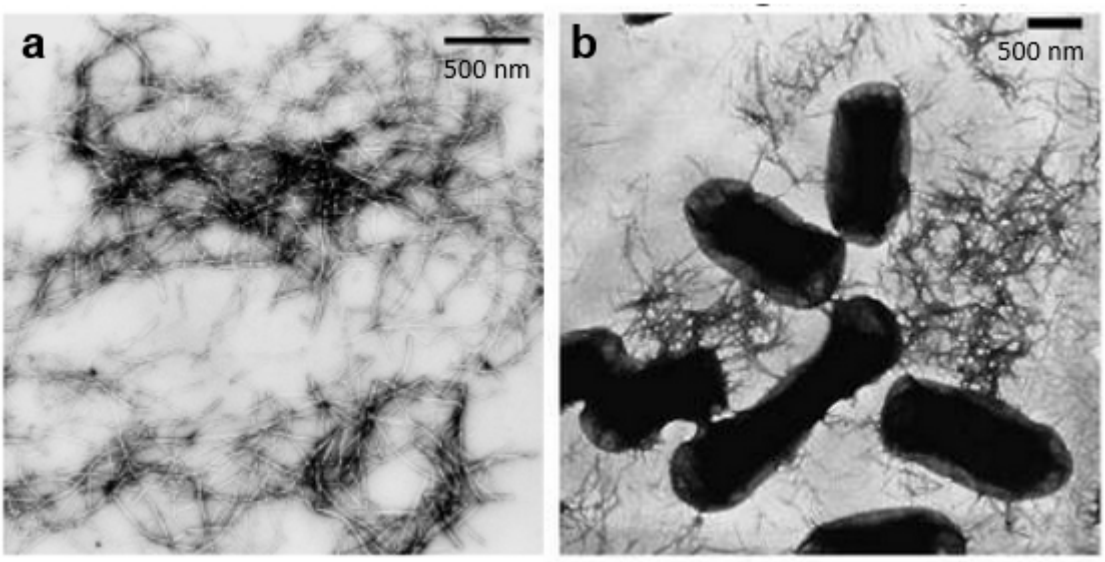


Figure 1.6: Morphology of Curli Fibres. a, TEM images of polymerised CsgA amyloid fibres. b, TEM images of mature Curli fibres on the surface of *E. coli* cells. Microscopy images from Wang *et al.,* 2008 **[109]**, scale bars are both 500 nm in size.

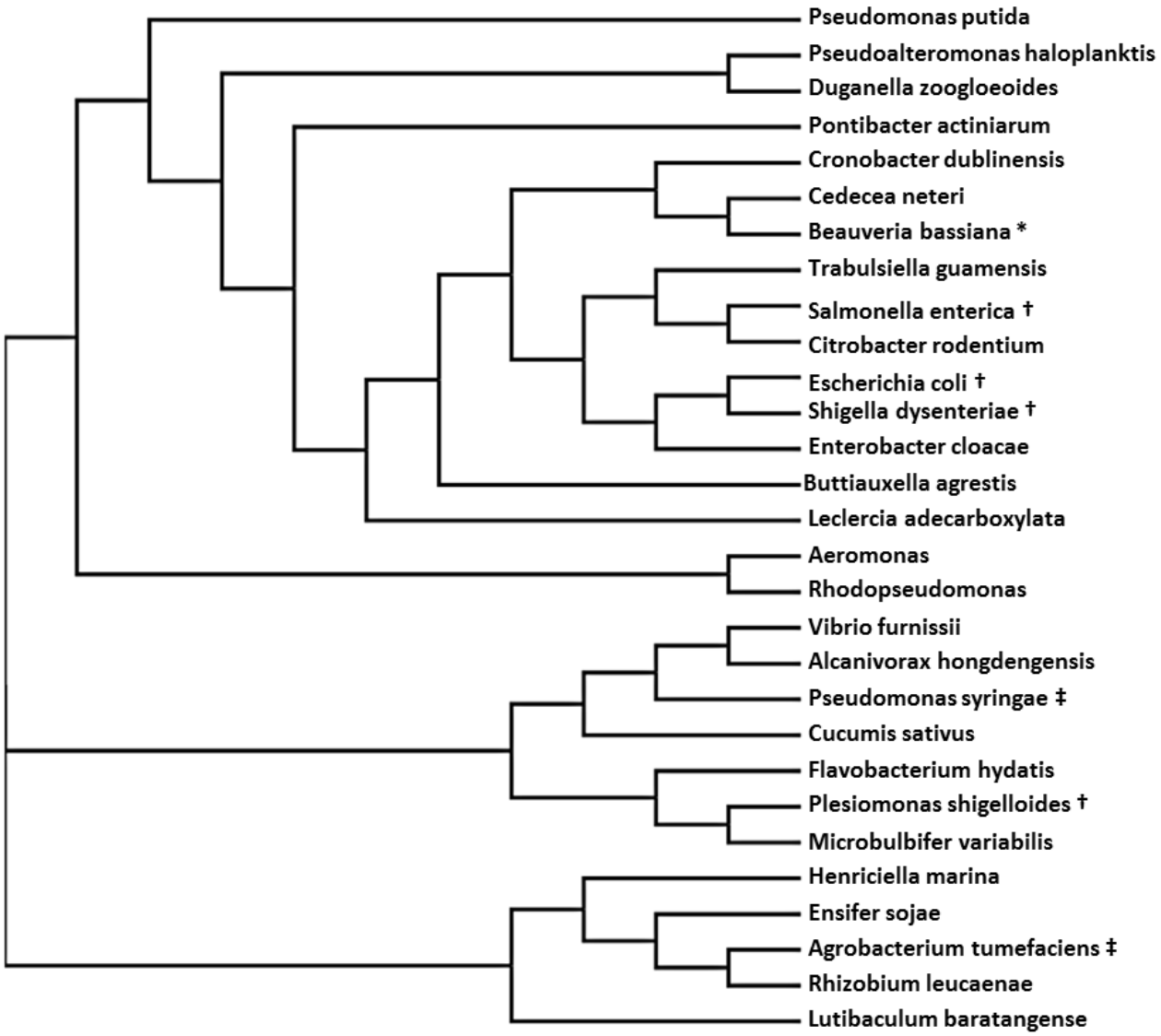


Figure 1.7: CsgA Phylogenetic Tree. Generated using PSI-BLAST and Clustal Omega, ClustalW2-Phylogeny. Homologs of CsgA were identified using PSI-BLAST [**110**] and representative sequences were selected, Clustal Omega was used to produce an alignment from which a phylogenetic cladogram was produced ClustalW2-Phylogeny. CsgA can be seen to be present across a broad range of species including several human pathogens (†), several plant pathogens (‡) and even a fungus which is pathogenic to insects (*).

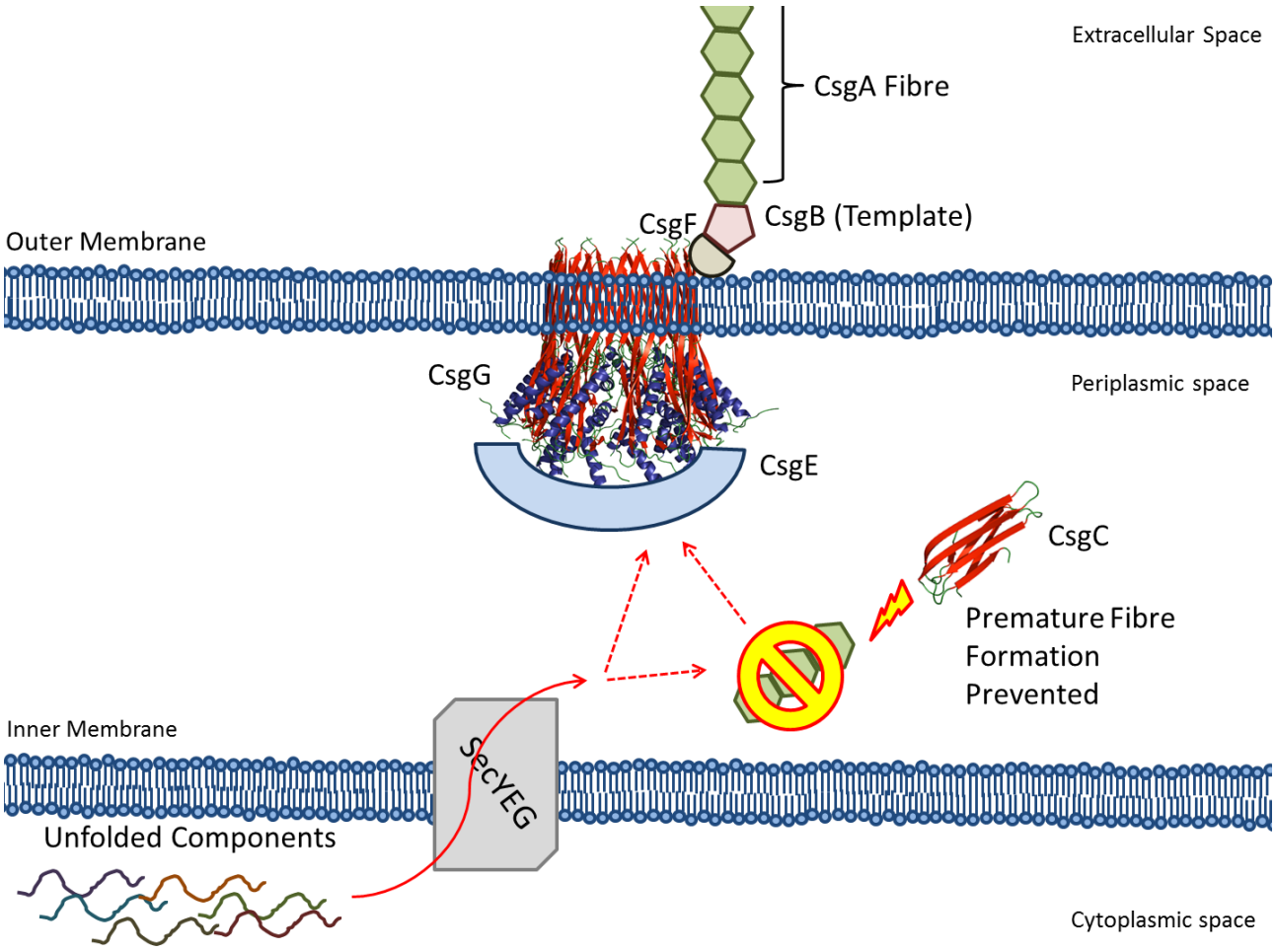


Figure 1.8: The Curli System. Cartoon Diagram illustrating current opinion on the molecular operation of the Curli system. The curli components are transported into the periplasm via the SEC machinery. In the periplasm the amyloid components are prevented from polymerising by the action of CsgC. The monomeric fibre subunits are transferred through the multimeric pore CsgG in a process which may involve CsgE acting on the periplasmic side. On the cellular surface CsgA is templated by CsgB which is anchored to the cell via CsgF. CsgG structure (**PDB: 3X2R**) from Cao et al., 2014 **[111]**. CsgC structure (**PDB: 2Y2Y**) from Taylor et al., 2011 **[106]**.

1.4.3 Curli Fibre Polymerisation

The kinetics of Curli fibre assembly have been extensively studied in vitro, CsgA can be purified under denaturing conditions and readily forms amyloid at a broad range of pH values or ionic conditions, these fibres can be visualised by transmission electron microscopy (TEM) or measured by ThT binding [**112**]. CsgA polymerisation has been shown to be concentration dependent and can be divided into three phases, an initial lag phase where nucleation occurs, and elongation phase during which the fibres polymerise to form filaments and finally a stationary phase, where the pool of monomeric CsgA required for polymerisation is presumably expended (**Figure 1.9**) [**112, 113**]. The mature CsgA protein has an N-terminal sequence required for outermembrane secretion (22 amino acids) and a C-terminal amyloid core domain containing five imperfect repeats which have been described as R1 to R5 (**Figure 1.10A**) [**114**]. These repeats form the core of the amyloid, the repeating unit is Ser – X_5 – Gln – X – Gly – X – Gly – Asn - X – Ala - X_3 -Gln, with the amyloid core possessing a conserved secondary structure of β-sheet -turn- β sheet, these repeating units are arranged in the cross-B structure to align the Gln and Asn residues in stacks to stabilise the fold [**114**, **115, 116**]. Mutations of these residues in R1 and R5 extend the lag phase of CsgA polymerisation dramatically (~100x) compared to wild type [**113, 116**]. The conserved Gly and Ala residues in the repeats are also stacked in a consistent manner suggesting they have a role in amyloid assembly [81]. The effects of the individual repeats have been carefully studied individually as peptides and found to have different amyloidogenic properties, repeat R1, R3 and R5 are amyloidogenic, rapidly assembling into amyloid fibres in vitro, while the intervening repeats R2 and R4 do not [**109**]. Deletion of the first and last repeats (CsgAΔR1 & CsgAΔR5) do not assemble into curli fibres in vivo and cannot be seeded in vitro by CsgA or CsgB [**117**], this indicates that R1 and R5 are important for both amyloid formation and CsgA-CsgA and CsgA-CsgB interactions. In addition to the conserved elements which promote amyloid formation several residues have been identified as "gatekeeper" residues (Gly78, Asp80, Gly82, Gly 123, Asp127) which repress the amyloid forming properties of the repeats [**89**]. Mutation of all the gatekeeper residues to their R1/R5 alternatives produce a variant of CsgA which polymerises without an observable lag phase in vitro and interestingly is capable of forming curli in vivo in a CsgB/CsgF independent manner, this indicates that the gatekeeper residues exert a form of intramolecular control over CsgA polymerisation [**117**].

CsgB, the nucleator protein has many features in common with CsgA, including 30 % sequence identity, an N-terminal domain and a C-terminal amyloid core. The amyloid core is similar to that of CsgA (**Figure 1.10**), with the initial four imperfect repeats possessing a similar sequence to those of CsgA: Ala – X_3 – Gln -X – Gly - X_2 – Asn - X – Ala – X_3 – Gln, however the fifth repeat diverges significantly from the pattern with several residues no longer conserved and the

addition of four positively charged residues which appear to have a role in association of CsgB with the cell surface [**118, 119**]. Deletion of the fifth repeat has no impact on CsgB's ability to nucleate CsgA and the protein is able to form amyloid in vitro. Repeats 1-3 can be deleted without any apparent effect on amyloid formation or localisation, suggesting repeats 4 and 5 are critical for both amyloid nucleation and cell surface association [**120**]. How CsgB is associated with the membrane is still unclear but another curli system protein, CsgF, is required [**104**].

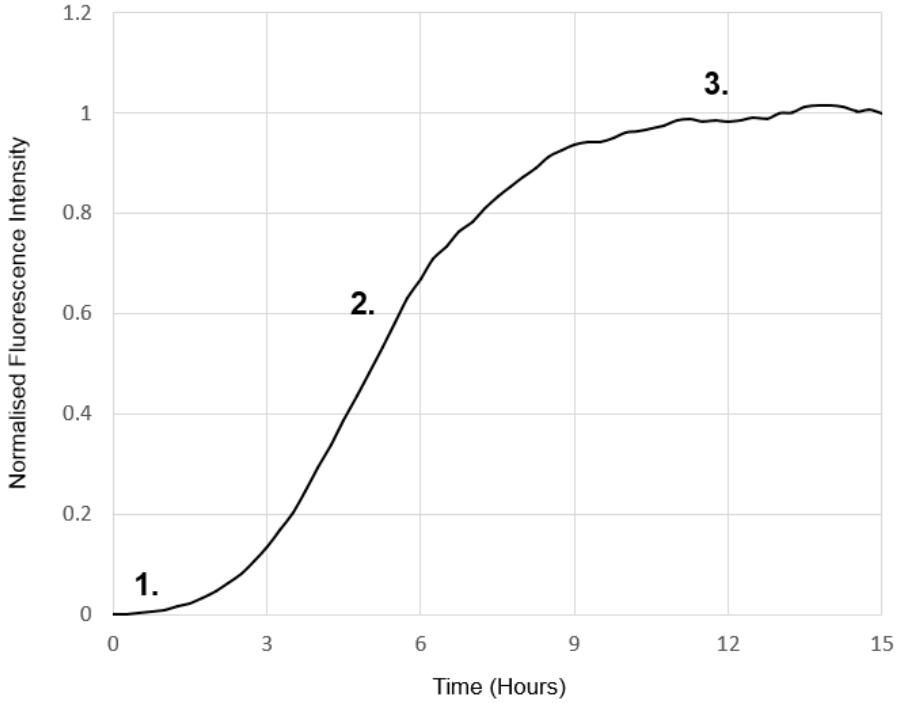


Figure 1.9: Amyloid Fibre Fibrillation As Measured By ThT. This figure shows the stages of amyloid fibre fibrillation: 1) An initial lag phase where nucleation is occurring; 2) Elongation phase where fibres are polymerising rapidly; 3) Stationary phase where fibre polymerisation slows due to lack of available csgA. Fluorescence is normalised with the final intensity as 1. The curve is taken from my own data.

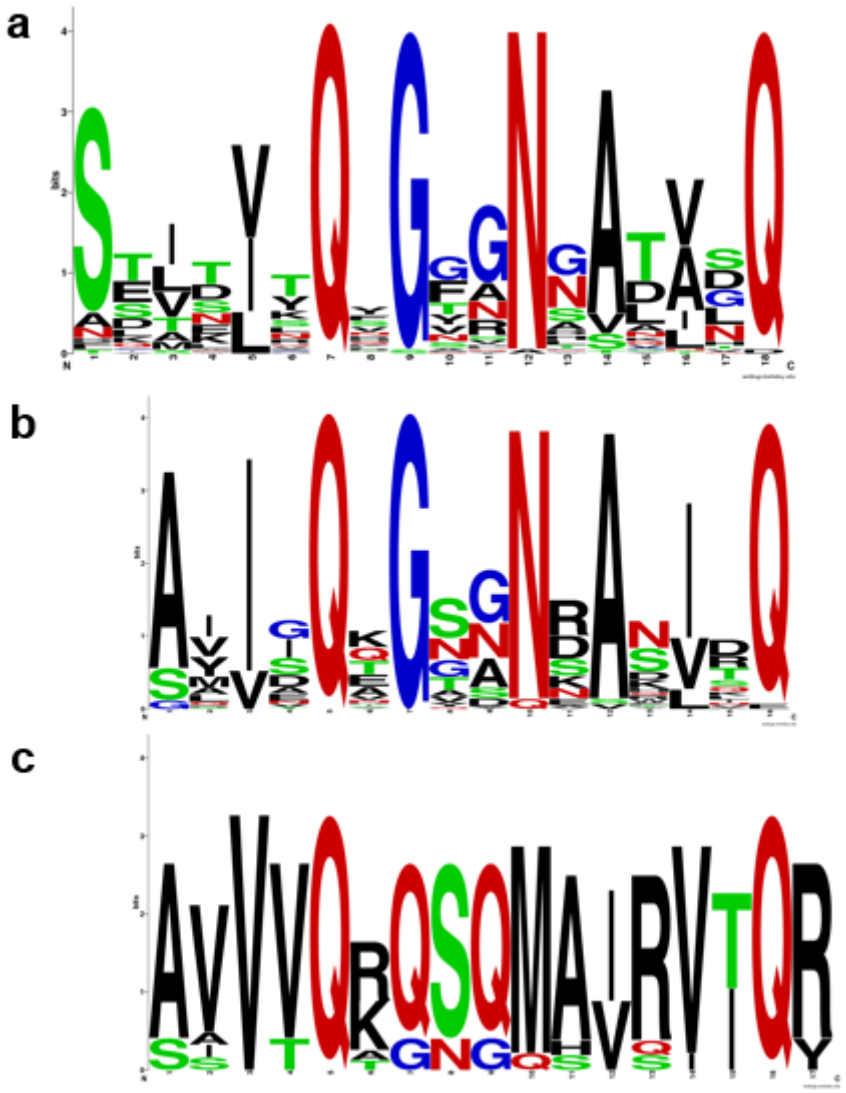


Figure 1.10: Amyloid Repeat Motifs of CsgA and CsgB. a, Diagram illustrating the conservation of residues in the five repeats of CsgA from species containing closely homologous sequences. **b**, Diagram illustrating conservation of residues in the first four repeats in CsgB from species containing closely homologous sequences **c**, Diagram illustrating conservation of residues in the divergent fifth repeat of CsgB from species containing closely homologous sequences. Homologs were identified using PSI-BLAST **[110]**, aligned with clustal omega **[121]** and then their conservation plotted using webLOGO software. The main repeats of both proteins clearly share a Q-X-G-X_2-N-X-A-X_3-Q motif, while the fifth repeat of CsgB is clearly quite distinct; nevertheless all contain the Q-X_{10}-Q motif, by which the three logo diagrams are aligned. Interestingly repeat 5 of CsgB is also the best conserved between species.

1.4.4 Structural & Functional Insights Into Curli

The structure of CsgG (**Figure 1.11**),was elucidated recently by two separate groups, interestingly the protein was found to form a large nonameric transport complex, with each monomer contributing 4 strands to a central 36 stranded B-barrel, forming the transmembrane channel, with a large domain extending into the periplasm [**105, 111]**. The central channel of CsgG is broad apart from a narrow chokepoint created by a series of loops projecting towards the centre of the barrel (**Figure 1.11B**). The large empty cavity facing the extracellular space would allow for the possibility of CsgF binding into the barrel to allow CsgB to associate with the CsgG complex [**122**]. The chokepoint introduced by the loops would allow some control over substrate selection and is wide enough to allow some secondary structure elements (such as β-strands) [**123**]. Since CsgG is embedded in the outer membrane without adenosine triphosphate (ATP) or a chemical gradient to power the mechanism, it is currently unclear how the necessary energy is provided for curli secretion.

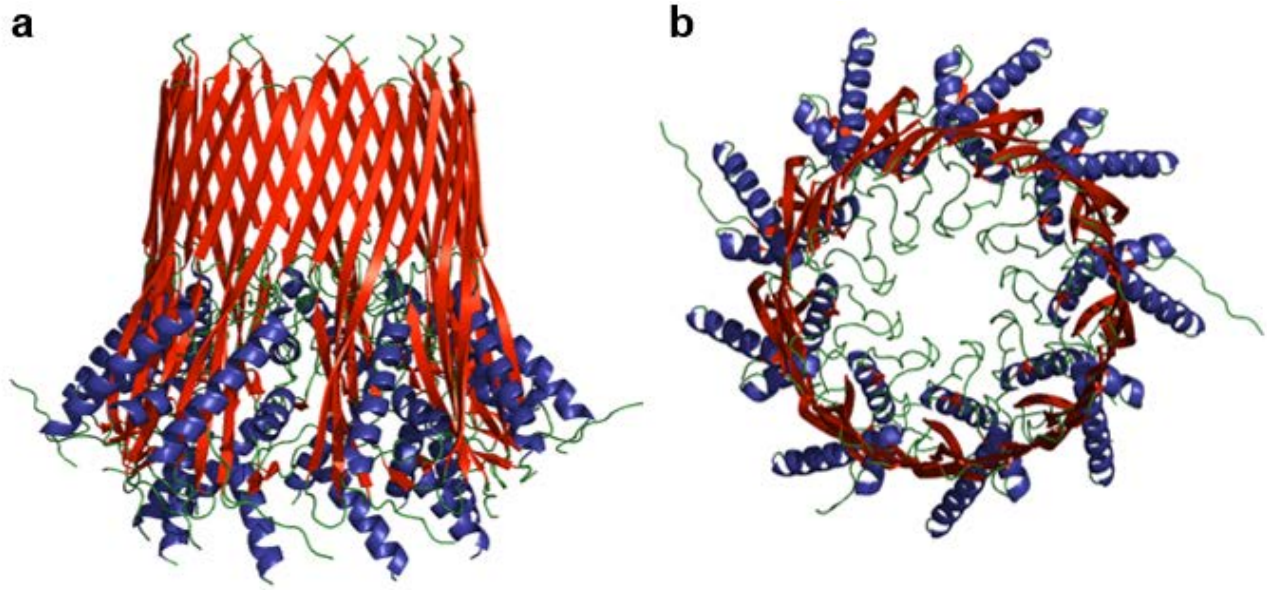


Figure 1.11: Crystal Structure of membrane inserted CsgG. a, CsgG structure (**PDB: 3X2R**) from Cao et al., 2014 [**111**] viewed **a**, from the side (Parallel to membrane), **b**, from the top (extracellular side); coloured by secondary structure with strands in red, helices in blue and loops in green. CsgG forms a large nonameric B-barrel structure with a large domain extending into the periplasm. The channel is constricted by a series of loops clearly visible in the top down view, due to the narrow channel and based on molecular modelling it was suggested that the substrates pass through the barrel as fully extended polypeptide chains **[105]**.

Together with the CsgG structure Goyal et al., 2014 also conducted cryo-em and single channel current recordings which indicated that an oligomer of CsgE caps the channel, on the periplasmic side to regulate access, together with the knowledge that CsgE is capable of recognising CsgA/CsgB via the N-terminal or curli repeats [**22**], this data was used by the authors to suggest some putative models for amyloid secretion (**Figure 1.12**) [**105**]. It should be noted however that CsgE is not essential for the export of CsgA, as co-expression of CsgG and CsgA together is sufficient for export and CsgE is not conserved in all curli operons [**108, 123**]. Interestingly CsgE has been shown to have an inhibitory effect on amyloid formation, an activity common to several chaperones, but it is significantly less efficient than CsgC suggesting that amyloid inhibition is not the primary function of the protein [**103**, **124, 125**].

The structure of the CsgC protein was solved by X-ray crystallography in 2011 (**Figure 1.13**), the overall structure was an immunoglobulin-like (beta sandwich with seven strands forming two sheets [**106]**. The sequence of CsgC contains an interesting conserved CxC motif, due to the position of this motif in the "complementarity-determining region" of the Immunoglobulin (IG) fold, a common site of target recognition, and the structural homology to the redox protein DsbD it was suggested that the protein acts via a redox activity and prior to the elucidation of the structure of CsgG it was suggested that residue C230 in the membrane protein was an exposed potential target [**105, 106**]. Functional data has since revealed that CsgC is a potent inhibitor of amyloid formation, capable of inhibiting the polymerisation of both CsgA and α-synculein based amyloid fibres. The inhibition of CsgA fibrillation is effective at molar ratios as low as 1:500 (CsgC:CsgA), via a process which does not disassemble existing fibres but inhibits the transition of the pre-amyloid structure into β-sheet [**107**]. It was shown that CsgC inhibition worked on a range of amyloid forming polypeptides, which Evans et al., 2015 suggested share a D-Q-ϕ-X0,1-G-K-N-ζ-E motif [**107**]; however observation of CsgA sequence conservation indicates this motif is not conserved and so may not be as specific as the authors suggest. The authors also found that CsgC does not direct CsgA into non-amyloid aggregates or degrade CsgA by proteolysis [**107**]. Based on the available data it appears that CsgC is likely to have a role in preventing the amyloid via unknown, novel mechanism.

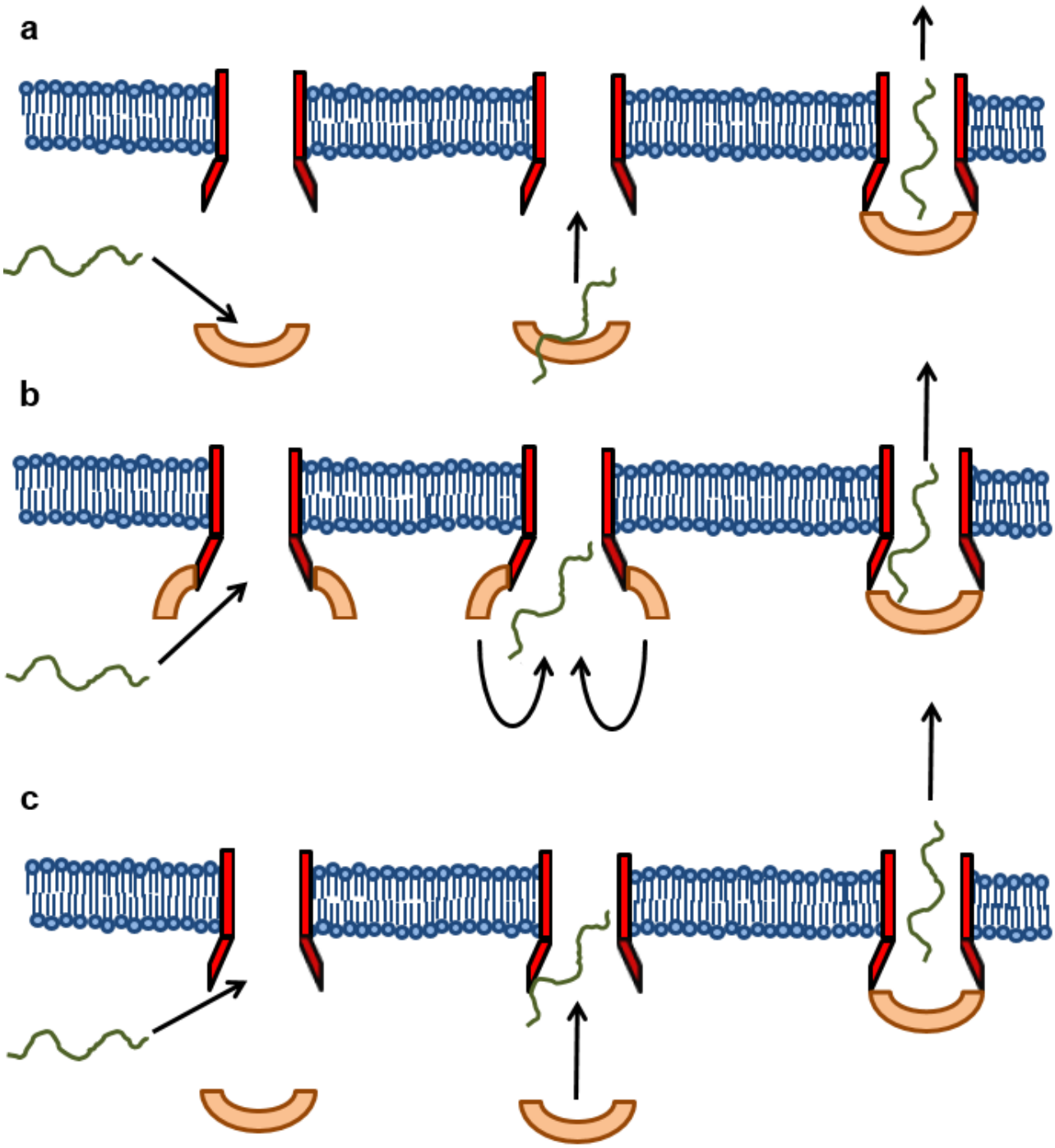


Figure 1.12: Suggested Mechanisms for CsgG-CsgE mediated export of CsgA. Cartoon Diagrams illustrating suggested mechanisms for CsgG-CsgE mediated export of CsgA. The CsgG oligomeric B-barrel is shown here in red, CsgA in green and CsgE in tan. In the paper by Goyal et al., 2014, **[105]** the authors suggested 3 possible models for assembly and substrate recruitment of the CsgG complex. **a**, 'Dock and trap' CsgE binds CsgA and the complex associates with CsgG. **b**, A CsgE, CsgG constitutive complex recruits CsgA in what might be described as 'Hook-and-Trap' **c**, 'Catch-and-cap' CsgA associates with CsgG and then CsgE associates with the complex to promote translocation.

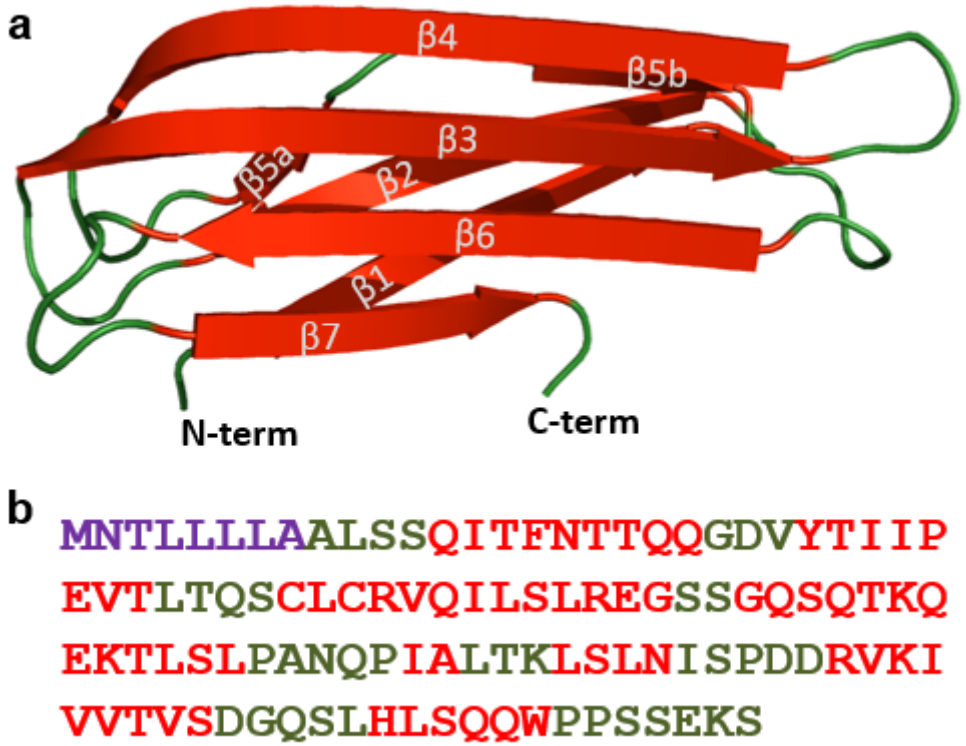


Figure 1.13: Structure of CsgC. a, Crystal structure of CsgC (**2Y2Y**, Taylor et al., 2011 **[106]**) with coils in green and β-strands in red, the N-terminal and C-terminal are indicated with N-term and C-term respectively, note that the signal peptide, and short regions of the N-terminal coil and C-terminal coil are not present in the crystal structure, strands are labelled by their order from N-terminus to C-terminus. **b**, Primary & secondary structure of CsgC showing the protein sequence with the signal sequence indicated in purple, secondary structure is indicated by colour based on the crystal structure with coils in green and β-strands in red Figure based on structural information elucidated by Taylor et al., 2011 **[106]**.

1.4.5 CsgH

Studies of the Curli operon's conservation between different bacterial species by Dueholm et al., 2012 led to the identification of an additional gene present in some curli operons which they termed CsgH [**108**]. CsgH possesses no sequence similarity to other members of the curli operon but due to its presence in several strains lacking CsgE and CsgF it was suggested it may play a similar role to those genes [**108**]. The predicted secondary structure for CsgH is similar to the known secondary structure of CsgC (**Figure 1.14**) and CsgC is also often absent in strains containing CsgH [**108**]. If these predictions are accurate it is possible that CsgH possesses a similar function to CsgC. A structural similarity between the two proteins, particularly conservation of β-strands, is especially interesting because it has been suggested that the β-rich structure of CsgC may allow it to inhibit CsgA via an amyloid like interaction [**107**]. If CsgH is another amyloid inhibiting protein like CsgC further study of the protein will be useful for comparison and study of amyloid regulation and inhibition in functional amyloid.

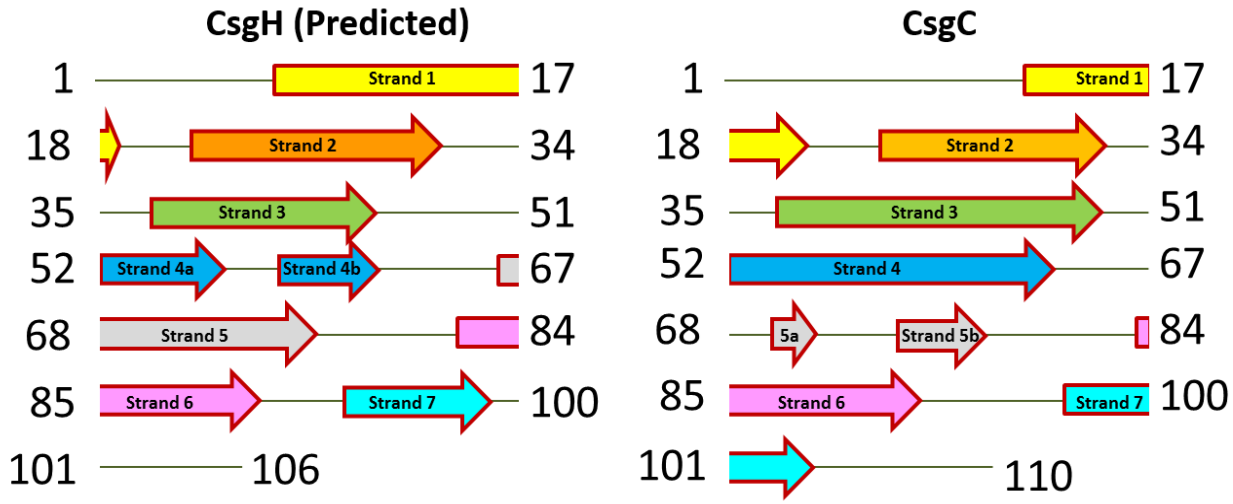


Figure 1.14: Predicted secondary structure of CsgH compared to CsgC. This cartoon diagram illustrates the similarity between the predicted secondary structure of CsgH (PSIPRED **[126]**) and the secondary structure of CsgC determined by Taylor et al., 2011 **[106]**. The strands are indicated with arrows and the loops by straight lines, the strands have been colour coded to indicate strands which are suggested to match each other between the two proteins.

1.5 *Pseudomonas* Amyloid-like fibres (Alf)

1.5.1 *Pseudomonas, fap*ABCDEF and Alf

A novel amyloid and biofilm forming system of proteins which was shown to produce fibres morphologically similar to curli has been identified in *Pseudomonas* and described as Amyloid-like fibres (Alf) **[92]**. *Pseudomonas* is a Gram-negative, aerobic, bacterial genus, whose members cover a broad range of environmental niches, and possess a great deal of metabolic diversity, with many non-pathogenic species and including other pathogenic species such as the plant pathogen *Pseudomonas syringae*. *Pseudomonas* persistence is aided by its ability to form biofilms as well as its inherent resistance to a broad range of antibiotics and to readily acquire further resistance by mutation **[127, 128, 129, 130]**. *Pseudomonas aeruginosa* is a significant pathogenic bacterium involved in opportunistic and hospital acquired infections, indeed the species is responsible for around 20 % of all cases of nosocomial pneumonia **[131]** and is also the primary cause of chronic respiratory infections, at around 60 % of cystic fibrosis patients for whom the infection often leads to death by respiratory failure **[132]**. It was previously thought that *Pseudomonas* bacteria were obligate aerobes but it has been shown that in biofilm *Pseudomonas aeruginosa* engages in anaerobic metabolism [**133**]. While *Pseudomonas syringae* is a significant plant pathogen, well studied for its Type III secretion system, *Pseudomonas fluorescens* and *Pseudomonas putida* are both used to promote healthy plant growth, with the latter also showing potential for bioremediation [**134, 135**].

Duelholm *et al.*, 2010 found that the Alf system itself is encoded by a single operon of six genes *fap*ABCDEF **(Figure 1.15)**, this operon is conserved among the pseudomonads **[92]** and is present in a range of bacterial species (**Figure 1.16).** Transformation of the *fap*ABCDEF operon, in the vector pMMB190Ap, into *E. coli* was shown to be necessary and sufficient to produce pellicle formation, a sign of biofilm formation [**92**]; Electron microscopy of the *E. coli* cells showed extensive amyloid fibrillation on the surface of the *E. coli* cells, very similar to those observed on *Pseudomonas* (**Figure 1.15**) [**92**].

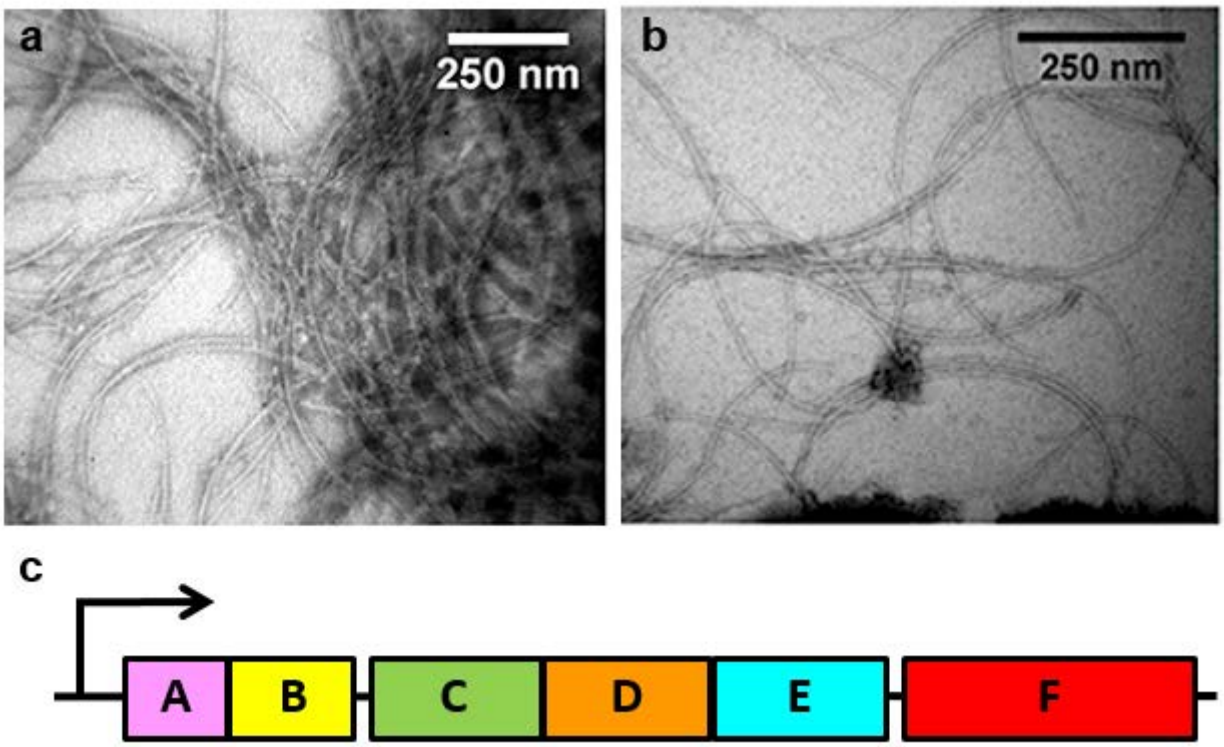


Figure 1.15: *Pseudomonas* Amyloid-like fibres produced by the *fap*ABCDEF operon. Negatively stained transmission electron microscopy images from Dueholm et al., 2010 **[92]**: **a**, of purified *Pseudomonas* UK4 amyloid fibres **b**, of purified amyloid fibres from *E. coli* expressing the fapABCDEF operon showing that *Pseudomonas* produces fibres morphologically similar to curli and that similar fibres can be produced by *E. coli* expressing the *fap*ABCDEF operon [**92**]. **c**, Diagram showing the topography of the *fap*ABCDEF operon which was necessary and sufficient for amyloid formation [**92**].

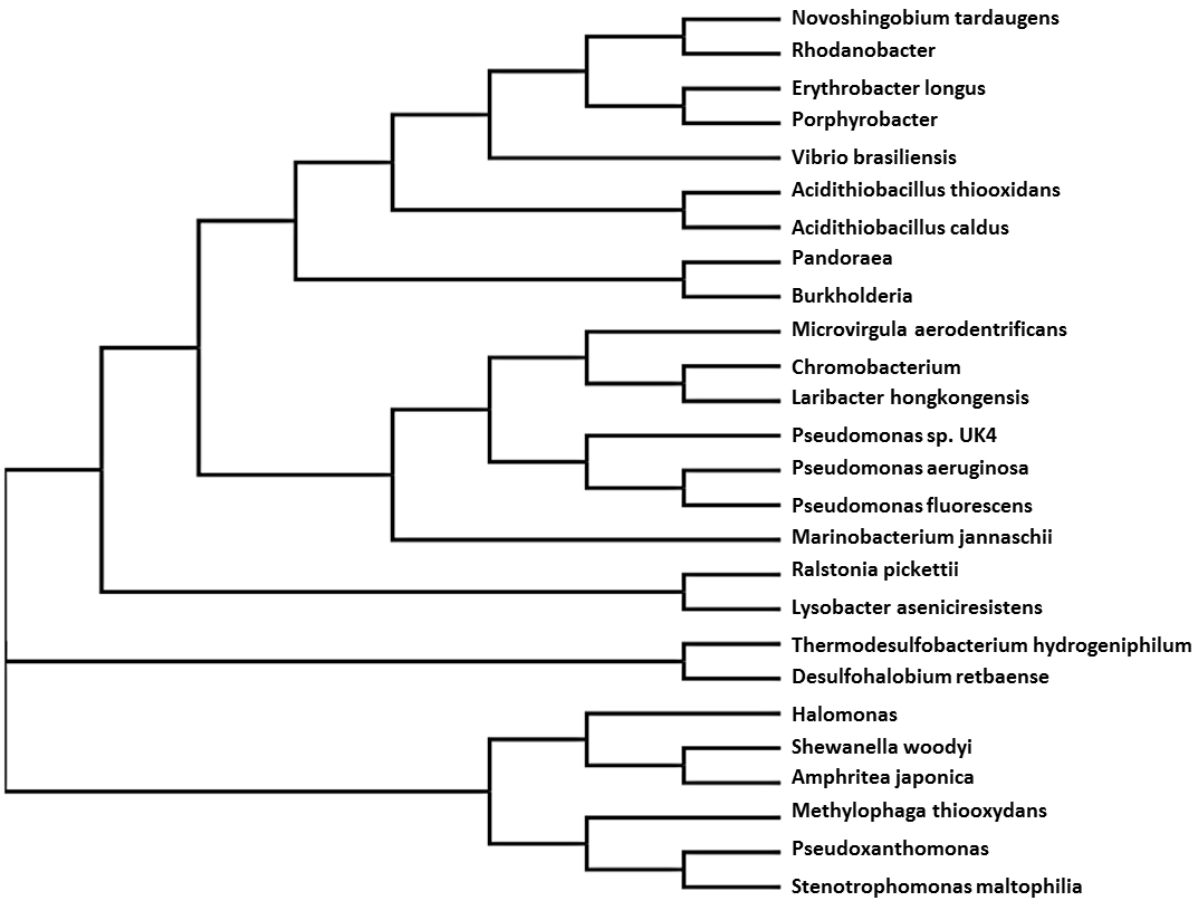


Figure 1.16: FapC conservation among the Pseudomonad species. Phylogenetic Tree produced using PSI-BLAST and Clustal Omega **[121]**, ClustalW2-Phylogeny **[137]**. Homologs of FapC were identified using PSI-BLAST **[110]** and representative sequences were selected, Clustal Omega was used to produce an alignment from which a phylogenetic cladogram was produced ClustalW2-Phylogeny. FapC homologs can be seen to be present across a broad range of gram negative bacteria and found in a diverse range of environments however it appears to be less widespread than CsgA (**Figure 1.7**).

1.5.2 The Alf System

FapC was found to be the main component of the amyloid fibres analogous to CsgA in the curli system, similarly to CsgA it also contained a series of repeat motifs containing conserved glutamine and asparagine residues, connected by less well conserved loops, however FapC contains three 37 residue repeats rather than five 17 residue repeats and the loops between the strands are significantly longer [**92**] (**Figure 1.17, 1.18**). FapB is homologous to FapC and was shown to also contain the repeat motifs [**92**] comparison with the Curli system would suggest therefore that FapB is analogous to CsgB and serves as the nucleator for amyloid formation. It has been suggested that the FapC proteins contain a N-X-A-X_4-S-X_5-G-N-X-G-X-N-X-$_3$G-X_2-N-Q-Q-X-N-X_7 [**92**] or a more minimalistic X_{15}-G-X_4-N-X_3-G-X_6-N-X_7 repeat [**138**], this latter result is supported by alignment of the homologous sequences found by PSI-BLAST using the FapC *Pseudomonas* UK4 sequence as query (**Figure 1.18**). Dueholm et al., 2013 also showed that FapC from several strains may contain a conserved C-terminal CxxC motif and their respective FapE proteins contained C-terminal Cys residues which they suggested may have a role in inter-molecular interactions [**138**]. The presence of this CxxC motif is interesting when you consider that CsgC contains a CxC motif which has been suggested to perform a functional role [**106**]. FapB's repeats are similar to the core repeat of FapC with a X_{15}- G – X_4 – N – X_3 – G – X_6 – N- X_7 motif (**Figure 1.17, Figure 1.18**), while the third repeat still has a similar motif it is more conserved and notably in the first position in **Figure 1.18C** where the other repeats generally contain a negatively charge Q/D residue the third repeat generally contains a positively charged arginine. It is possible that this positive charge is important for cell surface association because conserved positive charges in the final repeat of CsgB were found to be required for cell surface association [**118**].

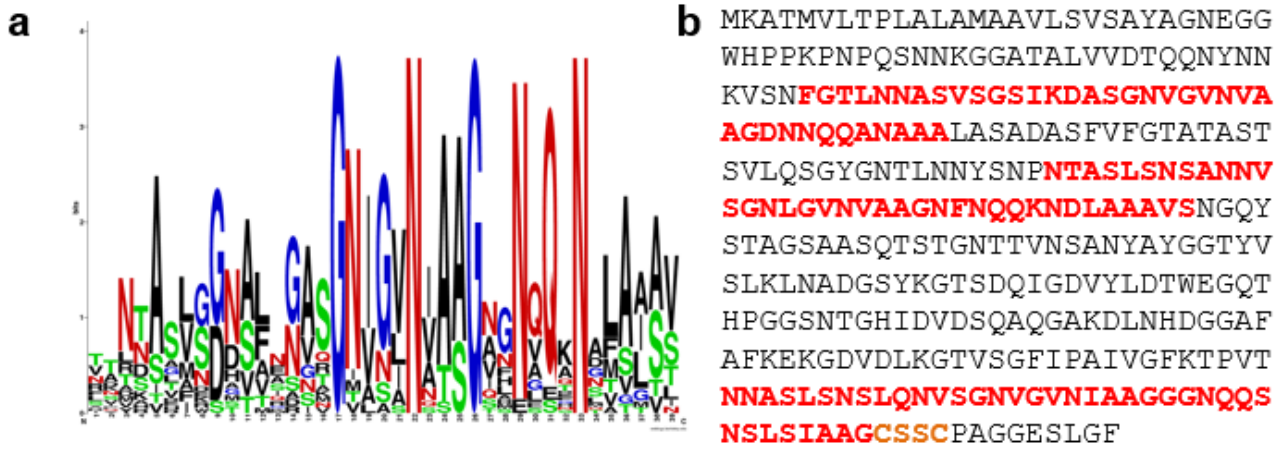


Figure 1.17: Amyloid Repeat Motif of FapC. a, LOGO Diagram illustrating the conservation of residues in the three repeats of FapC Homologs were identified using PSI-BLAST [**110**], aligned with clustal omega [**121**] and then their conservation plotted using webLOGO software. **b**, Sequence of FapC from *Pseudomonas aeruginosa* PAO1 with the amyloid repeats illustrated in red, the CxxC motif is indicated in orange. The loop region between the second and third amyloid repeats is particularly large.

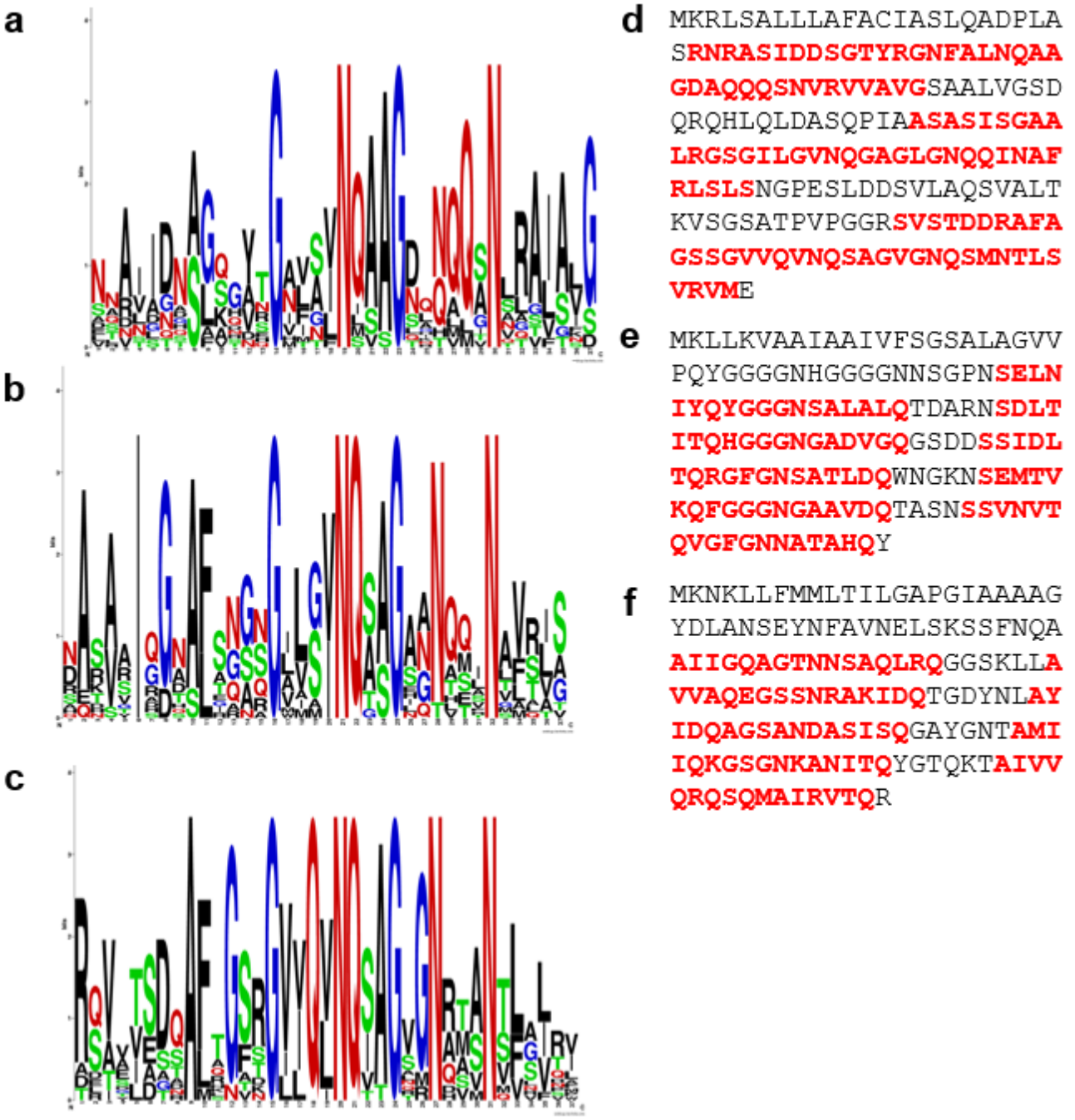


Figure 1.18: Amyloid Repeat Motifs of FapB. a, Diagram illustrating conservation of residues in the first repeat of FapB. **b**, Diagram illustrating conservation of residues in the middle repeat of FapB. **c**, Diagram illustrating conservation of residues in the last repeat of FapB. Homologs were identified using PSI-BLAST [**110**], aligned with clustal omega [**121**] and then their conservation plotted using webLOGO software. **d**, Sequence of FapB from *Pseudomonas aeruginosa* PAO1 with the three amyloid repeats illustrated in red. **e,** Sequence of CsgA from *E.coli* with the five amyloid repeats illustrated in red.**f,** Sequence of CsgB from *E.coli* with the five amyloid repeats illustrated in red.

Although the Curli and Alf systems are genetically distinct both of the nucleators possess a set of repeats similar in number and sequence to the main amyloid component but with a more conserved and slightly different C-terminal repeat containing conserved positive charge. The sequence of the repeats is very distinct from the repeats found in CsgA however it can be observed that all four motifs themselves contain a Q/N – X_{10} – Q/N motif which is perhaps important in determining the similar behaviour and ultrastructure of the two fibres (**Figure 1.19**).

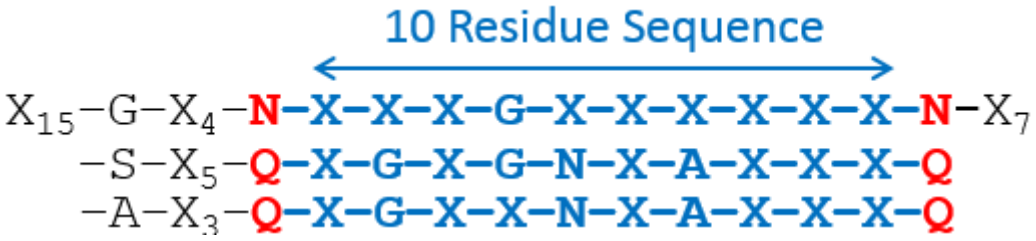


Figure 1.19: Potential conserved amyloid repeat between the Alf and Curli amyloid systems. Core amyloid motifs of FapB, FapC, CsgB and CsgA have been aligned to show a potential Q/N – X_{10} – Q/N motif common to all four proteins.

FapD and FapF, were shown to possess homology to proteins of known function: FapF showed strong homology to β-barrel membrane proteins and presumably is the pore through which the amyloid fibre components exit the cell, analogous to CsgG of the curli system; FapD appeared to be a member of the C39-like peptidase family, this family of cysteine proteases are often found in ABC transporters and are involved in bacteriocin processing and Quorum sensing cleaving a double-glycine motif [**92** , **139, 140, 141**]. Notably a double-glycine motif was identified in FapF by Dueholm et al., 2010 which the authors suggested may encode some form of signal peptide or bacteriocin **[92].** This prediction should be treated cautiously however as even using a generously broad double glycine motif (L-S-X_2-E-L-X_2-I-X-G-G) to search for the site it could not be identified in many FapF sequences from various strains of *Pseudomonas*. It should be noted however that some C39 peptidases have been shown to be inactive proteolytically but are believed to have a role in substrate recognition **[142]** so FapD could be acting as a chaperone or it could be processing and/or regulating FapF or other Alf components.

FapA and FapE did not have sufficient homology to any proteins of known function to help identify their function; however a point mutant, introducing a premature stop codon, in FapA was shown to significantly alter the properties of the mature fibres, with FapB becoming the predominant and FapC the minor component of the resulting amyloid, which appeared as short fibril-like aggregates [**138**]. Study of the composition of the Alf amyloid using mass spectrometry showed that small amounts of FapE were present in the mature fibre, possibly modulating the function [**138**].

Although our present knowledge of the Alf system is limited the predicted functions and current functional data can be used together with our knowledge of the curli system to suggest a

basic view as to the molecular basis for *Pseudomonas* Alf formation (**Figure 1.20**). The Alf components are transported across the inner membrane via the SEC machinery, in the periplasm the fibre components are processed, chaperoned or recognised by a combination of FapA, FapD and FapF. FapD may also process FapF to release a small signal peptide or bacteriocin which is likely to be secreted into the extracellular space. FapB, FapC and FapE are found in the amyloid fibres and are presumably secreted into the extracellular space via FapF. On the surface of the cell FapC is probably nucleated by FapB in a process similar to the curli system, FapE associates with the amyloid fibre in some way. Notably there is no obvious equivalent to CsgF and so the means by which FapB associates with the cell is unclear. Further study of this alternative Type VIII secretion system should provide useful insight into the operation of this devoted export machinery and bacterial functional amyloid.

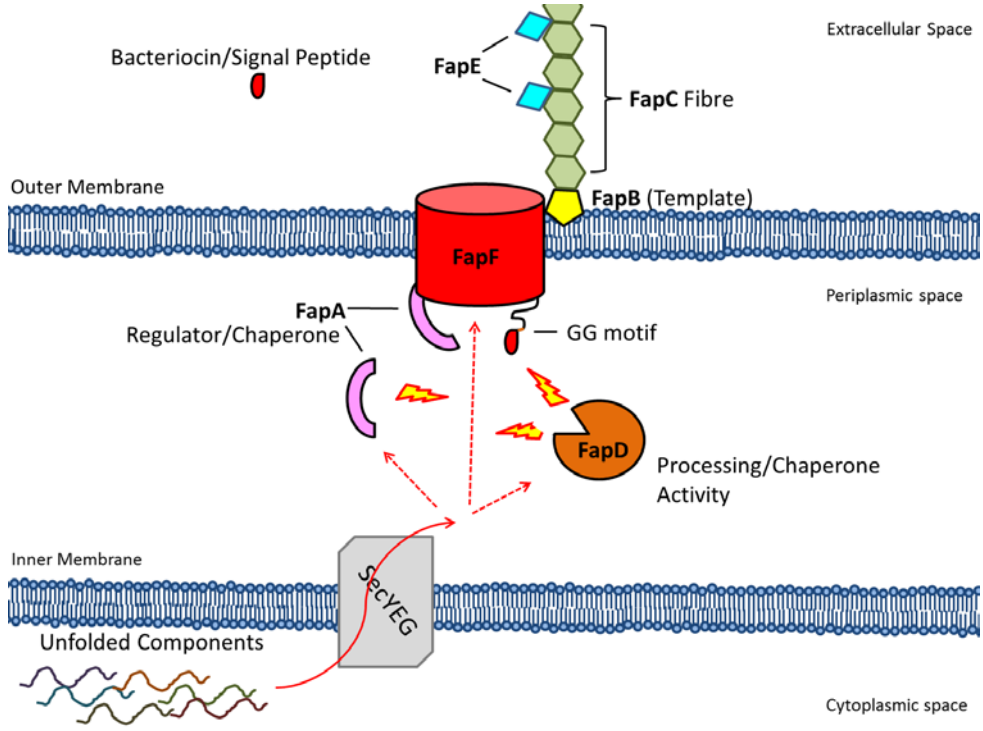


Figure 1.20: Speculative Cartoon Diagram illustrating the suggested molecular operation of the Alf system. The suggested roles of the Alf components is based on the functional roles suggested by Dueholm et al., 2010 [**92**] & Dueholm et al., 2013 [**138**] and comparison with the curli system (**Figure 1.8**). The Alf components are in to the periplasm via the SEC machinery, they are then processed, chaperoned or recognised by a combination of FapA, FapD and FapF. FapD may process FapF to release a signal peptide or bacteriocin. FapB, FapC and FapE are then secreted into the extracellular space via FapF where FapB nucleates the FapC fibre in a process similar to the curli system. FapE associates with the amyloid fibre.

1.6 Project Aims

1.6.1 *Pseudomonas* Alf

Current knowledge on the structure of *Pseudomonas* Alf components is limited with little known about the role of the various proteins encoded by the *fap*ABCDEF operon. As a starting point for structural studies we have aimed to produce purified proteins recombinantly and used these samples to probe the structure and function of Alf components. Since many components of the system were likely to be unstructured and challenging to study with structural techniques we looked to focus on structured components for NMR and crystallography.

1.6.2 CsgH

CsgH was recently identified as a novel component of the curli system which may be related to the amyloid inhibitory protein CsgC. In this research we have investigated whether the tertiary structure and function of CsgH are similar to CsgC by solving the structure of CsgH using NMR and testing the wild type protein and mutants' ability to inhibit CsgA polymerisation.

2 Materials and Methods

2.1 Materials

2.1.1 Plasmids

pET28c, pRSF-1B and pET46EK-LIC plasmids were all provided by novagen. pNIC28-Bsa4, pNIC-CTH, pNH-TrxT and pNIC-GST plasmids were kindly provided by the Structural Genomic Consortium. The pMMB190Ap:UK4fapA-F full operon plasmid [**92**] was generously provided by Professor Otzen. Dr Sebastian Lambert provided both the pHAT2*csgH*10-106 and pET28*csgH*10-106 plasmids. Dr Jonathan Taylor provided the pET28*csgA* plasmid. pLemo plasmid was extracted (2.2.7) from pLemo21 cells (NEB). pSHOM-1, a vector containing the OmpA sequence, was provided by Dr Stephen Hare.

2.1.2 Media

Lysogeny Broth (LB): (1 % (*w/v*) tryptone, 0.5% (*w/v*) Yeast Extract, 1 % (*w/v*) NaCl. Terrific Broth (TB): 1.2% (w/v) tryptone, 2.4% (w/v) Yeast Extract, 55 mM NaPO4, 15 mM KH2PO4, 50 mM NH4Cl, 5 mM Na2SO4, 10 mM NaCl, 2mM MgSO4, 0.5 % (w/v) Glycerol. M9 Minimal Media: 42.3 mM Na_2HPO_4, 22 mM KH_2PO_4, 0.7 M NaCl, pH 7-7.4, 13 mM NH_4Cl, 2 mM $MgSO_4$, 10 µM $CaCl_2$, 0.02% (*w/v*) Glucose, Vitamins solution (0.2 mg Choline Chloride, 0.5 mg Folic acid, 0.5 mg Pantothenic acid, 0.5 mg Nicotinamide, 1 mg Myo-inositol, 0.5 mg Pyridoxal HCl, 0.5 mg Thiamine HCl 0.05 mg Riboflavin, 1 mg Biotin), 1ml micronutrients solution (3 µM Ammonium Molybdate, 400 µM H_3BO_3 30 µM $CoCl_2$ 10 µM $CuSO_4$ 80 µM $MnCl_2$ 10 µM ZnSO4), 66 µM $FeSO_4$) supplemented with ^{15}N NH_4Cl, ^{13}C Glucose and/or Deuterium Oxide as appropriate. Bacterial plates were grown on LB media or M9 minimal media 1.5 % (*w/v*) agar plates as appropriate.

2.1.3 Cell Strains

Several different *Escherichia coli* strains were used for cloning and expression purposes. Primarily NEB5α (NEB) was used for cloning methods. For protein expression Shuffle (NEB), BL21 (NEB) and Lemo21 (2.2.9) were used as indicated in the results. *Pseudomonas aeruginosa* (PAO1) was used in genomic extraction (2.2.8) to provide the genomic DNA (gDNA) template for the Fap operon of genes from PAO1.

2.2 Molecular Biology

2.2.1 Cloning

2.2.1.1 Cloning Using LIC-system

2.2.1.1.1 In-House LIC-System

Polymerase chain reaction (PCR) was conducted using Pfu DNAP (Promega) according to the manufacturer's instructions, supplemented with 10 (*v/v*) % Dimethyl sulfoxide (DMSO), 0.4 ng/µl gDNA (*Pseudomonas* PAO1 Self-Extracted (**2.2.8**) or *Pseudomonas* PA7 provided by Dr James Garnett as appropriate and the appropriate primer pair from **Tables 2.1** and **2.2**. The product was visualised by agarose gel electrophoresis (**2.2.3**) DNA fragments of the correct length were gel extracted (**2.2.4**). 5 µl of the insert was treated with 1 unit T4 DNAP in NEB Buffer 2 (NEB), 10mM deoxycytidine triphosphate (dCTP), 2.5 mM DTT, 1 mg/ml Bovine Serum Albumin (BSA) and made up to 10 µl with ddH2O. The reaction was incubated at 22°C for 30 min and then the enzyme was denatured by incubation at 75°C for 20 min. 1 µg of pNIC vector was digested with BsaI-HF (NEB, UK) according to manufacturer's instructions. The digested vector was analysed by agarose gel electrophoresis (**2.2.3**) and then gel extracted (**2.2.4**) before treatment with T4 DNAP polymerase: 10 units T4 DNAP in NEB Buffer 2 (NEB), 10mM deoxyguanosine triphosphate (dGTP), 2.5 mM Dithiothreitol (DTT), 1 mg/ml Bovine Serum Albumin and made up to 100 µl with ddH2O. The vectors and inserts were mixed in 1:3 molar ratio and incubated for 1h at room temperature before transformation into DH5α (**2.2.6**). The transformed plasmids were then purified and sequenced (**2.2.7**).

2.2.1.1.2 Using pET46-EK LIC Kit

PCR was conducted using Q5 DNAP (NEB) according to the manufacturer's instructions, using 0.5 ng/µl *Pseudomonas* PA7 gDNA or pMMB190Ap:UK4fapA-F plasmid (UK4 genes) as appropriate using the appropriate primer pair from Table **2.3**.

2.2.1.2 Cloning Using Restriction Enzymes

PCR was conducted using Q5 DNAP (NEB) according to the manufacturer's instructions and using the appropriate primer pair **Table 2.4** and the appropriate template, *Pseudomonas* PA7 provided by Dr James Garnett, pMMB190Ap:UK4fapA-F for UK4 and pSHOM-1 for OmpA signal sequence. The product was visualised by agarose gel electrophoresis (**2.2.3**) DNA fragments of the correct length were gel extracted (**2.2.4**). The vector and inserts were treated with the appropriate restriction enzyme pair according to the manufacturer's instructions (NEB), the mixtures were incubated at 37°C for 3 h, the products were then visualised by agarose gel electrophoresis (**2.2.3**) DNA fragments of the correct length were gel extracted (**2.2.4**). The vectors and inserts were then combined at a 1:4 molar ratio and treated with T4 DNA ligase (NEB) according to the manufacturer's instructions. The mixture was incubated for 1 h at room temperature before transformation into DH5a (**2.2.6**). The transformed plasmids were then purified and sequenced (**2.2.7**).

Table 2.1: PCR Primers used for initial investigations into Alf in *Pseudomonas* PAO1. Annealing temperatures used for PCR are indicated. The LIC sequence is indicated in blue while the annealing sequence is indicated in black, stop and start codons are shown in lower case. These primers were used to produce pNIC-NTH, pNIC-GST and pNIC-TRX constructs which are vectors which provide an N-terminal poly-histidine tag (NTH), Glutathione S-transferase (GST) tag and thioredoxin (TRX) tag respectively. ΔSS indicates that the signal sequence has been removed.

Construct	Direction	Sequence 5' to 3'	Annealing
FapA	RV	TATCCACCTttaCTG tca TTG ACC CCC GAG C	54
FapA ΔNT	FW	TACTTCCAATCC atg GCG ACT CTC AAC CAA GGC	54
FapB	RV	TATCCACCTttaCTG tca TTC CAT GAC CCT GAC GCT	55
FapB ΔNT	FW	TACTTCCAATCC atg GCT CTG GTT CAC CTG CAC CAC G	55
FapC	RV	TATCCACCTttaCTG tca GAA GCC AAG GCT CTC GC	55
FapC ΔNT	FW	TACTTCCAATCC atg CTG GTT GCC GCC ACC GGC	55
FapD	RV	TATCCACCTttaCTG tca GAA GAA ATC GCT CTG GAT G	53
FapD	FW	TACTTCCAATCC atg CGC ACG CTC ATC CTT	53
FapE	RV	TATCCACCTttaCTG tca GTA GCC GCT GGG GCG	56
FapE ΔNT	FW	TACTTCCAATCC atg GAG CAT CAC CTC GAG CGC CGC	56
FapF	RV	TATCCACCTTTACTG tca GAA GTA GTA GGG GAA TTT C	55
FapF ΔNT	FW	TACTTCCAATCC atg GGG CAC GAT GCT GAA CCT GTT G	55
FapF FL	FW	TACTTCCAATCC atg ACC CAG ACA CTC	55

Table 2.2: Table of PCR Primers for pNIC-NTH *Pseudomonas* PAO1 constructs. The LIC sequence is indicated in blue while the annealing sequence is indicated in black, stop and start codons are shown in lower case. The template used was the appropriate gene cloned into pNIC-NTH using the primers in **Table 2.1**. These primers were used to produce pNIC-NTH constructs. Annealing temperature was 55˚C for all primers. Bracketed numbers indicate specific amino acid residue the primer commences at it the native sequence.

Primer	Direction	Sequence
FapA (24)	FW	TACTTCCAATCC atg CTCCCCGCGCATGCGGC
FapA (120)	RV	TATCCACCTttaCTG GACCGTCGCCGGGTTCGCC
FapA (98)	RV	TATCCACCTttaCTG GCGAATCCCGGAACCGCTGG
FapA (113)	RV	TATCCACCTttaCTG CTGGCTGGCGAAACCCGGC
FapA (86)	RV	TATCCACCTttaCTG GGCATCGCCCAGTTCGCGG
FapB (80)	FW	TACTTCCAATCC atg CAG CCG ATC GCA GCC AGC
FapB (86)	RV	TATCCACCTttaCTG GGCGCTGGCTGCGATCG
FapB (120)	RV	TATCCACCTttaCTG GCTGAGACTCAGCCGGAAGGC
FapB (137)	RV	TATCCACCTttaCTG CAGCGCCACGCTTTGCGC
FapB (100)	RV	TATCCACCTttaCTG GGC AGC GGC ATC CTC GGC
FapB (100)	FW	TACTTCCAATCC atg GGCGTCAACCAGGGCGCCG
FapC (268)	RV	TATCCACCTttaCTG GTCGCCCTTTTCCTTGAAGGCGAAC
FapC (268)	FW	TACTTCCAATCC atg GACGTCGACCTGAAAGGCACGG
FapC (149)	FW	TACTTCCAATCC atg CTG GGC GTG AAC GTC GCC G
FapC (148)	RV	TATCCACCTttaCTG GTTGCCCGAGACGTTGTTGGCC
FapC (320)	RV	TATCCACCTttaCTG GTTGCTCTGCTGGTTGCCGCC
FapC (82)	FW	TACTTCCAATCC atg GTCGGGGTCAACGTCGCCGC
FapC (77)	RV	TATCCACCTttaCTG ATCCTTGATCGAGCCGCTCACC
FapC (106)	FW	TACTTCCAATCC atg CAG CTT CGT GTT CGG CAC CG
FapC (171)	RV	TATCCACCTttaCTG GTTGGAGACGGCGGCGGCC
FapC (191)	FW	TACTTCCAATCC atg GTC AAC AGC GCC AAC TAC GCC TAT
FapC (300)	RV	TATCCACCTttaCTG CTGCAACGAGTTGCTCAGGCTG
FapC (125)	FW	TACTTCCAATCC atg AAT ACG CTG AAC AAC TAC TCC AAC CCC
FapE (90)	RV	TATCCACCTttaCTG GAGGACCGGGCGCGCCAG
FapE (150)	RV	TATCCACCTttaCTG CGGCGCCTCGCCGTGG
FapE (165)	FW	TACTTCCAATCC atg CTC AGC GGC GAG ACC GGC
FapE (185)	RV	TATCCACCTttaCTG GATCGCCAGTTGCAGGGCGC
FapF (130)	FW	TACTTCCAATCC atg CTCACCTACAGCCACTACGACACCC
FapF (99)	FW	TACTTCCAATCC atg GAC GAT GGC GCG CCC GC
FapD (23)	FW	TACTTCCAATCC atg TTTCCCGCCCTGCCCGGC
FapD (46)	FW	TACTTCCAATCC atg GAC CTG GTC GAG CAG AAG ACC G

Table 2.3: PCR Primers for pET46 (All annealed at 50 C). The LIC sequence is indicated in blue while the annealing sequence is indicated in black, stop and start codons are shown in lower case. These primers were used to produce pET46-EK/LIC constructs for the *Pseudomonas* UK4 strain unless otherwise indicated. ΔSS indicates that the signal sequence has been removed, ΔR indicates the numbered amyloid repeat has been removed, NTD indicates the N-terminal domain, bracketed numbers indicate specific amino acid residue the primer commences at it the native sequence. PA7 indicates that the gene was the *Pseudomonas* PA7 version of the gene.

Primer	Direction	Sequence
FapA ΔSS	FW	GAC GAC GAC AAG atg GAG GAC GGC ATC ATC GTG AC
FapA	RV	GA GGA GAA GCC CGG tta tca TGG CTG GAC CCC CAT CAT C
FapB ΔSS	FW	GAC GAC GAC AAG atg GAC AGC AAC AAT CAG GCC CTG
FapB	RV	GA GGA GAA GCC CGG tta tta CTT GAT AGT TAC ACC GAG GGT G
FapB ΔR1	FW	GAC GAC GAC AAG atg GCC AGT GCG GCA ATC CAG
FapB ΔR3	RV	GA GGA GAA GCC CGG tta tag AGG GGG ATC GCG AGT C
FapC ΔSS	FW	GAC GAC GAC AAG atg GGT CCT GCC GAA AAA TGG AAA CC
FapC	RV	GA GGA GAA GCC CGG tta tca GAA CGC CAG TTT GCT GC
FapC ΔR1	FW	GAC GAC GAC AAG atg AAG GGC ACG CAA AAC AAC GC
FapC ΔR2	RV	GA GGA GAA GCC CGG tta TTT GCC GTC AGC CGT CAC
FapD ΔSS	FW	GAC GAC GAC AAG atg GGG CAA ATG GCG ATT TCC GC
FapD	RV	GA GGA GAA GCC CGG tta cta GAA GAA GTC GCT CTG GAT G
FapD 58	FW	GAC GAC GAC AAG atg AGC ATC CGC GAA CGC CG
FapD CT 200	RV	GA GGA GAA GCC CGG tta CAC GGC GAA GAC GAT GCC
FapE ΔSS	FW	GAC GAC GAC AAG atg GCC TCG GCG TTC AAG CC
FapE	RV	GA GGA GAA GCC CGG tta tca ATA ACC GAG CG TGC GTA G
FapF ΔSS	FW	GAC GAC GAC AAG atg GCG CCA GAC GTG GAT ATC G
FapF	RV	GA GGA GAA GCC CGG tta tta GAA GTA GTACGGGAATTTCAGGC
FapF (107)	FW	GAC GAC GAC AAG atg AAG GAT GAT TCG GAG CCG GC
FapF (117)	FW	GAC GAC GAC AAG atg AGC AAC TTG TAC AAC GAA GCC AG
FapF NTD	RV	GA GGA GAA GCC CGG tta CTT GAG CGA TTG CCC GTA GG
FapD PA7 ΔSS	FW	GAC GAC GAC AAG atg ACG GTC GAC ACC AAG CCG
FapD PA7 (44)	FW	GAC GAC GAC AAG atg TCC GTG CAG CTG GAA CCG
FapD PA7 (57)	FW	GAC GAC GAC AAG atg AAC GTC ATC CGC CAG GCC
FapD PA7 CT (243)	RV	GA GGA GAA GCC CGG tta ACC GCT GCC CAA CTC ACG G
FapD PA7	RV	GA GGA GAA GCC CGG tta CTA CTT GGT GCG CAG GTG TAT C
FapF PA7 ΔSS	FW	GAC GAC GAC AAG atg GAG GAG GCC TCG GTC GAC C
FapF PA7 (45)	FW	GAC GAC GAC AAG atg GAG GAG GCC TCG GTC GAC C
FapF PA7 (56)	FW	GAC GAC GAC AAG atg AAG GCG CTG GAA GAG GTC TTC
FapF PA7 (63)	FW	GAC GAC GAC AAG atg AAG AGC TAC ACC TTG CTG AAG AAG
FapF PA7	RV	GA GGA GAA GCC CGG tta tta CTG GGC CTT CAG CCC TTT G

Table 2.4: PCR Primers for Restriction Cloning. (All annealed at 50˚C). The restriction sites are colour coded: green for NcoI, purple for SpeI, blue for HindIII, orange for BamHI; the his tag encoding sequence is coloured red and the G/C overhang residues to improve enzyme binding are shown in pink, while the annealing sequence is indicated in black capitals, stop and start codons are shown in lower case. ΔSS indicates that the signal sequence has been removed, ΔR indicates the numbered amyloid repeat has been removed, NTD indicates the N-terminal domain, bracketed numbers indicate specific amino acid residue the primer commences at it the native sequence.

Primer	Direction	Sequence
WJH_OmpA	FW	GG CC atg GCC atg AAA AAG ACA GCT ATC GCG ATT G
WJH_OmpA_R1	RV	ATG ATG ATG ATG ATG ATG ACT AGT GGC CTG CGC TAC GGT AGC
WJH_OmpA_R2	RV	CC GGTACC ATGATGATGATGATGATG ACTAGT GG
FapA dSS	FW	GGCC GGTACC GAG GAC GGC ATC ATC GTG AC
FapA	RV	CCGG AAGCTT tca TGG CTG GAC CCC CAT CAT C
FapB ΔSS	FW	GGCC GGTACC GAC AGC AAC AAT CAG GCC CTG
FapB	RV	CCGG AAGCTT tta CTT GAT AGT TAC ACC GAG GGT G
FapB ΔR1	FW	GGCC GGTACC GCC AGT GCG GCA ATC CAG
FapB ΔR3	RV	CCGG AAGCTT tta TAG AGG GGG ATC GCG AGT C
FapC ΔSS	FW	GGCC GGTACC GGT CCT GCC GAA AAA TGG AAA CC
FapC	RV	CCGG AAGCTT tca GAA CGC CAG TTT GCT GC
FapC ΔR1	FW	GGCC GGTACC AAG GGC ACG CAA AAC AAC GC
FapC ΔR3	RV	CCGG AAGCTT tta TTT GCC GTC AGC CGT CAC
FapD ΔSS	FW	GGCC GGTACC GGG CAA ATG GCG ATT TCC GC
FapD	RV	CCGG AAGCTT cta GAA GAA GTC GCT CTG GAT G
FapE ΔSS	FW	GGCC GGTACC GCC TCG GCG TTC AAG CC
FapE	RV	CCGG AAGCTT tca ATAACCGAGCGTGCGTAG
FapF ΔSS	FW	GGCC GGTACC GCGCCAGACGTGGATATCG
FapF	RV	CCGG AAGCTT tta GAAGTAGTACGGGAATTTCAGGC
FapF (107)	FW	GGCC GGTACC AAG GAT GAT TCG GAG CCG GC
FapF (117)	FW	GGCC GGTACC AAG GAT GAT TCG GAG CCG GC
FapD PA7 ΔSS	FW	GGCC GGTACC ACG GTC GAC ACC AAG CCG
FapD PA7	RV	CCGG AAGCTT tta CTA CTT GGT GCG CAG GTG TAT C
FapF PA7 ΔSS	FW	GGCC GGTACC GAG GAG GCC TCG GTC GAC C
FapF PA7	RV	CCGG AAGCTT tta tta CTG GGC CTT CAG CCC TTT G

2.2.2 Mutagenesis

2.2.2.1 Quikchange Mutagenesis

PCR was conducted using Pfu DNAP (Promega) according to the manufacturer's instructions using the appropriate plasmid and primer pairs from table **2.5**. The PCR product was purified by PCR purification (**2.2.5**). 10 units DpnI were then added to the eluted PCR products and incubated at 37°C for 1 h. 1 µL of the mixture was then transformed into NEB5α (**2.26**). The transformed plasmids were then purified and sequenced (**2.2.7**).

Table 2.5: Quikchange mutagenesis primers for pET46-Fap plasmids. Genes from UK4 unless otherwise indicated. Annealing step of PCR was conducted using touchdown PCR from 72 to 55 C and with the inclusion of 10 % DMSO in the Reaction.

Primer	Direction	Sequence
FapA M43-	FW	CAATGTGCAAGGCTTCTAAGTCGGCCGGCCTTCC
FapA M43-	RV	GGAAGGCCGGCCGACTTAGAAGCCTTGCACATTG
FapB M17-	FW	CTTGGCTGCTCAGCCGCCTAAGCCGACAGCAACAATCAG
FapB M17-	RV	CTGATTGTTGCTGTCGGCTTAGGCGGCTGAGCAGCCAAG
FapC M18-	FW	CTCTGGCTGCGCTCTAAGCTGTTGCTGCACAG
FapC M18-	RV	CTGTGCAGCAACAGCTTAGAGCGCAGCCAGAG
FapC C237A C240A	FW	CACCCTGGGCAGCGGCGCAACAGTAGCAGCAGCGGGCACTGGCAG
FapC C237A C240A	RV	CTGCCAGTGCCCGCTGCTGCTACTGTTGCGCCGCTGCCCAGGGTG
FapD M46-	FW	CAAATGGCGATTTCCGCCTAACCCGGCGGTGCGGTGATC
FapD M46-	RV	GATCACCGCACCGCCGGGTTAGGCGGAAATCGCCATTTG
FapD C76A	FW	GAAAACCGATTTCAGCGCAGGTGCTGCTGCCCTC
FapD C76A	RV	GAGGGCAGCAGCACCTGCGCTGAAATCGGTTTTC
FapE M60-	FW	GTTTCGGCATTGTCTAGAGTACCACCTGGAC
FapE M60-	RV	GTCCAGGTGGTACTCTAGACAATGCCGAAAC
FapE C241A	FW	GGGGGCCTTGAACGCCAATCTGGATCAAC
FapE C241A	RV	GTTGATCCAGATTGGCGTTCAAGGCCCCC
FapF Y23-	FW	GAGTCTGTTGTAGGCAGCGCCAGAC
FapF Y23-	RV	GTCTGGCGCTGCCTACAACAGACTC
FapF PQ213AA	FW	GCCAGTGGCGGAGACGCAGCGGCCACTTCAGAAGAG
FapF PQ213AA	RV	CTCTTCTGAAGTGGCCGCTGCGTCTCCGCCACTGGC
FapF EE218AA	FW	CCCAGGCCACTTCAGCAGCAAGCGTGAGCCGGGATC
FapF EE218AA	RV	GATCCCGGCTCACGCTTGCTGCTGAAGTGGCCTGGG
FapD PA7 C67A	FW	GGCCTACGACTACAGCGCAGGCTCGGCGGCCCTTAC
FapD PA7 C67A	RV	GTAAGGGCCGCCGAGCCTGCGCTGTAGTCGTAGGCC

2.2.2.2 Q5 Mutagenesis

Q5 mutagenesis was carried out using the Q5 mutagenesis kit (NEB) according to the manufacturer's instructions using the primer pairs described in **table 2.6**.

Table 2.6: Q5 mutagenesis primers for pET28*csg*H$^{10\text{-}106}$ plasmid. Designed and used according to the manufacturer's instructions (NEB). Capital letters indicate annealing sequence, lower case the sequence encoding the substitution(s).

Primer	Direction	Sequence
K32A	FW	AGGTAGCTATgcaCTGGCAGTTGATAAAG
K32A	RV	GCACCATCACGAACTGCG
R45S	FW	AGGCACCAGCagtATTAAACAGG
R45S	RV	GCTGCACCGGCTTTATCA
K47E	FW	CAGCCGTATTgaaCAGGCAGGCG
K47E	RV	GTGCCTGCTGCACCGGCT
D35S	FW	ACTGGCAGTTtctAAAGCCGGTG
D35S	RV	TTATAGCTACCTGCACCATC
K32A D35S	FW	agtttctAAAGCCGGTGCAGCAGGC
K32A D35S	RV	gccagtgcATAGCTACCTGCACCATCACG
E57K	FW	CGCAATTGCAaaaCAGCGTGTTA
E57K	RV	GTAAATGCGCCTGCCTGTTTAATA
D93A	FW	GTGTAATCTGgctCCGGAAACCG
D93A	RV	TGAATGGTAACGCTACCAAAG
E95K	FW	TCTGGATCCGaaaACCGTTAAATAAG
E95K	RV	TTACACTGAATGGTAACGC
R25E	FW	TACCGCAGTTgaaGATGGTGCAG
R25E	RV	ATAACACCCTGCAGTGTAAC
N64W	FW	TACCGTTGGTtggGTTGTTCTGGATTATAG
N64W	RV	ACACGCTGTTCTGCAATTG
E95K	FW	GTGTAATCTGaaaCCGGAAACCGTTAAATAA
E95K	RV	TGAATGGTAACGCTACCAAAG
F83A	FW	GGATGTTAGCgctGGTAGCGTTACC
F83A	RV	AGACGTGCTGCATAACGA
R45S K47E	FW	tgaaCAGGCAGGCGCATTTACCG
R45S K47E	RV	atactGCTGGTGCCTGCTGCACC
K32A D35S R45S K47E	FW	GTGATGGTGCAGGTAGCTATgcactggcagtttcaaaagccgg tgcagcaggcaccagctccattgatCAGGCAGGCGCATTTACC
K32A D35S R45S K47E	RV	CACTACCACGTCCATCGATAcgtgaccgtcaaagttttcggcca cgtcgtccgtggtcgaggtaactaGTCCGTCCGCGTAAATGG
R45K	FW	AGGCACCAGCaaaATTAAACAGGCAG
R45K	RV	GCTGCACCGGCTTTATCA
H14A	FW	TAGCGGTGGTgctGTTACACTGC
H14A	RV	ACTGCTGCTTCAACTTCAC

2.2.3 Agarose Gel Electrophoresis

Gels were composed of 1 % (*w/v*) agarose in Tris-acetate-EDTA (TAE) Buffer (40 mM Tris-acetate, 1 mM Ethylenediaminetetraacetic acid (EDTA), pH 8.0), 0.05 µl/ml sybrsafe) were used to visualise DNA and separate it by molecular weight. Samples were loaded with 6 x loading buffer (40% (*w/v*) sucrose, 0.25% (*w/v*) bromophenol blue). The gels were run in TAE buffer at 120 V. Size and concentration of DNA were estimated by running 2-Log (NEB), 1KB (NEB) and 100bp (NEB) ladders as appropriate.

2.2.4 Gel Extraction

DNA was extracted from agarose gels using GeneJet Gel Extraction Kit (Thermo Scientific) according to the manufacturer's instructions.

2.2.5 PCR Purification

PCR products were purified using the GeneJet PCR Purification Kit (Thermo Scientific) according to the manufacturer's instructions.

2.2.6 Transformation

The appropriate competent cells and plasmids were mixed and incubated on ice for 20 min. The cells were then heat shocked at 42°C for 1 min, incubated on ice for 10 min. The cells were mixed with LB and incubated for 1h at 37°C before plating on LB/M9 agar plates overnight at 37°C.

2.2.7 Plasmid Purification and Sequencing

Plasmids were purified using GeneJet plasmid miniprep kit (Thermo Scientific) according to the manufacturer's instructions and sequenced by Source Bioscience (UK) or GATC Biotech (Germany) using their T7 and pET-reverse primers.

2.2.8 Genomic DNA Extraction

gDNA was purified using Genomic DNA Purification Kit (Thermo Scientific) according to the manufacturer's instructions.

2.2.9 Making pLemo21 cell strain for expression.

pLemo plasmid was extracted from the purchased pLemo21 strain (NEB) (**2.2.7**) and transformed into BL21 DE3 cells (NEB) (**2.2.6**). The cells were then made chemically competent for transformation (**2.2.10**)

2.2.10 Making Competent Cells

The selected strain of *E. coli* was streaked onto a LB-agar plate with no-antibiotic and grown overnight at 37 °C. A single colony was then picked from the plate and inoculated into 10 mL LB media and incubated shaking overnight at 37 °C. The overnight culture was then used to inoculate 500 mL of LB at 1:100 dilution. The cells were then grown to an Optical Density at 600 nm (OD_{600}) of approximately 0.4 before pelleting by centrifugation 3000g 10 min 4 °C. The cell pellet was resuspended on ice into cold 0.1 M CaCl and incubated on ice for 30 min. The cells were then pelleted again at 3000g 10 min 4 °C and then resuspended on ice into cold 0.1 M CaCl 15 % Glycerol. The resuspended cells were then aliquoted in 100 µL volumes into 1.5 mL microcentrifuge tubes and flash frozen in liquid nitrogen to be thawed when needed for transformation.

2.3 Protein Expression and Purification

2.3.1 Protein Expression Trials

10 mL LB cultures were inoculated with the desired expression strain transformed with the appropriate plasmid and incubated overnight (~16h) at 37 °C. The OD_{600} of the overnight was measured and diluted to a standard OD_{600} of 0.05 into wells containing 5ml LB media with the appropriate antibiotic (and the appropriate concentration of Rhamnose if Lemo21) in a 24-deepwell block (GE healthcare). The cultures were incubated shaking at 37 °C until they reached OD_{600} ~0.5, the temperature was then reduced to the induction temperature and the cultures were induced with 0.1-1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6-0.8 and incubated overnight at the induction temperature. 1 mL samples were then taken and analysed using Bacterial Protein Extraction Reagent (Thermo Scientific) according to the manufacturer's instructions to fractionate soluble and insoluble expression. The resulting samples were then analysed by Sodium dodecyl sulfate- Poylacrylamide Gel Electrophoresis (SDS-PAGE) (2.4.1).

2.3.2 Large Scale Protein Expression

2.3.2.1 Unlabelled Proteins

10 mL LB cultures were inoculated with the desired expression strain transformed with the appropriate plasmid and incubated overnight (~16h) at 37 °C. These overnights were inoculated 1:1000 into LB or TB media with the appropriate antibiotic. The cultures were incubated shaking at 37 °C until they reached OD_{600} 0.6-0.8, they were then cooled to induction temperature and induced with 0.5-1 mM IPTG and incubated overnight at the induction temperature.

2.3.2.2 Isotopically Labelled Proteins

100 mL M9 cultures were inoculated with the desired expression strain transformed with the appropriate plasmid and incubated overnight (~16h) at 37 °C. . These overnights were inoculated 1:100 into M9 media with the appropriate antibiotic and labelled components. The cultures were incubated shaking at 37 °C until they reached OD_{600} 0.6-1.0, they were then cooled to induction temperature and induced with 0.5-1 mM IPTG and incubated overnight at the induction temperature.

2.3.3 Standard Protein Purification

Cells were pelleted at 4500x, 6 °C for 15 minutes, resuspended in resusupension buffer (300 mM NaCl, 50 mM TrisHCl, pH 8), with phenylmethanesufonylfluoride (PMSF) or β-mercaptoethanol as required, and the cells lysed by sonication at 70 % amplitude for 15 min. The lysate was pelleted at 15,000x, 4 °C for 1 h. The supernatant was then used for native purifications. While the pellet was dissolved in 8 M urea or 8 M guanadinium hydrochloride overnight for denaturing purifications and then clarified by spinning down at 15,000x 16 °C for 30 min. The

proteins of interest were then purified from these supernatants using Ni affinity chromatography 2.3.7.

2.3.4 Purification of FapF

2.3.4.1 Denaturing Purification

The protein was expressed using the standard method (2.3.2) with an induction temperature of 30 °C and an IPTG concentration of 1 mM. Cells were pelleted at 4500x, 6 °C for 15 minutes, resuspended in resusupension buffer (300 mM NaCl, 50 mM TrisHCl, 1 mM PMSF, pH 8) and lysed using a Constant Systems TS series cell disruptor (Constant Systems) at 25 kPsi. The lysate was pelleted at 15,000x, 4 °C for 1 h, the supernatant was then discarded and the pellet dissolved in 8 M urea overnight and then clarified by spinning down at 15,000x 16 °C for 30 min. The protein was then refolded by pulse refolding (2.3.8.1) into 5 % n-dodecyl-N,N-Dimethylamine-N-Oxide (LDAO), 300 mM NaCl, 20 mM TrisHCl pH 8, PMSF. The refolding mixture was then dialysed into 0.5% LDAO, 300 mM NaCl, 20 mM TrisHCl pH 8, PMSF overnight 8 °C. Refolded FapF was then purified out by gel filtration (2.3.9) using an HiLoad 16/600 superdex S200 column (GE Healthcare) into 0.1 % LDAO, 300 mM NaCl, 20 mM TrisHCl pH 8.

2.3.4.2 Membrane Purification

10 mL LB cultures were inoculated with the Lemo21 cells transformed with the appropriate plasmid and incubated overnight (~16h) at 37 °C. These overnights were inoculated 1:1000 into autoinducing TB medium 100 μg/ml carbenicillin, 25 μg/ml chloramphenicol, 0.2 mM rhamnose. The cultures were incubated shaking at 37 °C for 4 hours before reducing temperature to 25 °C for incubation overnight. Cells were pelleted at 4500x, 6 °C for 15 minutes, resuspended in 300 mM NaCl, 50 mM TrisHCl, 1 mM PMSF, pH 8 1 mM PMSF 10 μg/ml DNase I. Cells were lysed using the Constant Systems TS series cell disruptor (Constant Systems) at 25 kPsi. The lysate was pelleted at 15,000x, 4 °C for 1 h, discard the pellet and transferred the supernatant to ultracentrifuge tubes, spin down at 41,000x 1 hour, discard supernatant, resuspened the pellet in in 300 mM NaCl, 50 mM TrisHCl 0.5 % (*w/v*) sarcosine. Incubate shaking for 1 hour 8 °C. Ultracentrifuge again 41k 1.5 hours, discard s/n and resuspend pellet in 1% (*w/v*) detergent, 200 mM NaCl, 20 mM TrisHCl pH 8.0, incubate overnight, stirring 8 °C. Centrifuge 41k 25 °C 30 mins. The proteins of interest were then purified from these supernatants using Ni affinity chromatography 2.3.7 with the inclusion of the appropriate detergent at 2x critical micelle concentration (CMC).

2.3.5 Purification of CsgA

The protein was expressed using the standard method (2.3.2) with an induction temperature of 37 °C, 3h incubation after induction and an IPTG concentration of 0.5 mM. Cultures were separated into 300 mL aliquots and cells pelleted at 4500x, 4 °C for 15 minutes, and flash frozen in liquid nitrogen. On the day CsgA was required the cells were thawed and resuspended in 300 mL Guanadinium HCl, 50mM KPO4 pH 7.8, 100 mM NaCl, sonicated and then 70% Amplitude 5 min before incubation shaking at room temperature for 30 min. The lysate was then clarified by centrifugation 15,000x 16 °C 30 min and sonicated again at 70% Amplitude 5 min. 500 uL TALON Metal Affinity Resin (Clontech) was then added to the supernatant and incubated shaking at room temperature for 1 h. Resin was pelleted at 500g 3 min 16 °C, supernatant was discarded and the resin transferred to a polypropelene column, the resin was washed in 3 mL of 8 M Guanadinium HCl 100 mM NaCl 50 mM KPO4, the column was then moved to 4 °C cold room and washed in 3 mL 1.6 M Guanadinium HCl 100 mM NaCl 50 mM KPO4 pH 7.5 and then 3 mL 100 mM NaCl 50 mM KPO4 pH 7.5 2 mM imidazole. The protein was then eluted 200 uL at a time in 2 mL 100 mM NaCl 50 mM KPO4 pH 7.5 500 mM imidazole. The elution was filtered through a 30 kDa Molecular Weight Cut-off (MWCO) concentrator (Generon) 4500x, 4 °C centrifugation. The flow-through was then de-salted into 50 mM KPO4 pH 7.5 using a Hi-trap desalting column (GE Healthcare). The concentration of CsgA was measured by Absorbance at 280 nm using a nanodrop 1000 spectrophotometer (Thermoscientific) before use in assays.

2.3.6 Purification of CsgH

Cells were pelleted at 4500x, 6 °C for 15 minutes, resuspended in resusupension buffer (300 mM NaCl, 50 mM TrisHCl, 1 mM PMSF, pH 8). Cells were lysed either using the sonicator, with the addition before of 0.1 mg/ml Lysozyme at 70 % amplitude, 10 min (Unlabelled) or using the cell disruptor at 25 kPsi (Labelled) adding 0.1 mg/ml Lysozyme after lysis. Lysate was clarified by centrifugation 4500x 1h 4 °C and CsgH purified by Ni-affinity Purification (2.3.7). The elution was then further purified by gel filtration (2.3.9) using a HiLoad 16/600 superdex S75 pg column (GE Healthcare) using the desired buffer for further experiments.

2.3.7 Ni-affinity Purification

2mL Ni-NTA superflow resin (Qiagen) was equilibrated in 300 mM NaCl, 20 mM TrisHCl pH 8 the sample mixture containing the protein was applied to the column and allowed to pass through by gravity flow, the column was then washed with 20 mL 300 mM NaCl, 20 mM TrisHCl, 5 mM Imidazole and then the protein eluted in 5 mL 300 mM NaCl, 20 mM TrisHCl pH 8, 500 mM imidazole. For denaturing purification all three buffers also contained either 8 M urea or 8 M guanadinium HCl.

2.3.8 Protein Refolding

2.3.8.1 Pulse Refolding

Protein samples for refolding were purified under denaturing conditions with either 8 M Urea or 8 M guanadinium HCl. The purified sample was dripped into a stirring solution of the chosen refolding buffer diluting the protein around 20 fold. The resulting solution was then concentrated on a Ni-IDA (Bio-Rad, USA) column and eluted in the desired buffer with the addition of 500 mM imidazole. The sample was then gel filtered (**2.3.9**) to remove aggregates.

2.3.8.2 Matrix Assisted refolding

Protein samples for refolding were purified under denaturing conditions with either 8 M Urea or 8 M guanadinium HCl, the purified sample was bound to a Ni-IDA column and then washed in 300 mM NaCl, 20 mM TrisHCl pH 8 with decreasing concentrations of denaturant in 10 mL volumes: 8 M, 7 M, 6 M, 5 M, 4 M, 3 M, 2 M, 1 M, 0 M. The protein was then eluted with 300 mM NaCl, 20 mM TrisHCl pH 8, 500 mM Imidazole. The sample was then gel filtered (**2.3.9**) to remove aggregates.

2.3.8.3 Refolding by Dialysis

Protein samples for refolding were purified under denaturing conditions with either 8 M Urea or 8 M guanadinium HCl, the purified sample was placed in dialysis tubing (Spectrum, USA) and then dialysed into 100 fold volume buffer 300 mM NaCl, 20 mM TrisHCl pH 8 with decreasing concentrations of denaturant: 8 M, 7 M, 6 M, 5 M, 4 M, 3 M, 2 M, 1 M, 0 M. The resulting solution was then concentrated on a Ni-IDA (Bio-RAD, USA) column and eluted in the desired buffer with the addition of 500 mM imidazole. The sample was then gel filtered (2.3.9) to remove aggregates.

2.3.9 Gel Filtration

Gel filtration was conducted on AKTA purifier, AKTA prime or AKTA pure gel filtration systems (GE Healthcare) using the appropriate size exclusion column for the protein size in accordance with the manufacturer's instructions (GE Healthcare). The column was equilibrated in the appropriate buffer before use, the protein sample was then applied in around 1-4 mL volume and the flow rate was 0.5-1.5 mL/min depending on time constraints.

2.4 Analytical Techniques

2.4.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein samples for analysis were mixed with 6x loading buffer (250 mM Tris-HCl pH 6.8, 10% Sodium dodecyl sulfate (SDS), 0.05 (*w/v*) % bromophenol blue, 5 % (*v/v*) β-mercaptoethanol and 30 % (*v/v*) glycerol) to produce a 1x final concentration. The mixture was heated at 100 ºC for 5 minutes and spun down before loading onto the gel. To allow MW to be estimated the gel was also loaded with Mark12™ Standard (Invitrogen, USA) or Protein Marker Broad Range (2-212 kDa) (NEB).Gels used were either self-cast and composed of stacking (4.5% polyacrylamide (Acrylamide:Bisacrylamide = 37.5:) 0.35 M bis-Tris HCl pH 6.7, 0.67% (w/v) ammonium persulphate (APS), 0.33% (v/v) Tetramethylethylenediamine (TEMED)) and a resolving gel (12% polyacrylamide (Acrylamide:Bisacrylamide = 37.5:) 0.35 M bis-Tris HCl pH 6.7, 0.38% (w/v) APS, 0.1% (v/v) TEMED); or were pre-cast 10, 12 or 16 % RunBlue gels (Expedeon). The self-cast gels were run in MOPS buffer (250 mM MOPS, 250 mM Tris, 5 mM EDTA , 0.5% (*w/v*) SDS, pH 6.5) at 150 V and the precast gels were run in accordance with the manufacturer's instructions (Expedeon). To visualise the protein, gels were stained with InstantBlue (Expedeon).

2.4.2 Western & Dot Blotting

For western blotting SDS-Gels were blotted onto Polyvinylidene fluoride (PVDF) membrane (GE Healthcare, USA) in transfer buffer (Tris-glycine-SDS (TGS) (25 mM Tris, 192 mM Glycine, 0.1 % (*w/v*) SDS, pH 8.6), 20% Methanol) at constant current of 400 mA for 1 hour in a BioRad Mini Trans-Blot Cell (BioRad, USA). For the dot blot 10 µL of sample was applied directly to the PVDF membrane and allowed to dry. For either method the membrane was then blocked with Phosphate Buffered Saline (PBS), 0.05 % (*v/v*) Tween20, 2 % (*w/v*) BSA) overnight, shaking at 4 °C, and then incubated with 1:1000 Anti-polyHistidine-Peroxidase labelled antibody in PBS, 0.05 % (*v/v*) Tween20 (Sigma, UK) for 1 h at room temperature excess antibody was removed by washing three times with PBS, 0.05 % (*v/v*) Tween20 10 min at room temperature. The membrane was then developed using ECL Western Blotting Reagents (GE Healthcare, USA) and imaged on a LAS-3000 Fujiimager (Fujifilm, Japan).

2.4.3 Limited Proteolysis

FapF was digested in a mixture containing 1 mg/ml FapF, 100 µg/ml of protease in 0.1 % (*w/v*) LDAO, 300 mM NaCl, 20 mM TrisHCl pH 8.0, samples were taken periodically as indicated and analysed by SDS-PAGE (2.4.1).

2.4.4 Size Exclusion Chromatography Multiple Anomalous Light Scattering (SEC-MALS)

Size Exclusion Chromatography Multiple Anomalous Light Scattering (SEC-MALS) was conducted using 1260 infinity (Agilent Technology), S200 analytical column (GE Healthcare), MiniDawn TREOS (Wyatt Technology) and OptiLab T-rEX (Wyatt Technology) in various buffer conditions described in results.

2.4.5 Amyloid Assays

2.4.5.1 ThT Fibrillation Assay

Freshly purified amyloid protein ~20 μM was combined with a stock solution of ThT (20 μM final concentration) and a concentrated stock of any other additives being used to produce 100 uL samples which were loaded on a 96 well clear bottomed plate for analysis of a spectromax spectrophotometer. The plate was incubated at 22 °C, with readings every 15 min with periodic shaking between readings. The emission was at 495 nm and the excitation was at 438 nm with a cut-off at 455 nm unless otherwise described.

2.4.5.2 Congo Red Assay

To observe amyloid formation in *E. coli* specialised YESCA agar plates (10 g/L casamino acids, 1 g/L yeast extract, and 20 g/L agar) were prepared with 50 μg/L Congo Red dye and varying concentrations of IPTG (0-1 mM). Cells were incubated over a few days at room temperature to allow time for cell growth and protein expression.

2.4.6 Circular Dichroism (CD)

The samples for analysis by Circular Dichroism were dialyzed into 20 mM phosphate buffer pH 8.0, 150mM NaF. The samples were then loaded into a quartz cuvette 100-QS with 1mm path length. The spectrum was obtained on Chirascan CD Spectrometer (Applied Photophysics) at 22°C, wavelengths 180 nm to 260 nm, with intervals 0.5 nm, 1 nm bandwidth and 1 sec time per point.

2.4.7 Differential Scanning Fluorimetry (DSF)

Protein samples to be analysed by Differential Scanning Fluorimetry (DSF) were prepared at 2 μM in their specific buffer to be tested and then combined with SYPRO orange (Invitrogen) according to the manufacturer's instructions to make 40 μL samples. Samples were transferred into a 96 well plate and analysed using a Stratagene Mx3005P (Agilent Technologies) with Excitation 492 nm and Emission 610 nm over a temperature range 25 °C to 95 °C.

2.5 Structural Techniques

2.5.1 NMR

2.5.1.1 NMR spectroscopy

The majority of NMR experiments were performed on a sample containing 0.4 mM protein in 10 mM MES (pH 6.5), 150 mM NaCl and 10% D2O, while the ^{13}C Nuclear Overhauser Effect Spectroscopy Heteronuclear Single Quantum Coherence (NOESY-HSQC) and H-H Nuclear Overhauser Effect Spectroscopy (NOESY) were carried out in 100 % D2O. All NMR spectra were acquired at 292 K on Bruker Avance-III HD 950, Avance-II 800 and Avance-III 600 spectrometers. Triple resonance HNCACB, CBCA, HNCO and HN(CA)CO spectra were used together with a ^{1}H-^{15}N HSQC spectrum to obtain the backbone assignments. Side chain chemical shift assignments were obtained using HBHA(CO)NH, CC(CO)NH and HCCH Total Correlation Spectroscopy (HCCH-TOCSY), ^{1}H-^{13}C HSQC and H-H NOESY spectra. Structural distance constraints were obtained from ^{15}N NOESY-HSQC, ^{13}C NOESY-HSQC and H-H NOESY spectra (each with a mixing time of 100 ms). Spectra were processed using NMRPipe [**143**].

2.5.1.2 Structure calculation

Spectra were analysed and chemical shifts were assigned using CcpNmr 2.4.0 [**144**], MARS 1.1 [**145**] and a modified version of NMRview 5.2.2 [**146**]. Dihedral angles were obtained from the chemical shift assignments using TALOS + [**147**], the resulting predictions were used for the structure calculation. NOE peaks were picked manually in CcpNmr 2.4.0 without any manual assignments. Automatic NOE assignment and structure calculation was carried out using ARIA 2.3 [**148**]/CNS 1.1 [**149**] software packages. An additional constraint was included to direct the formation of a disulphide bond between Cys4 and Cys90 in the protein structure. 300 structures were calculated in the final iteration and the 30 best structures were refined in water from which the 20 lowest-energy structures were selected and then validated using the MolProbity structure-validation web service [**150**, **151**]. The coordinates of CsgH have been deposited to the Protein Data Bank (www.rcsb.org) under PDB-ID accession number [**2N59**].

2.5.1.3 Calculating changes in chemical shift

The co-ordinates of the peaks in the two spectra being compared were recorded and the difference in Nitrogen chemical shift (δN) and proton chemical shifts (δH) extracted for each residue assigned in both spectra. The overall displacement of the peak (D) was calculated using **Equation 2.1**. Note that the difference in Nitrogen is weighted 0.14 relative to the proton, this has been suggested to be an appropriate general weighting to compensate for the difference in chemical shift range between the different nuclei [**153**].

Equation 2.1

$$D = \sqrt{\delta H^2 + 0.14\delta N^2}$$

2.5.2 X-ray Crystallography

2.5.2.1 Crystal Screens

2.5.2.1.1 Initial Screens

Crystal trays were laid in sparse matrix screens (Imperial College Screens ICL1-14) using a Mosquito Nanodrop crystallisation robot (TTP LabTech, UK) under conditions indicated in the results. Crystal trays were incubated at 20 or 4 °C as indicated.

2.5.2.1.2 Optimisation Screens

Crystal conditions from the initial screens were then used to design optimisation screens which were set up manually in 48 well MRC Maxi Plates (Molecular Dimensions) as indicated in the results.

2.5.2.1.3 Lipdic Cubic Phase (LCP)

Lipidic Cubic Phase (LCP) samples were prepared using 60 % (*w/w*) Monoolein using the method described by Caffrey & Cherezov 2009 [**154**] and were laid in sparse matrix screens (Imperial College Screens ICL6 & ICL9) using a Mosquito Nanodrop crystallisation robot (TTP LabTech, UK) and incubated at 20 °C.

2.5.2.1.4 Bicelles

Bicelles were prepared from 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine (DMPC) (Affymetrix) and 3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxyl-1-propanesulfonate (CHAPSO) (Affymetrix) at a 2.8:1 molar ratio to produce a 40 % (*w/v*) bicelle solution. To combine with detergent-solubilised FapF, the bicelle mixture was then added to the protein solution to produce an 8 % (*w/v*) bicelle, 0.05 % (*w/v*) LDAO solution.

3) Results

3.1 Bioinformatics *Pseudomonas aeruginosa* (PAO1) Alf

Little was known about the structure or function of most of the components of the Alf system, to help guide construct design and experimental approaches the sequences of the genes were used in a range of bioinformatics programs: Phyre2 [**155**], I-TASSER V4.3 [**156**], The PSIPRED Protein Sequence Analysis Workbench [**157**], BLAST [**110**], SignalP [**158**] and PRED-TMBB [**159**]; the results are summarised in this section.

3.1.1 FapA

FapA is predicted to be an unstructured protein with ~60 % of the protein predicted to be disordered by Phyre2 (secondary structure predictions are ~65 % accurate for proteins with limited homologs [**155**]), with no significant homologs of known function, limited secondary structure elements and a signal sequence from residues 1-28 (**Figure 3.1**). A construct was designed with the signal sequence removed and replaced with an NTH-tag for expression to the cytoplasm and to ease purification.

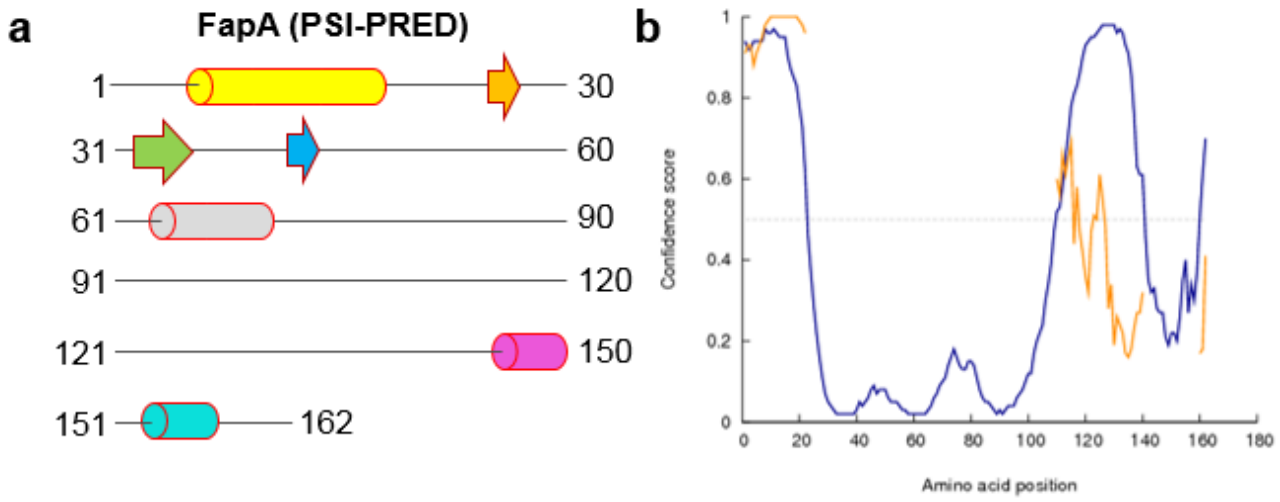


Figure 3.1: Summary of Bioinformatics results for FapA. a, Diagram illustrating PSIPRED **[126]** predicted secondary structure of FapA with helices indicated with a cylinder, B-strands as arrows and disordered regions as lines.**b**, Chart illustrating likelihood that the residues are in a disordered region of the protein generated using DISOPRED **[160]**. The likelihood that the residue is disordered is shown in blue, probability of a disordered region being involved in protein binding based on evolutionary sequence conservation and amino acid composition is shown in orange.

3.1.2 FapB

FapB is predicted to be an unstructured protein with ~66 % of the protein predicted to be disordered by Phyre2 (secondary structure predictions are ~65 % accurate for proteins with limited homologs [**155**]), with no significant homologs, few secondary structure elements and a signal sequence from residues 1-20 (**Figure 3.2**). Three amyloid repeats, could be identified by manual inspection based on the characteristic motif X_{15}- G – X_4 – N – X_3 – G – X_6 – N- X_7 (23-61, 82-120, 150-188). A construct was designed with the signal sequence removed and replaced with an NTH-tag for expression to the cytoplasm and to ease purification.

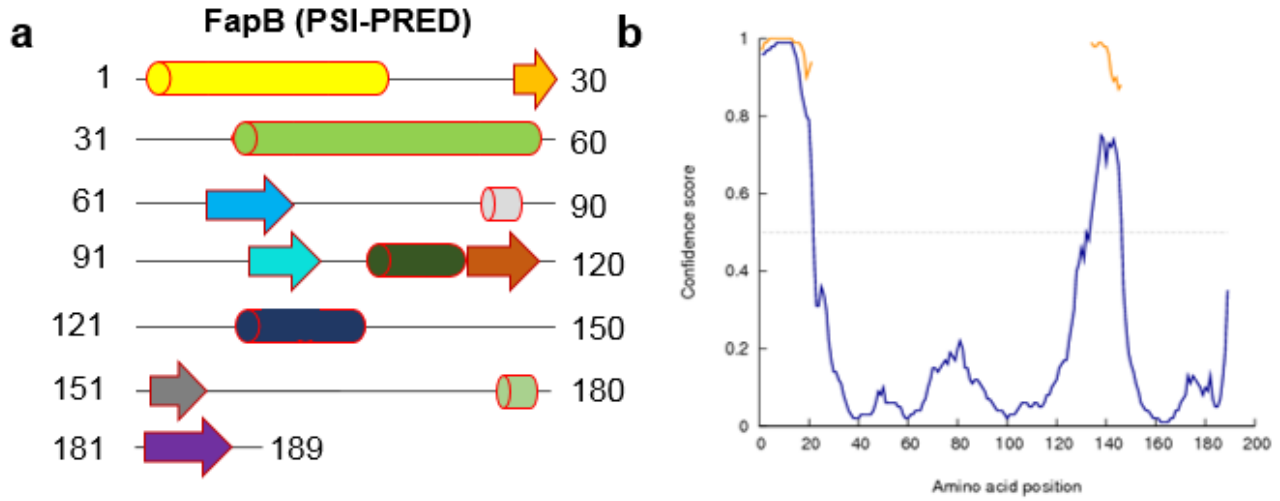


Figure 3.2: Summary of Bioinformatics results for FapB. a, Diagram illustrating PSIPRED **[126]** predicted secondary structure of FapB with helices indicated with a cylinder, β-strands as arrows and disordered regions as lines. **b**, Chart illustrating likelihood that the residues are in a disordered region of the protein generated using DISOPRED **[160]**. The likelihood that the residue is disordered is shown in blue, probability of a disordered region being involved in protein binding based on evolutionary sequence conservation and amino acid composition is shown in orange.

3.1.3 FapC

FapC is predicted to be a very unstructured protein with ~81 % of the protein predicted to be disordered by Phyre2 (secondary structure predictions are ~65 % accurate for proteins with limited homologs [**155**]), with no significant homologs, no major structural elements and a signal sequence from residues 1-29 (**Figure 3.3**). Three amyloid repeats could be identified by manual inspection based on the characteristic motif X_{15}-G-X_4-N-X_3-G-X_6-N-X_7 (67-103, 134-170, 291-327) (**Figure 3.3C**). A simple construct was designed with the signal sequence removed and replaced with an NTH-tag for expression to the cytoplasm and to ease purification

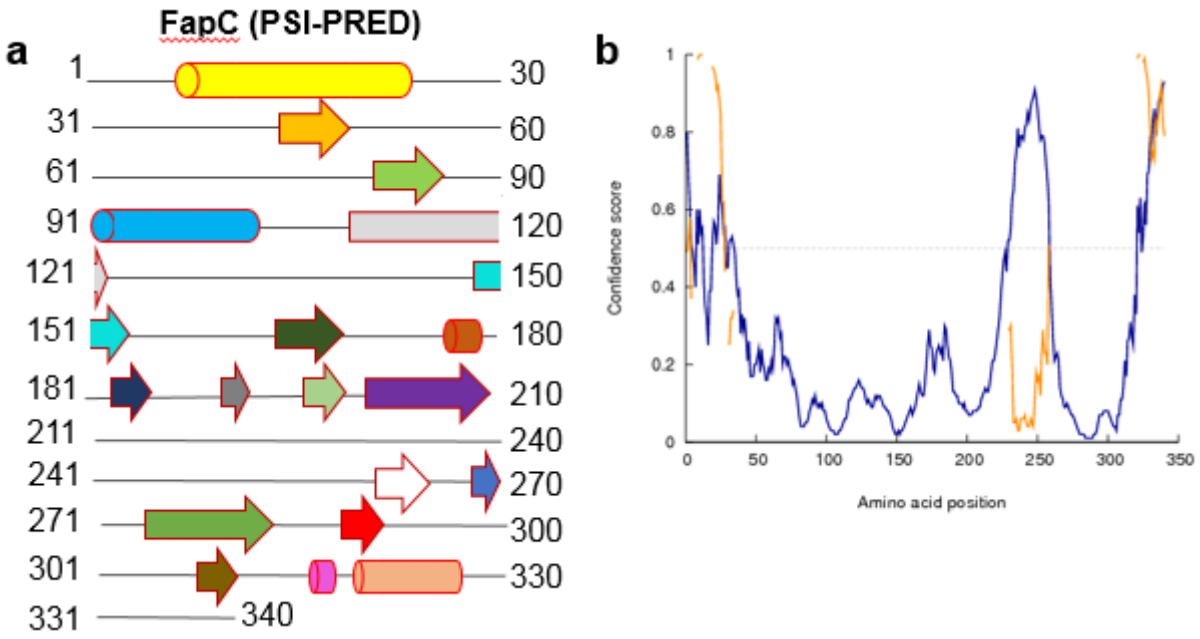
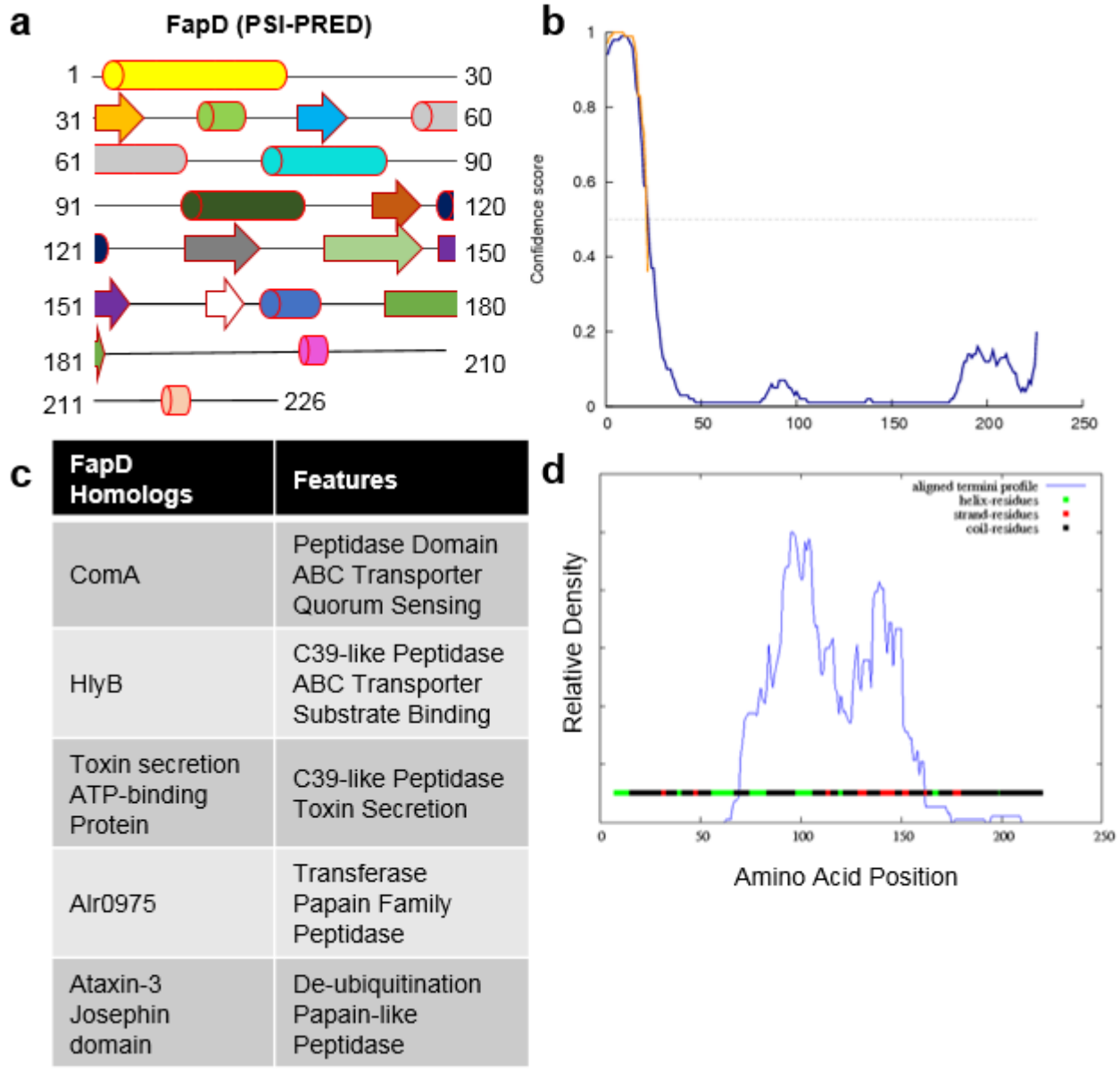


Figure 3.3: Summary of Bioinformatics results for FapC. a, Diagram illustrating PSIPRED **[126]** predicted secondary structure of FapC with helices indicated with a cylinder, β-strands as arrows and disordered regions as lines. **b**, Chart illustrating likelihood that the residues are in a disordered region of the protein generated using DISOPRED **[160]**. The likelihood that the residue is disordered is shown in blue, probability of a disordered region being involved in protein binding based on evolutionary sequence conservation and amino acid composition is shown in orange.

3.1.4 FapD

FapD is predicted to be a member of the C39-like peptidase family with homology to many members including the peptidase subdomain of the Streptococcus protein ComA (**Figure 3.4C**). Structural predictions indicate that FapD is a structured protein with a compact structured core flanked by a disordered N-terminus and a shorter but possibly also disordered C-terminus (**Figure 3.4**).



FapD Homologs	Features
ComA	Peptidase Domain ABC Transporter Quorum Sensing
HlyB	C39-like Peptidase ABC Transporter Substrate Binding
Toxin secretion ATP-binding Protein	C39-like Peptidase Toxin Secretion
Alr0975	Transferase Papain Family Peptidase
Ataxin-3 Josephin domain	De-ubiquitination Papain-like Peptidase

Figure 3.4: Bioinformatics results for FapD. a, Diagram illustrating PSIPRED **[126]** predicted secondary structure of FapD with helices indicated with a cylinder, β-strands as arrows and disordered regions as lines. **b**, Chart illustrating likelihood that the residues are in a disordered region of the protein generated using DISOPRED **[160]**. The likelihood that the residue is disordered is shown in blue, probability of a disordered region being involved in protein binding based on evolutionary sequence conservation and amino acid composition is shown in orange. **c**, Table of Selected FapD Homologs with features of interest based on Phyre2 results **[155]** and manual inspection **d**, Chart of Dompred **[161]** results for FapD where the aligned sequences are used to produce a plot of the density of aligned termini (Blue), this indicates that there is a conserved core domain to the protein from residues 55 to 160

Both Phyre2 and I-TASSER produced predicted structures with Phyre2 producing a model with 72 % of residues modelled at >90 % accuracy (**Figure 3.5A**), and I-TASSER producing a model with a C-score of -1.09 (**Figure 3.5B**). The core domain of FapD predicted by Phyre2 aligns well with the core domain of the I-TASSER model (**Figure 3.5C**), suggesting that this prediction is reliable. Both prediction servers suggest that the N-terminus of FapD is unstructured while the C-terminus appears to contain some secondary structure but may be extended or partially disordered. The suggested extended C-terminal helix in the I-TASSER model is interesting as it could conceivably be capable of forming extensive contacts with interaction partners with other components of the Alf system. FapD is predicted to be a cysteine peptidase from the C39-like peptidase family, these proteins contain an active site tetrad of proximal Glutamine, Cysteine, Histidine and Glutamate/Aspartate, with Cysteine and Histidine comprising the catalytic residues, these can be observed to be conserved in FapD (Q50, C56, H140, D156) and are positioned close together in the predicted structure (**Figure 3.5D**). Both I-TASSER and Phyre2 predict that C56 and H140 are the active residues in the protein. The protein is predicted by SignalP [**158**] to contain a signal peptide (Residues 1-19). The secondary structure prediction (**Figure 3.4A**) supports the suggestion that FapD has disordered regions at the N-terminus and C-terminus although the boundaries of the secondary structure suggest a larger central domain perhaps residues 30-185.

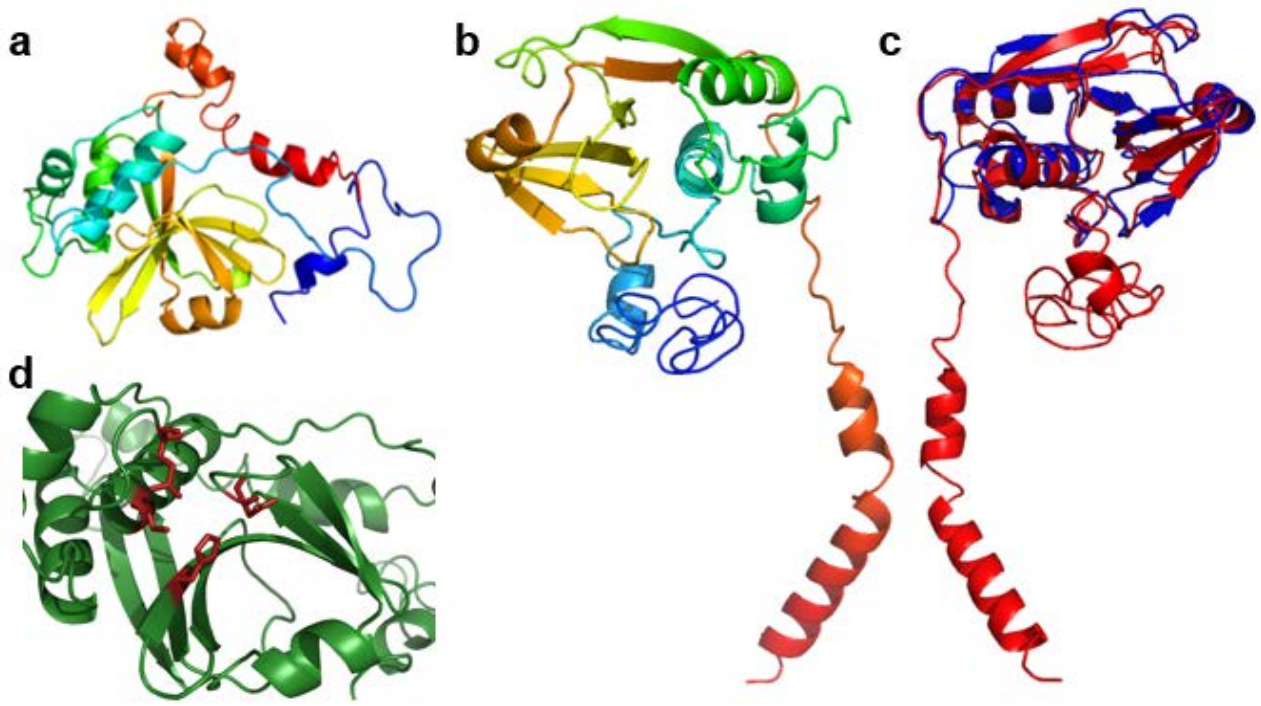


Figure 3.5: Further Bioinformatics Predictions for FapD. a, Cartoon model structure of FapD generated by Phyre2 **[155]** coloured blue to red from N-terminus to C-terminus. **b**, Cartoon model structure of FapD generated by I-TASSER [**156**] coloured blue to red from N-terminus to C-terminus. **c**, Overlay of FapD model of conserved core domain based on PDB structure **4RY2** [**162**] (Blue) and I-TASSER model structure (Red). **d**, Cartoon diagram of FapD structure predicted by Phyre2 **[155]** illustrating the predicted catalytic site residues in dark red sticks.

For initial work a simple construct was designed with the signal sequence removed and replaced with an NTH-tag for expression to the cytoplasm and to ease purification; additional constructs were designed with truncations to the N-terminus $FapD^{23-226}$ and $FapD^{46-226}$. The C-terminus of FapD in these constructs was not truncated because although PSIPRED suggests the region to be largely random coil (**Figure 3.5A**), DISOPRED does not confidently predict the region to be disordered (**Figure 3.4B**). The I-TASSER model also suggests a long helix or pair of helices suggesting some structure to the C-terminus (**Figure 3.5B**).

3.1.5 FapE

FapE is predicted to be an unstructured protein with ~64 % of the protein predicted to be disordered by Phyre2 (secondary structure predictions are ~65 % accurate for proteins with limited homologs [**155**]), with no significant homologs, limited secondary structure elements and a signal sequence from residues 1-28 (**Figure 3.6**). A construct was designed with the signal sequence removed and replaced with an NTH-tag for expression to the cytoplasm and to ease purification.

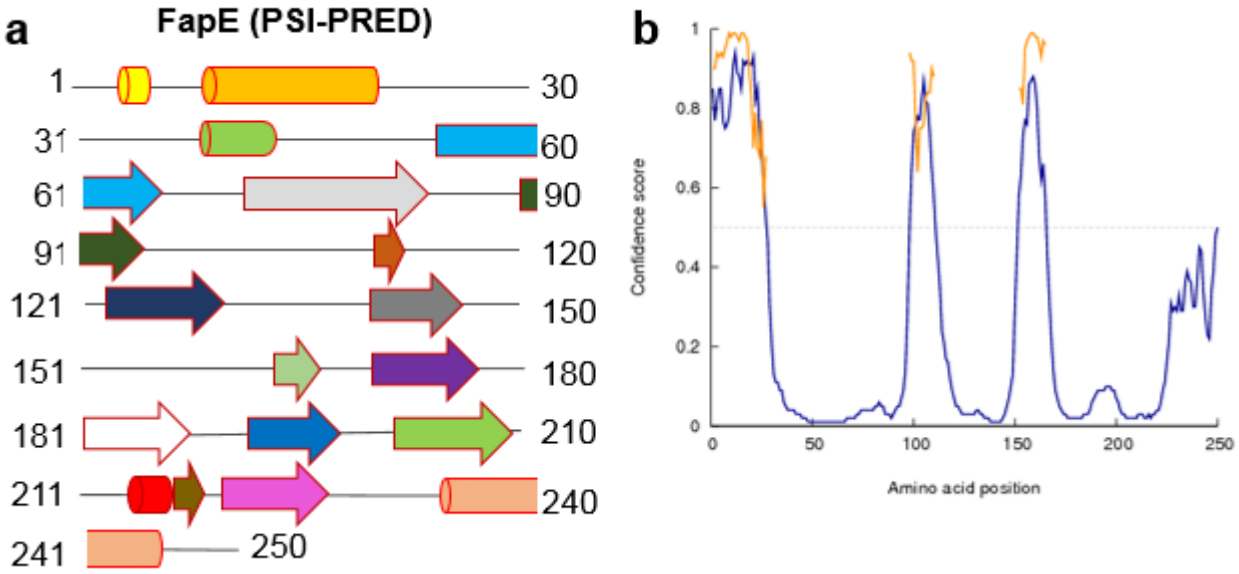
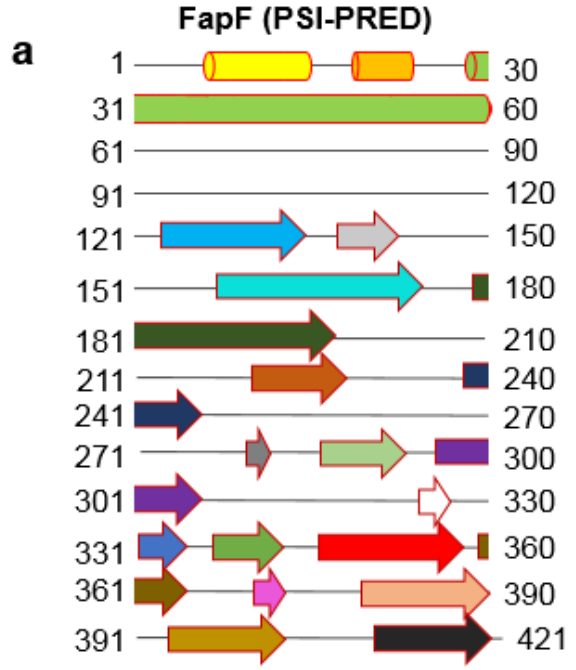
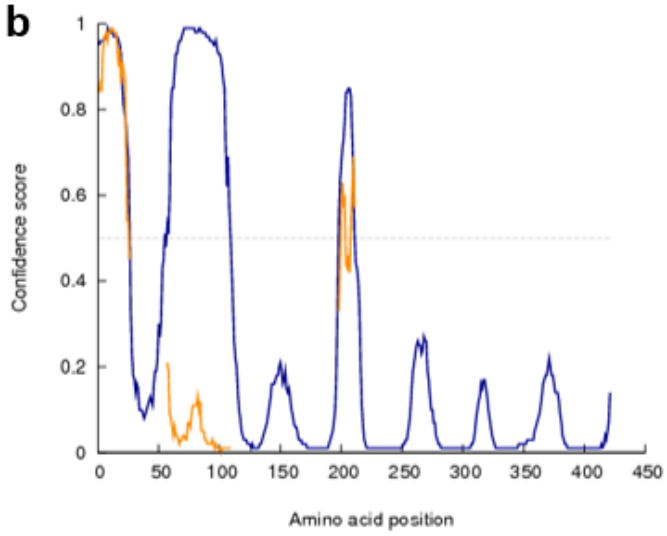


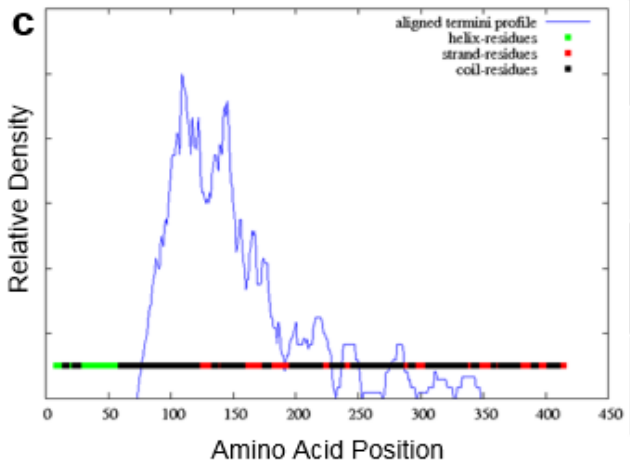
Figure 3.6: Summary of Bioinformatics results for FapE. a, Diagram illustrating PSIPRED **[126]** predicted secondary structure of FapE with helices indicated with a cylinder, β-strands as arrows and disordered regions as lines. **b**, Chart illustrating likelihood that the residues are in a disordered region of the protein generated using DISOPRED **[160]**. The likelihood that the residue is disordered is shown in blue, probability of a disordered region being involved in protein binding based on evolutionary sequence conservation and amino acid composition is shown in orange.

3.1.6 FapF

FapF is predicted to be a β-barrel membrane protein, with homology to many members of the FadL family of B-barrels involved in fatty acid transport (FadL, TodX, TbuX) (**Figure 3.7D**). Structural predictions indicate that FapF is a structure protein with a large C-terminal domain, likely to be the β-barrel, and an extended, largely disordered N-terminal region but which contains a long helix near the N-terminus after the predicted signal sequence (SignalP) is removed (**Figure 3.7**).







d

FapF Homologs	Features
FadL (*P. aeruginosa*)	β-barrel fatty acid transporter
TodX (*P. putida*)	β-barrel Toluene transporter
TbuX (*Ralstonia pickettii*)	β-barrel Toluene transporter
OmpW (*E. coli*)	β-barrel

Figure 3.7: Bioinformatics results for FapF. a, Diagram illustrating PSIPRED **[126]** predicted secondary structure of FapF with helices indicated with a cylinder, B-strands as arrows and disordered regions as lines. **b**, Chart illustrating likelihood that the residues are in a disordered region of the protein generated using DISOPRED **[160]**. The likelihood that the residue is disordered is shown in blue, probability of a disordered region being involved in protein binding based on evolutionary sequence conservation and amino acid composition is shown in orange. **c**, Chart of Dompred [**161**] results for FapF where the aligned sequences are used to produce a plot of the density of aligned termini (Blue), this indicates that there is a conserved core domain to the protein from residues 80 to 350 **d**, Table of Selected FapF Homologs with features of interest based on Phyre2 results **[155]** and manual inspection

Both Phyre2 and I-TASSER produced predicted structures with Phyre2 producing a model with 82 % of residues modelled at >90 % accuracy (**Figure 3.8A**), and I-TASSER producing a model with a C-score of -1.36 (**Figure 3.8D**). Alignment of the two model structures (**Figure 3.8E**) shows that the predictions for the β-barrel domain are consistent while the loops and N-terminal domain are not, consistent with the lack of conservation, structure and homology in these regions.

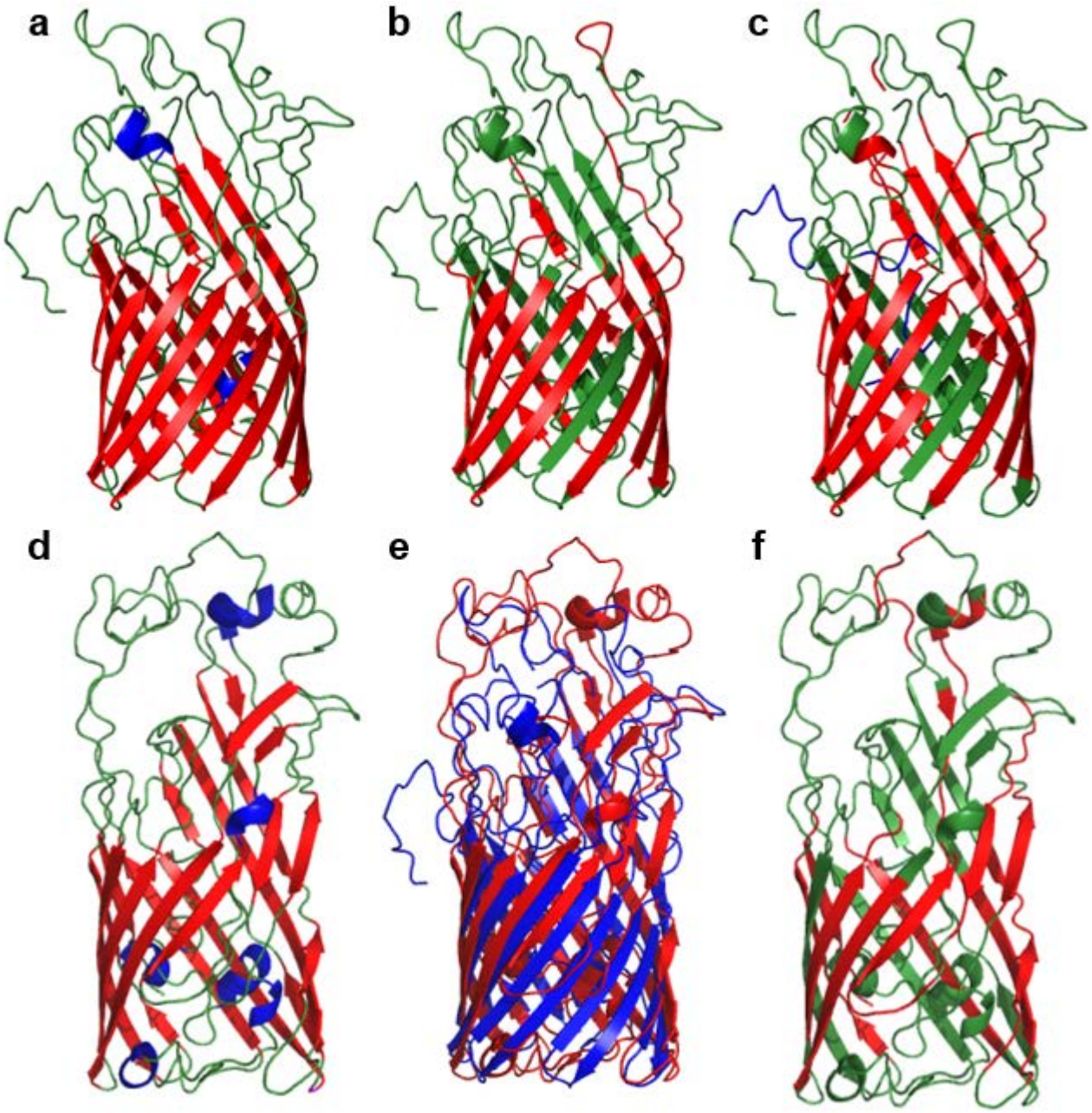


Figure 3.8: Further Bioinformatics results for FapF. **a,** Cartoon model structure of FapF generated by Phyre2 **[155]** coloured by secondary structure with loops in green, strands in red and helices in blue.**b,** Cartoon model structure of FapF generated by Phyre2 with transmembrane strands predicted by PRED-TMBB **[159]** coloured in red. **c,** Cartoon model structure of FapF generated by Phyre2 coloured by according the PSIPRED **[126]** secondary structure prediction with loops in green, strands in red and helices in blue. **d,** Cartoon model structure of FapF generated by I-TASSER **[156]** coloured by secondary structure with loops in green, strands in red and helices in blue. **e,** Overlay of FapF model generageted by Phyre2 (Blue) and I-TASSER model structure (Red) **f,** Cartoon model structure of FapF generated by I-TASSER with transmembrane strands predicted by PRED-TMBB **[159]** coloured in red

Comparison of the structural predictions for the β-barrel from different sources indicates that predictions of the transmembrane strands and the core domain of FapF are inconsistent (**Figure 3.8**). The homology based structure produce by Phyre2 [**155**] (**Figure 3.8A**) contains several strands which are predicted to be disordered or loops regions by PRED-TMBB (**Figure 3.8B**) and PSIPRED (**Figure 3.8C**); this inconsistency is also present in the I-TASSER predicted structure (**Figure 3.8F**).Conversely PRED-TMBB also predicts some sequences to be transmembrane strands which are not modelled as such by Phyre2 or I-TASSER. In general the structure predictions for FapF suggested a β-barrel with between 12 and 14 strands and containing a large disordered N-terminal domain. The domain prediction for FapF does indicate that there is a core domain starting somewhere around residue 80 but the C-terminal boundary is unclear, probably because the barrel domain is consistently predicted to extend to the end of the protein.

A construct was designed with the signal sequence removed and replaced with an NTH-tag, without the signal sequence FapF should not be transported into the periplasm and so we would not expect it to be integrated into the outer membrane. The resulting protein expression will be expected to be in inclusion bodies but it should be possible to refold the protein as B-barrels have proven amenable to refolding protocols [**163**]. A full length construct was also designed retaining the native N-terminal signal sequence which together with the use of a CTH tag may allow expression to the membrane.

3.1.7 Summary of Bioinformatics Results

Bioinformatics predictions are heavily reliant on homology to proteins of known structure or function, as a result proteins with no strong genetic relationships such as FapA, FapB, FapC and FapE predictions are limited to more basic analysis such as secondary structure, signal peptides and disorder predictions. These proteins were predicted to be highly disordered proteins with limited structural features, they were all predicted to be transported into the periplasm via the SEC machinery due to the presence of the characteristic signal peptide. Construct design of these proteins was aimed at simply producing the full length protein with the signal sequence removed and a hexahistidine-tag replacing it to allow purification from the cytoplasm. No functions could be reliably predicted for any of these proteins, however it is known that FapC and FapB are amyloid proteins [**92**] and the repeats could be identified by manual inspection, and additional constructs were therefore designed with particular repeats deleted.

For proteins with clear relationships to proteins of known structure and function more meaningful information can be obtained from bioinformatics, this was the case for FapD and FapF. FapD was predicted to be a structured protein from the C39-like peptidase family as reported by Dueholm et al., [**92**], these proteins are found in both proteolytically active and inactive forms, FapD is predicted to be active based on the presence of the catalytic tetrad. Domain, disorder and structural predictions all suggest that FapD has disordered regions in the C-terminus and N-terminus. Caution must be exercised when using the bioinformatics to design truncations without experimental evidence so a conservative, full length, construct was initially designed with the signal peptide replaced with a hexahistidine tag, later some additional constructs were produced with truncations to the disordered regions and with a point mutations targeted at the active cysteine residue. FapF is predicted to be a β-barrel protein as reported by Dueholm et al., [**92**]. Predictions suggest the possibility of an N-terminal disordered region which may be a plug domain, the exact pattern of β-strands in the β-barrel are not consistently predicted by the four software packages (I-TASSER, PSIPRED, TMBB-PRED and Phyre2), although the overall size and boundaries of the β-barrel domain are reasonably consistent. The signal peptide was removed and replaced with a hexahistidine tag, the consequence of which would be insoluble expression requiring refolding as the protein would be unable to integrate with the outer membrane, this decision was taken in the knowledge that β-barrels have been successfully refolded at high yields [**163**]. If refolding is unsuccessful a CTH construct retaining the signal sequence could be used for extraction of FapF from the membrane. The disorder predictions were incorporated with experimental data later during construct optimisation to produce truncated constructs.

3.2 Expression and Characterisation of *Pseudomonas* Alf

3.2.1 Expression Trials

Initially pNIC-NTH constructs of the individual Alf components with their signal sequences removed and replaced with NTH-tags were designed, cloned, transformed into BL21 and tested for soluble and insoluble expression over a range of temperatures. pNIC-NTH-FapB$^{19\text{-}189}$ showed both soluble and insoluble expression of protein at a range of temperatures, the highest levels of soluble expression were seen with 3 hour incubations at 30 or 37˚C, while the highest levels of insoluble (and overall) expression were seen at 30 or 37˚C overnight (**Figure 3.9**). Interestingly the FapB in the soluble fraction showed different migration properties to that in the insoluble, migrating more slowly on the gel, indicating either a difference in size or some difference in SDS-binding or denaturation (**Figure 3.9**). The other proteins only expressed in the insoluble fractions (**Figure 3.10, 3.11**) except for perhaps a small amount of soluble expression of FapC (**Figure 3.10**).

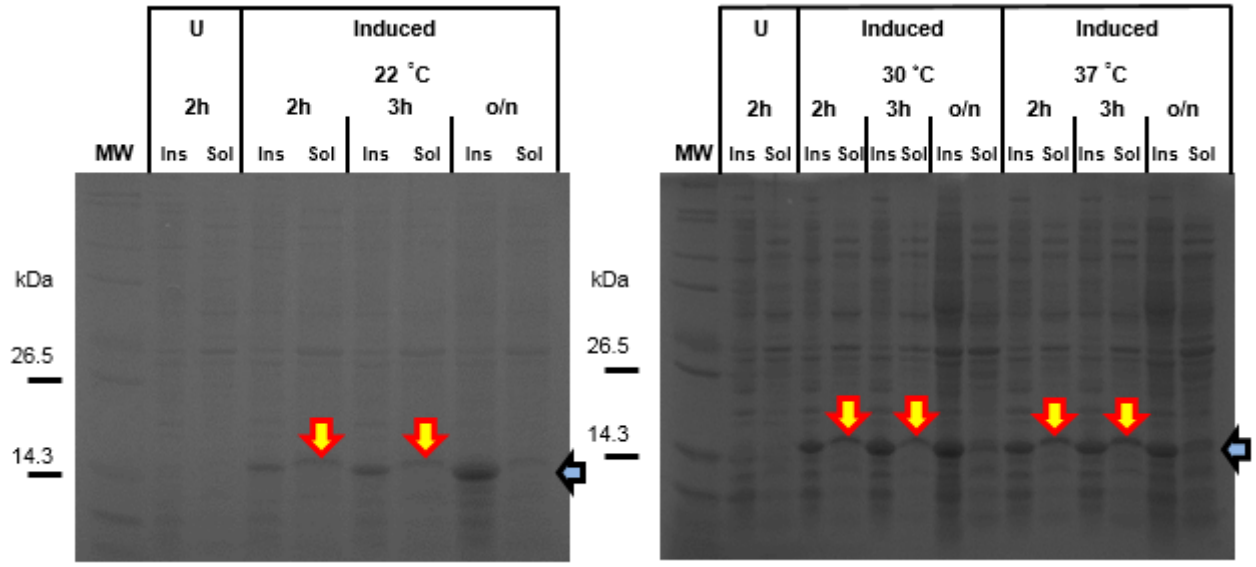


Figure 3.9 FapB Expression Trials. SDS-PAGE gels of FapB expression following induction at 22, 30 and 37 degrees after 2 hour (2h), 3 hour (3h) and overnight (o/n) incubation compared to uninduced (U) sample incubated at 30 degrees for 2 hours. The samples were fractionated into soluble (Sol) and insoluble (Ins) fractions. Relevant bands in the molecular weight ladder (MW) are indicated with their weight and the FapB bands are indicated with a blue arrow for the size of the insoluble bands (clearly visible) and with individual yellow arrows indicating the soluble FapB bands which are seen at a slightly different MW.

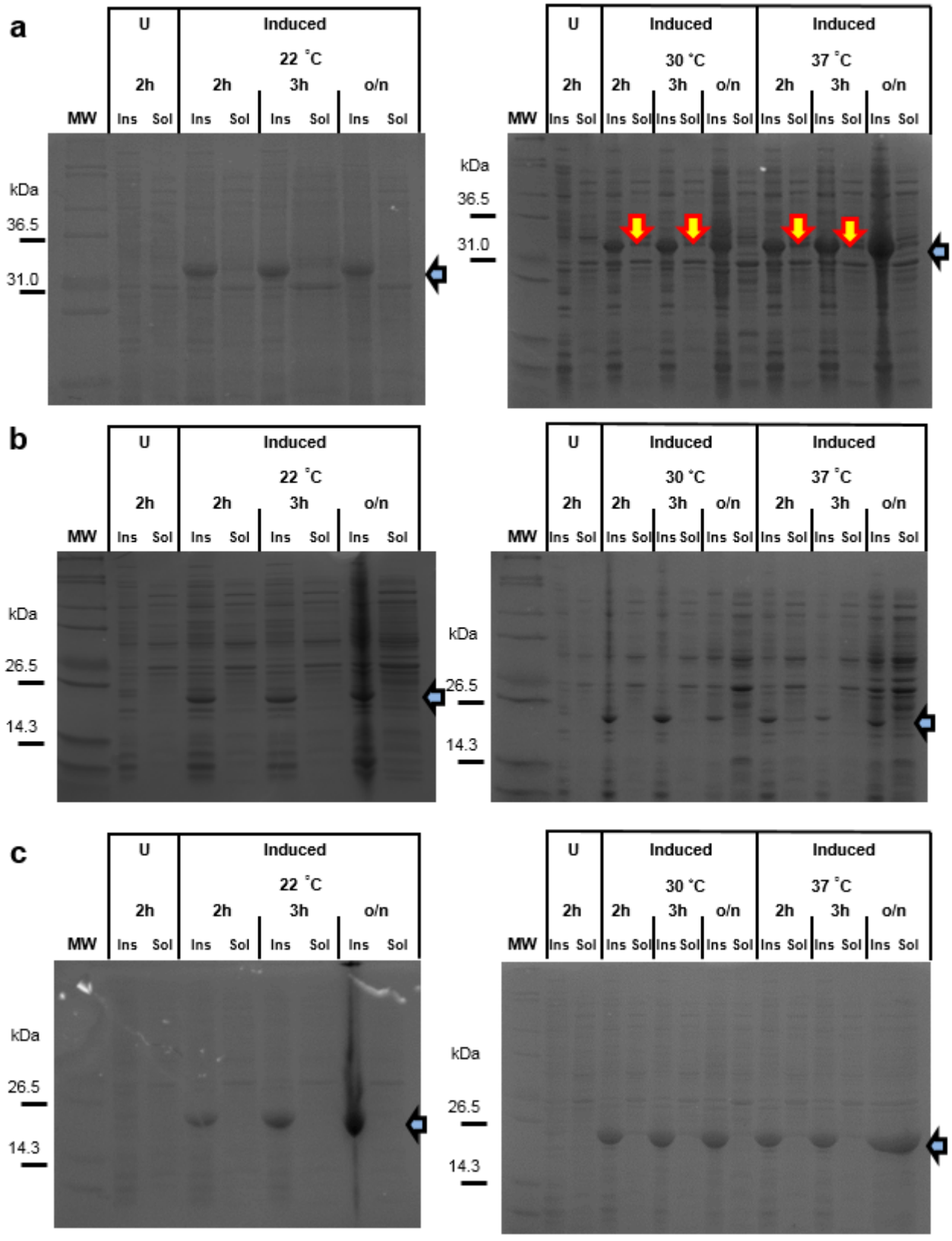


Figure 3.10 Expression Trials of FapC, FapD and FapE. a, FapC, **b**, FapD and **c**, FapE, showing protein expression over a range of indicated temperatures and incubation times. The samples were fractionated into soluble (Sol) and insoluble (Ins) fractions. Notably all the proteins expressed well insolubly as indicated with the blue arrow. FapC also showed some limited soluble expression, indicated with the yellow arrows. Notably, the final two lanes of the right FapE gel have merged due to a well defect. Molecular weight of relevant markers are indicated for each gel.

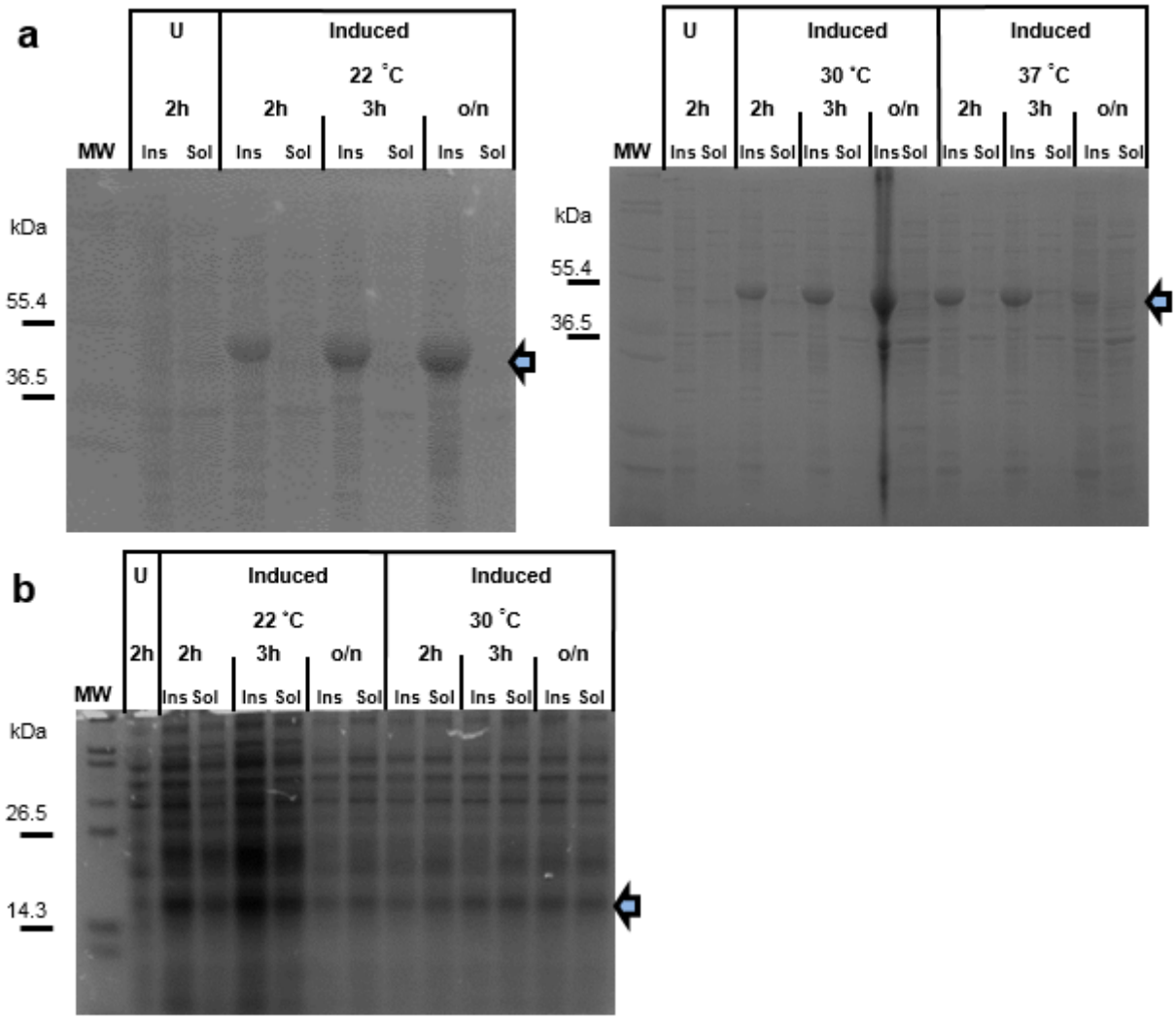


Figure 3.11 Expression Trials of FapF and FapA. Showing protein expression over a range of indicated temperatures and incubation times. The samples were fractionated into soluble (Sol) and insoluble (Ins) fractions. **a**, FapF expression trial, FapF shows strong expression in the insoluble fraction at all temperatures as indicated with the blue arrows. **b**, FapA expression trial, the FapA expression trial is less conclusive, since there appears to also be some expression in the uninduced sample (bands of correct molecular weight indicated with blue arrow). FapA expression was not tested at 37°C and the uninduced sample was whole cell lysate rather than fractionated.

3.2.2 Purification of FapB samples

Since some soluble FapB could be expressed the protein was purified using Ni-Immobilised Metal Affinity Chromatography (IMAC) under both native and denaturing conditions, both of which produced similar samples of FapB as visualised on an SDS-PAGE gel (**Figure 3.12**), unlike the FapB bands in the whole cell lysate (**Figure 3.9**). 1D NMR of the natively purified FapB indicated the protein was largely unstructured in solution (**Figure 3.13**).

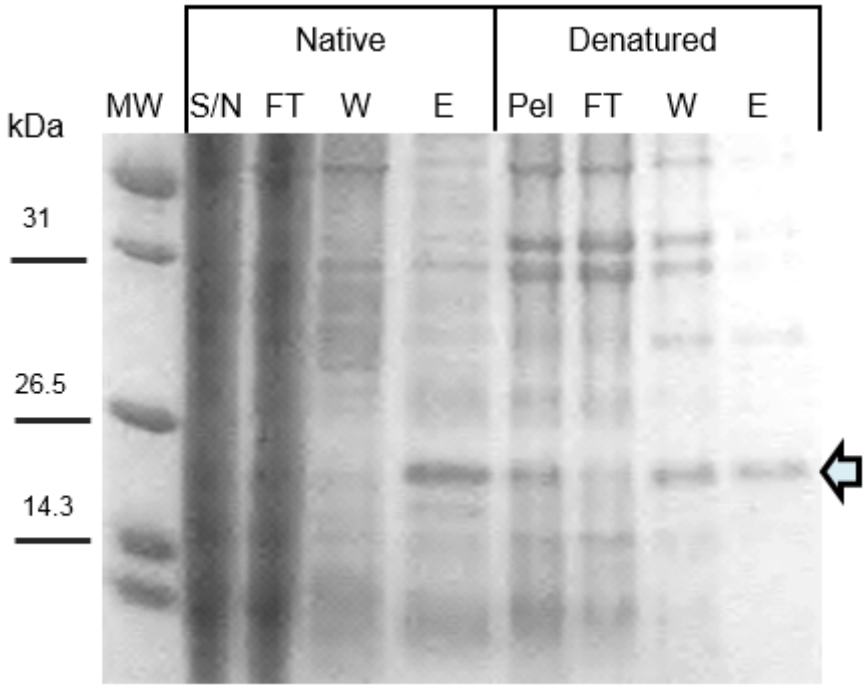


Figure 3.12 Purification of FapB under Native and Denaturing Conditions. SDS-PAGE gel of FapB purified under both Native and Denaturing (8 M Urea) conditions. Using the same culture of cells the protein was fractionated following lysis into Supernatant (S/N) and Pellet (Pel) fractions by centrifugation, these were used in native and denaturing purifications respectively and purified by Ni-IMAC. The fractions shown are the initial flow-through (FT), the wash fraction (W) and the final elution (E). Both elution fractions are enriched for FapB but the denaturing elution fraction did have greater purity than the native. The supernatant and flow through of the native purification had much greater concentrations of protein causing the overloading seen on the gel in these lanes.

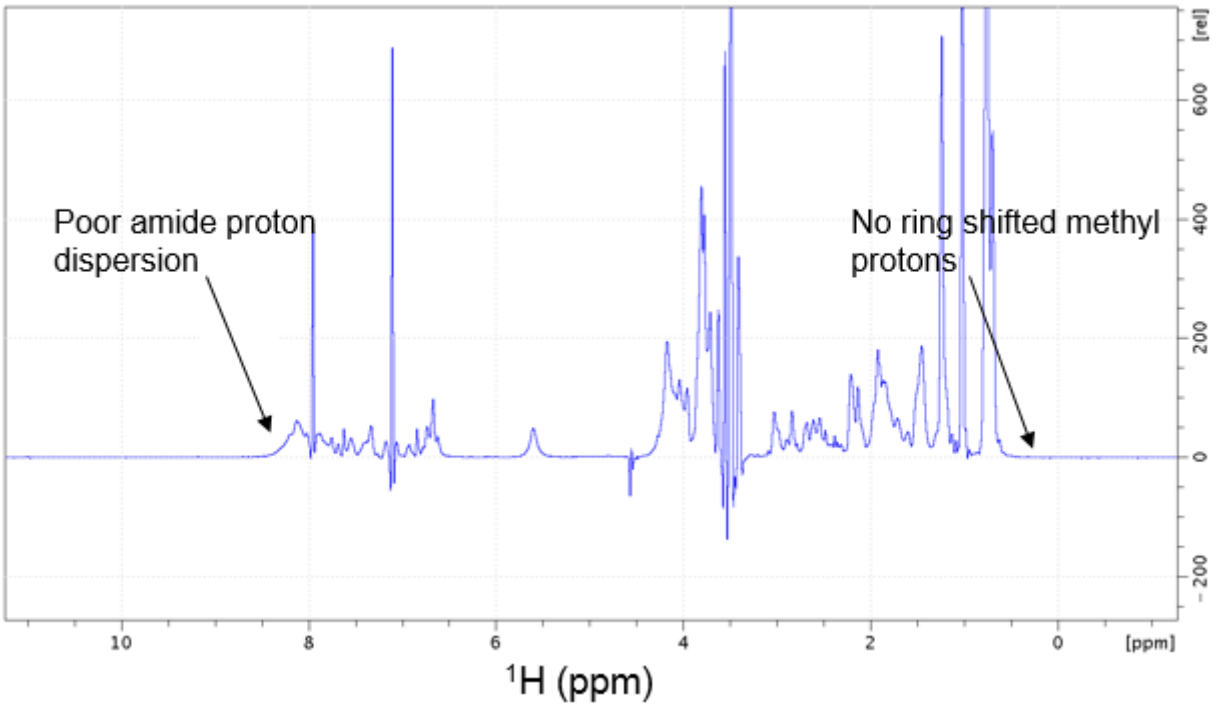


Figure 3.13 1D ^{1}H NMR of FapB purified under native conditions. The 1D spectra of FapB (300 mM NaCl, 20 mM TrisHCl pH 8), indicates that the protein is unfolded due to the poor dispersion of the amide proton peaks and the absence of ring shifted methyl proton peaks. The relevant regions have been labelled and indicated with arrows.

3.2.3 Denaturing Purification of Alf Components

Since all the proteins expressed largely in the insoluble fraction they were purified under denaturing conditions (8M urea) by Ni-IMAC, this produced protein samples were which were relatively pure and contained proteins of the expected molecular weight (**Figure 3.14**). These samples could be used for various experiments and attempts were made to refold the proteins (**3.2.4**).

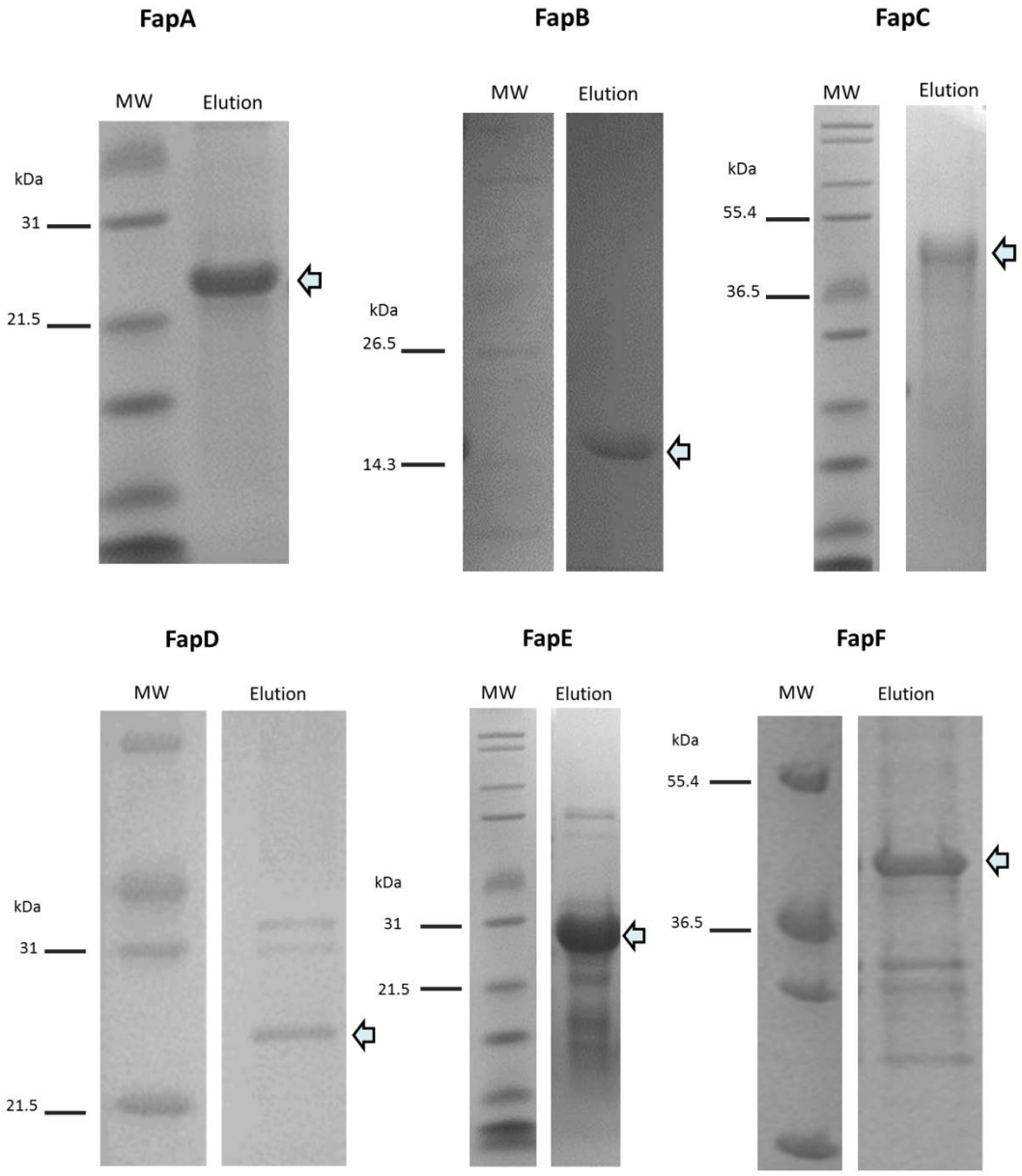


Figure 3.14: Samples of Alf components purified under denaturing conditions. Gels of elution fractions of Alf components with relevant molecular weight markers labelled with their size and with the protein of interest indicated with an arrow. Samples are reasonably pure with the protein of interest constituting the majority of the sample except for FapD where yields were lower.

3.2.4 Refolding PAO1 NTH Constructs

Various refolding methods were attempted with the purified Alf proteins, including pulse refolding, dialysis and matrix assisted refolding. Evidence of limited refolding could be observed for FapA, FapB and FapE as soluble protein could be produced by matrix assisted refolding (**Figure 3.15**), but was not as effective by dialysis or pulse refolding (data not shown). FapD PAO1 could not be successfully refolded, possibly due to degradation of the tag (**Figure 3.15**). These proteins were generally unstable however, precipitating or aggregating within hours or days of purification CD and NMR conducted on these samples shortly after purification indicated that although the proteins may have some degree of structure they were largely unfolded **Figure 3.16, Figure, 3.17, Figure 3.18**). FapF could be refolded by pulse refolding (**3.4.1**).

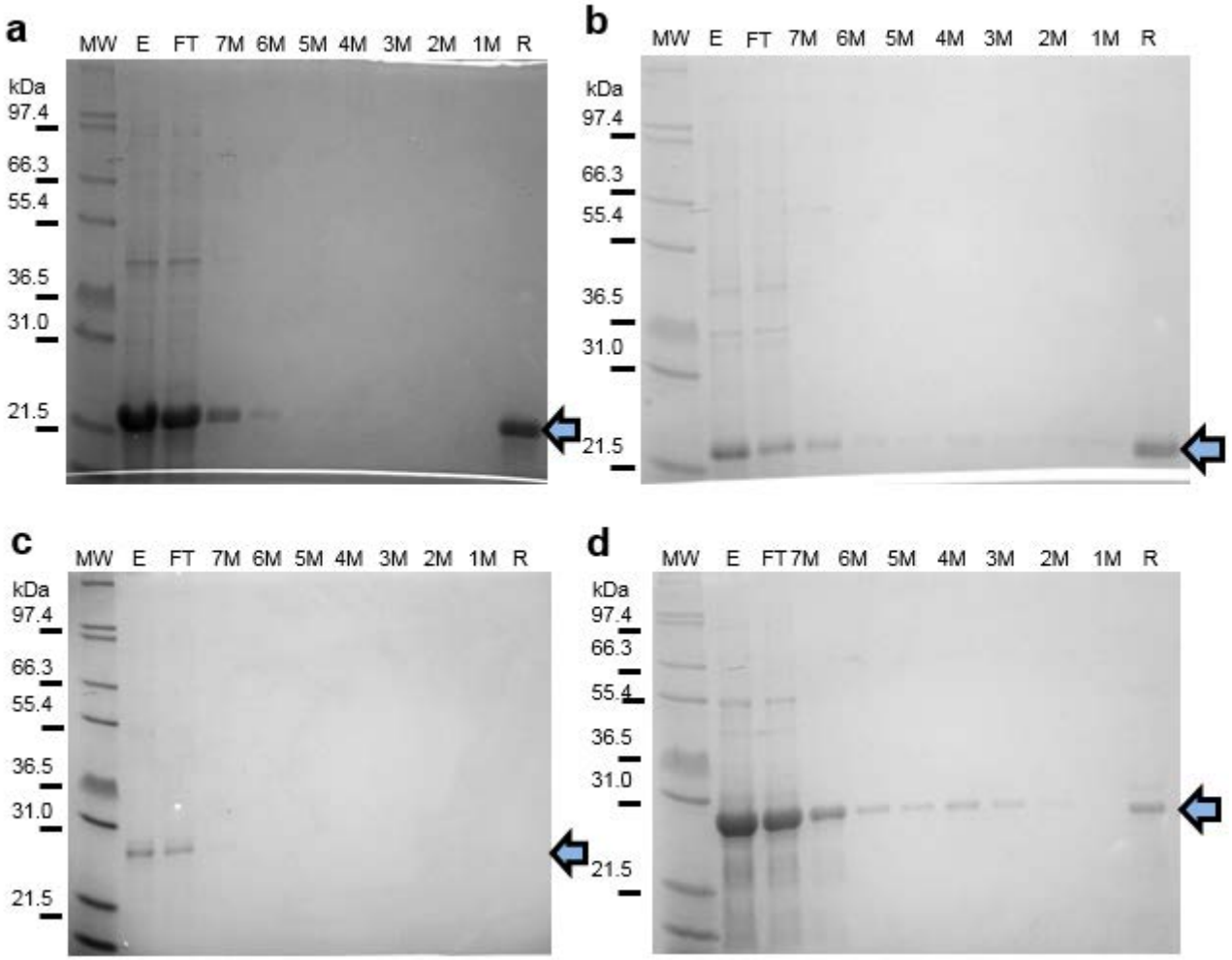


Figure 3.15: Matrix Assisted Refolding of Alf Components. a, FapA, **b**, FapB, **c**, FapD, **e**, FapE. SDS-PAGE gels showing the fractions collected during matrix assisted refolding of several Alf component proteins, E is the elution from a denaturing purification of the protein of interest after dialysis to remove the imidazole, FT is the flow through of the Ni column after application of the previous elution, 7M-1M are wash fractions of 7-1 M urea respectively while R is the final eluted fraction without any urea. The relevant MW of the protein of interest is indicated with a blue arrow.

3.2.5 CD and NMR of Alf Components

3.2.5.1 FapA

Refolded FapA was transferred into CD buffer by serial dilution and a CD spectra was collected. The accurate data collection was limited to a range from 200-250 nm (**Figure 3.16A**), secondary structure was estimated from the CD data using the BeStSel webserver [**174**]. The secondary structure based on the CD data was largely unstructured (~59 %) with some β-strand and α-helical structure (**Figure 3.16B**). Comparable results were obtained with the K2D webserver [**214**] which predicted the protein to be ~46 % disordered although it only used the data range from 200-240 nm. These results support the suggestion that FapA is an intrinsically unstructured protein. 1D ^{1}H NMR was also used to study the structure of FapA and indicated that the protein was unfolded with poor peak dispersion and no ring-shifted methyl peaks (**Figure 3.16C**).

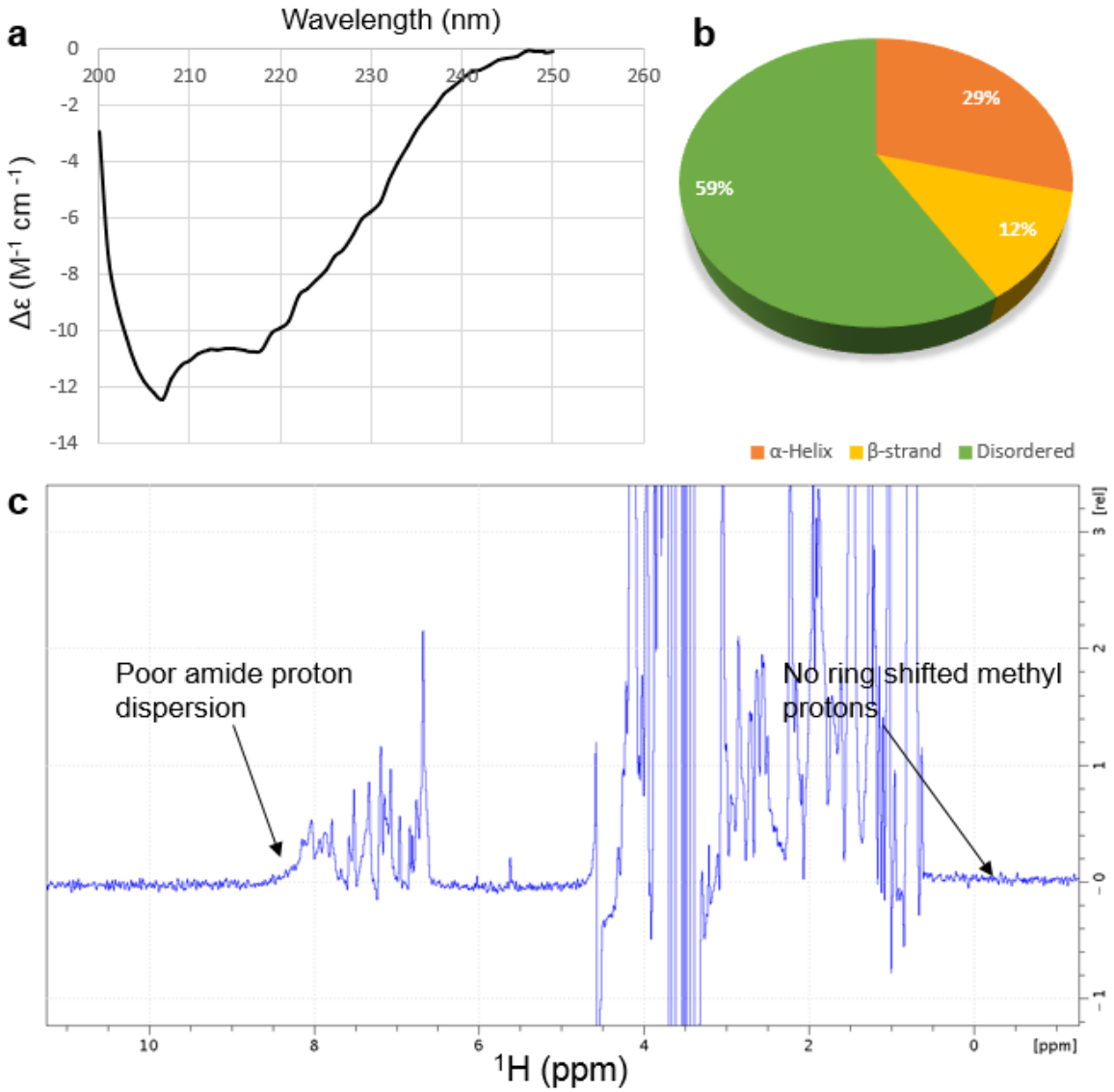


Figure 3.16: Structural studies of FapA. a, CD Spectra of FapA, 200 nm to 250 nm. **b**, Pie Chart showing the percentage of basic secondary structure elements (β-strand,α-helical and disordered) as estimated from the CD Spectra using BeStSel [**174**]. **c**, 1D ^{1}H NMR spectra of FapA (300 mM NaCl, 20 mM TrisHCl pH 8), poor amide proton peak dispersion and the absence of ring shifted methyl protein peaks suggests that the protein is unfolded.

3.2.5.2 FapB

Refolded FapB was transferred into CD buffer by serial dilution and a CD spectra was collected. The accurate data collection was limited to a range from 190-250 nm (**Figure 3.17A**) and this was used to predict the secondary structure using the BeStSel webserver [**174**]. The secondary structure based on the CD data was largely unstructured (~58 %) with some β-strand (22 %) and α-helical structure (20 %) (**Figure 3.17B**). Comparable results were obtained with the K2D webserver [**214**] which predicted the protein to be ~70 % disordered although it only used the data range from 190-240 nm. These results support the suggestion that FapB is an intrinsically unstructured protein in solution. 1D ^{1}H NMR was also used to study the structure of FapB and indicated that the protein was unfolded with poor amide proton peak dispersion and no ring-shifted methyl proton peaks (**Figure 3.17C**).

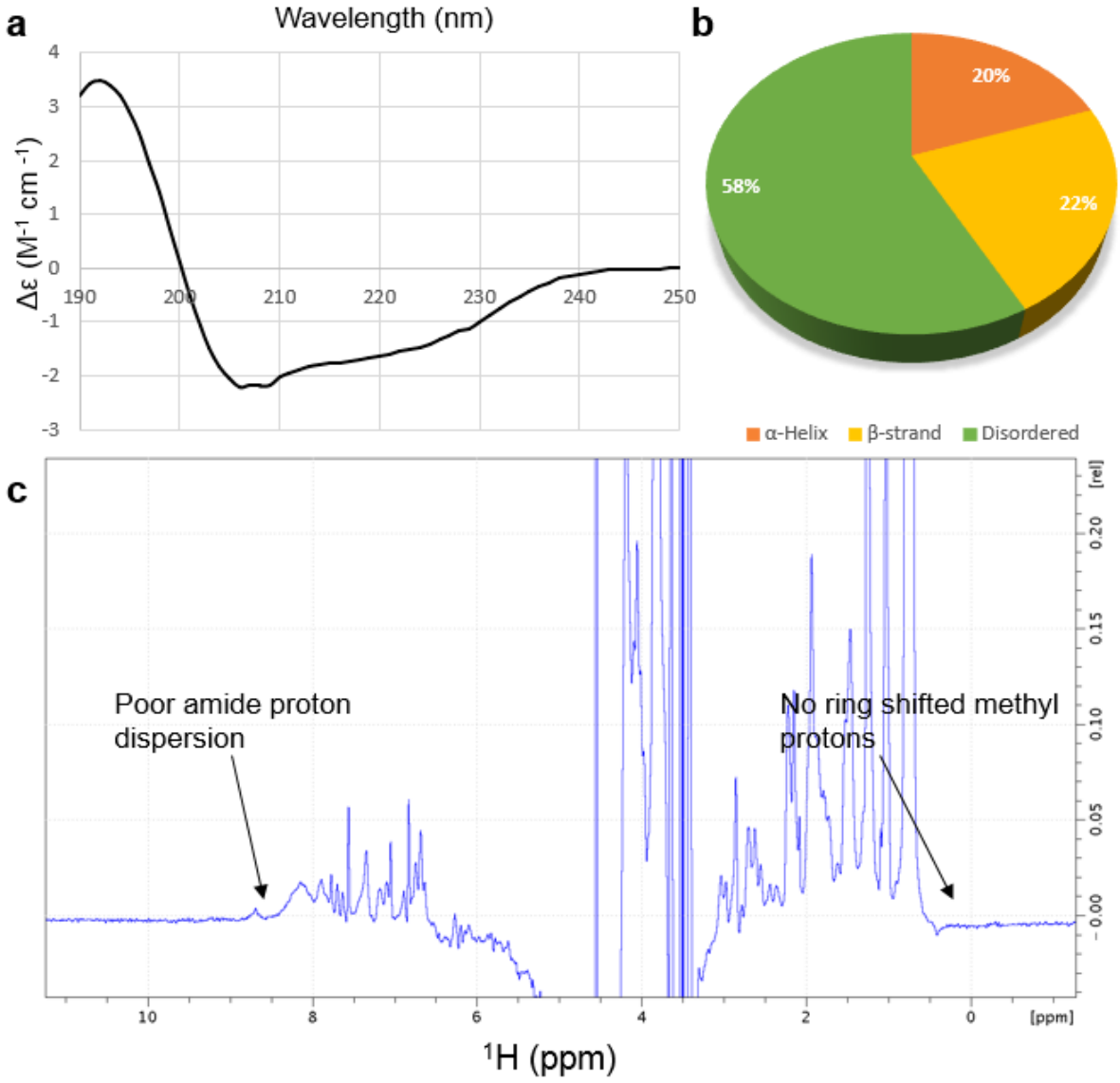


Figure 3.17: Structural studies of FapB. a, CD Spectra of FapB, 190 nm to 250 nm. **b**, Pie Chart showing the percentage of basic secondary structure elements (β-strand,α-helical and disordered) as estimated from the CD Spectra using BeStSel [**174**]. **c**, 1D ^{1}H NMR spectra of FapB, poor amide protein peak dispersion and the absence of ring shifted methyl proton peaks suggests that the protein is unfolded.

3.2.5.3 FapE

Refolded FapE was transferred into CD buffer by serial dilution and a CD spectra was collected. The concentration of FapE was lower than FapB and FapA resulting in a much lower signal for the CD experiment, accurate data collection was still limited to a range from 190-250 nm (**Figure 3.18A**) and this was used to predict the secondary structure using the BeStSel webserver [**174**]. The signal in CD was very low and so the results are less reliable however the secondary structure from the available data indicated the protein was largely unstructured (~66 %) with some β-strand (34 %) and α-helical structure (0.5 %) (**Figure 3.18B**). Comparable results were obtained with the K2D webserver [**214**] which predicted the protein to be ~50 % disordered although it only used the data range from 190-240 nm. These results support the suggestion that FapE is an intrinsically unstructured protein. 1D ^{1}H NMR was also used, however no protein peaks were observed structured or unstructured (**Data Not Shown**), either because the concentration was too low or the protein as aggregated for the signal to be detected.

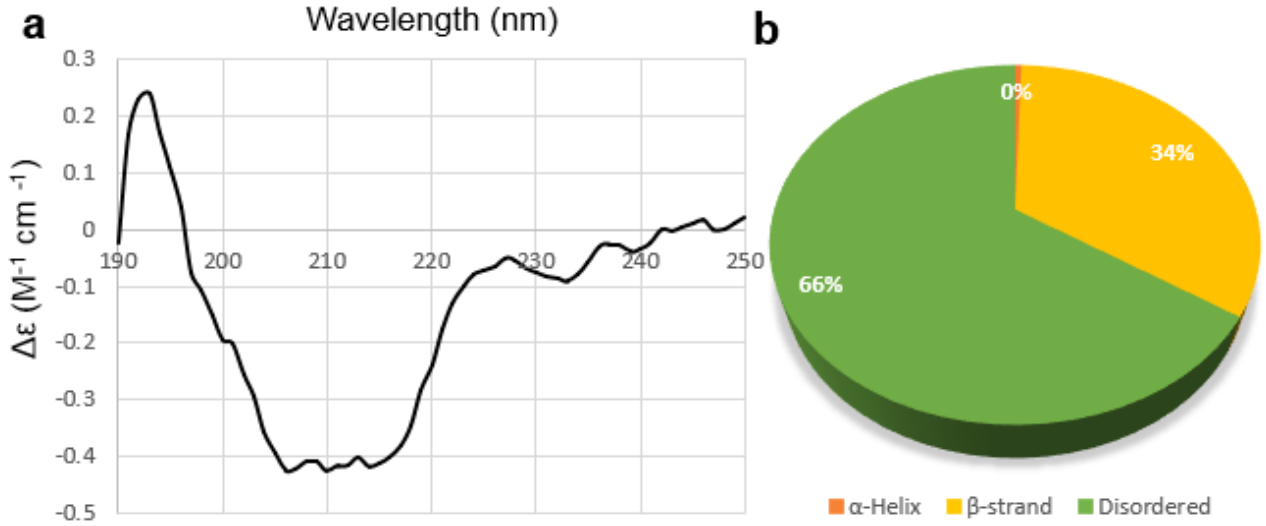


Figure 3.18: Structural studies of FapE. a, CD Spectra of FapE, 190 nm to 250 nm. **b**, Pie Chart showing the percentage of basic secondary structure elements (β-strand,α-helical and disordered) as estimated from the CD Spectra using BeStSel [**174**].

3.2.5.4 FapF

Refolded FapF was transferred into CD buffer, containing 0.05 % (v/v) LDAO by serial dilution and a CD spectra was collected, accurate data collection was still limited to a range from 190-250 nm (**Figure 3.19A**) and this was used to predict the secondary structure using the BeStSel webserver [**174**]. The secondary structure based on the CD data was largely structured with β-strand (40 %), α-helical structure (22 %) and disordered regions (38%) (**Figure 3.19B**), this supports the suggestion that FapF is a structured protein which may contain a largely disordered domain. Interestingly the K2D2 [**214**] prediction for FapF produced significantly different results with a prediction that the protein was ~84 % α-helical with only 0.6 % of the protein predicted to form β-strands. 1D ^{1}H NMR was also used to study the structure of FapF, although the relaxation properties of a large protein in a micelle caused line broadening the data still indicates that the protein is folded as evidenced by ring-shifted methyl peaks (**Figure 3.19C**).

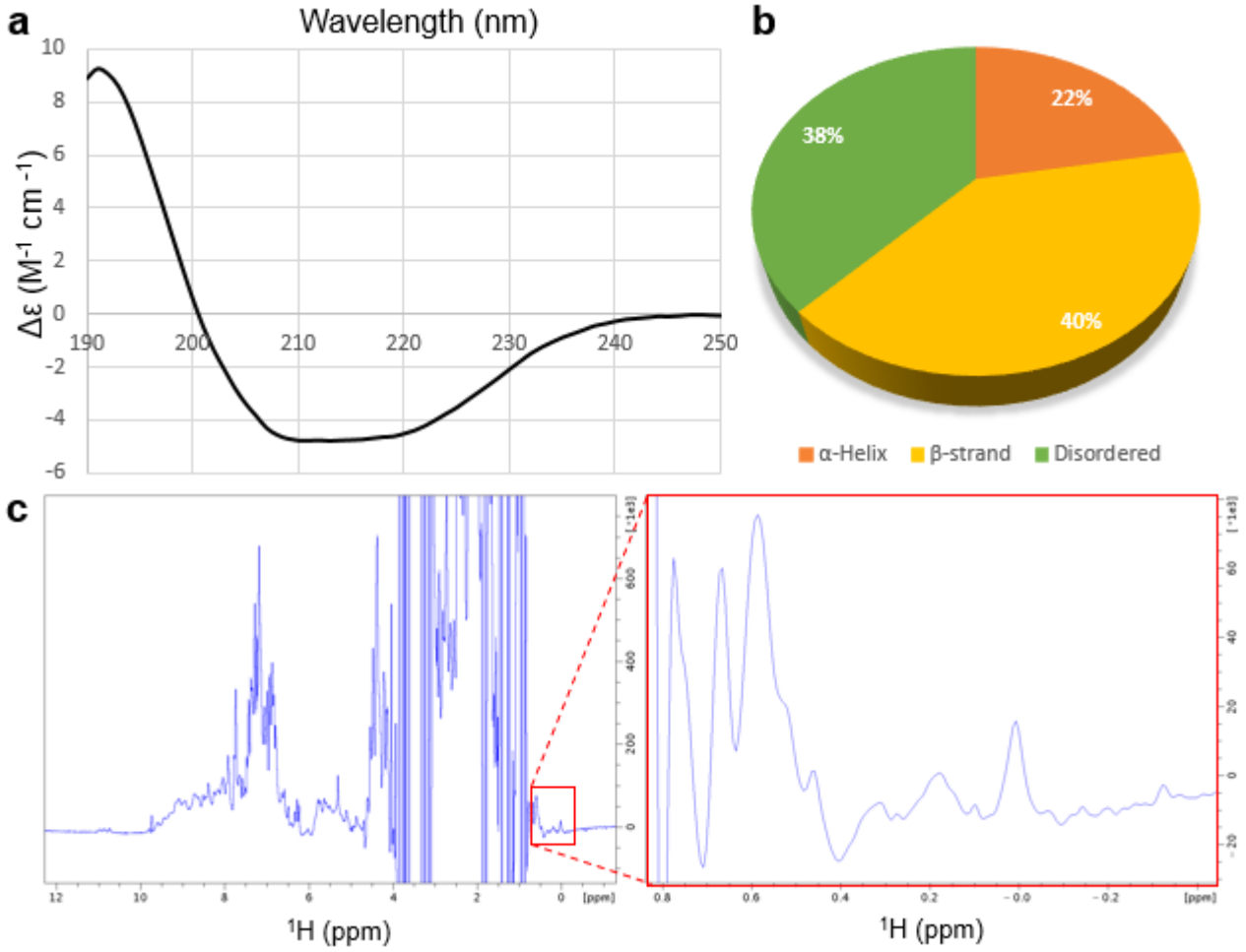


Figure 3.19: Structural studies of FapF. a, CD Spectra of FapF, 190 nm to 250 nm. **b**, Pie Chart showing the percentage of basic secondary structure elements (β-strand,α-helical and disordered) as estimated from the CD Spectra using BeStSel [**174**]. **c**, 1D ^{1}H NMR spectra of FapF (300 mM NaCl, 20 mM TrisHCl pH 8, 0.1 % (*w/v*) LDAO), the protein appears to be at least partially folded as evidenced by the presence of ring shifted methyl protons which have been highlighted in the inset indicated with the red box.

3.2.6 ThT Assay

3.2.6.1 ThT Assays by Dilution

To test whether FapB and FapC could form amyloid in vitro, FapB and FapC were purified under denaturing conditions and the ThT assay was used to measure the creation of amyloid fibres. The protein solutions were diluted directly into the ThT buffer and the fibrillation measure. An exponential curve of increasing fluorescence was observed for FapC, FapB and a sample containing both proteins, while the Chymotrypsin control showed no rise in fluorescence (**Figure 3.20**), this indicates that FapC and FapB are both capable of forming amyloid. Notably no lag phase is observed which indicates that nucleation of the fibres has already occurred, this means it is not possible to conclude whether FapB is a nucleator for FapC fibrillation or not. Since FapB is capable of forming amyloid alone but is only a minor fibre component in vivo [**138**] these results do support the hypothesis that FapB shares a similar role to CsgB as a fibre nucleator. The absence of a lag phase may indicate that the samples may contain oligomers before dilution leading to rapid nucleation.

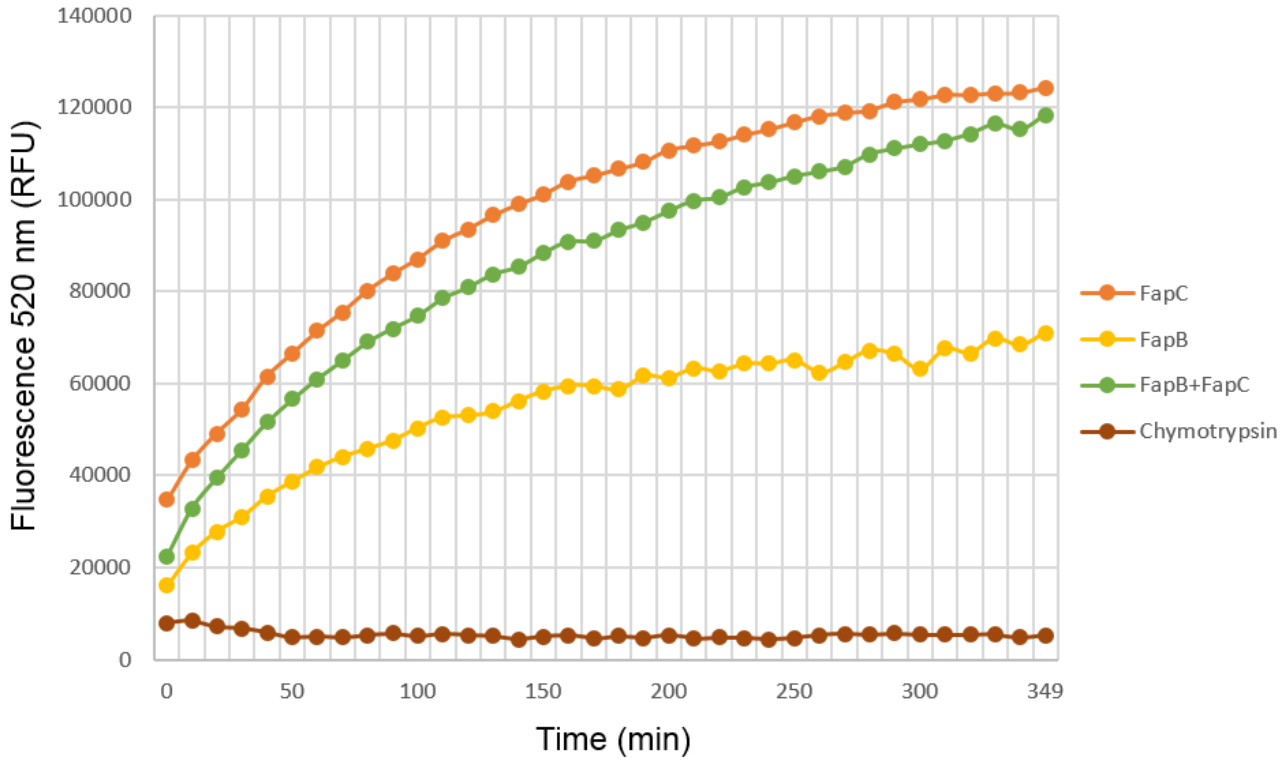


Figure 3.20: ThT Assay for FapB and FapC by Dilution from Urea. Graph showing the relative fluorescence of FapB, FapC and FapB + FapC together, along with the chymotrypsin control. The fluorescence of the buffer alone has been subtracted from all four curves. For this experiment excitation was at 485 nm and emission was measured at 520 nm due to the available filters. Protein samples were used at high concentrations ~5 mg/mL.

3.2.6.2 Fibrillation of FapC UK4

Purified FapC (UK4) was generously provided by Dr Maria Andreasen in 8 M Guanadinium HCl, the sample was transferred into phosphate buffer using a desalting column 75 µM ThT added and then fibrillation measured using the ThT assay. FapC (UK4) shows a sigmoidal curve of ThT fluorescence with a lag, exponential and stationary phase similar in nature to the curves seen for CsgA (**Figure 3.2A**). Similarly a time course experiment was conducted using NMR, a series of 1D ^{1}H NMR spectra were collected of FapC in solution over time, (**Figure 3.21B)**, these showed a decline in signal with no change in peak position presumably reflecting amyloid fibrillation progressing rapidly from monomer to aggregate with no detectable change to an intermediate structure.

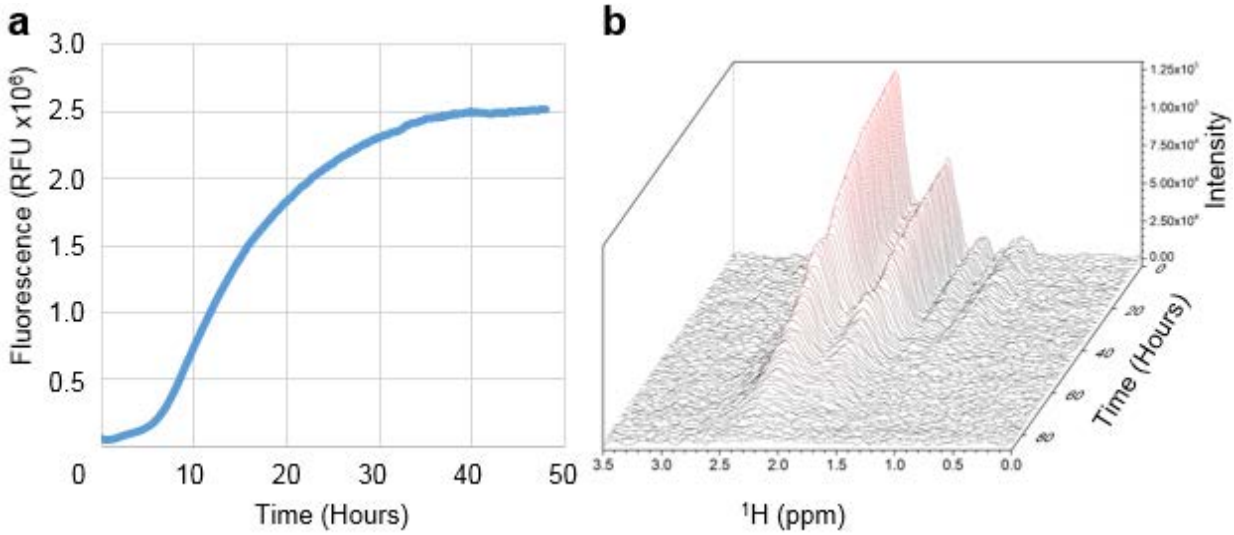


Figure 3.21: Fibrillation of FapC (UK4). a, Graph showing the ThT fluorescence of FapC (UK4) over time showing that FapC forms amyloid. **b**, ^{1}H NMR timecourse showing the decline in peak intensity of selected peaks from FapC over time, presumably reflecting fibrillation.

3.2.7 Interaction Studies

3.2.7.1 Pulldown experiments with pNIC-GST and pNIC-NTH constructs.

To probe the possible interactions between the components and potentially form complexes which might be more amenable to structural studies attempts were made to refold the pNIC-NTH constructs of proteins from the *fap*ABCDEF operon together with pNIC-GST constructs, however most showed no evidence of interactions (**data not shown**). There was some evidence of FapB and FapE interacting in a pull down experiment where the denatured proteins were refolded by rapid dilution together and then purified by Ni-affinity which appeared to support the suggestion that FapB and FapE interact (**Figure 3.22**). Since FapE-GST still appears in the wash the result is not conclusive.

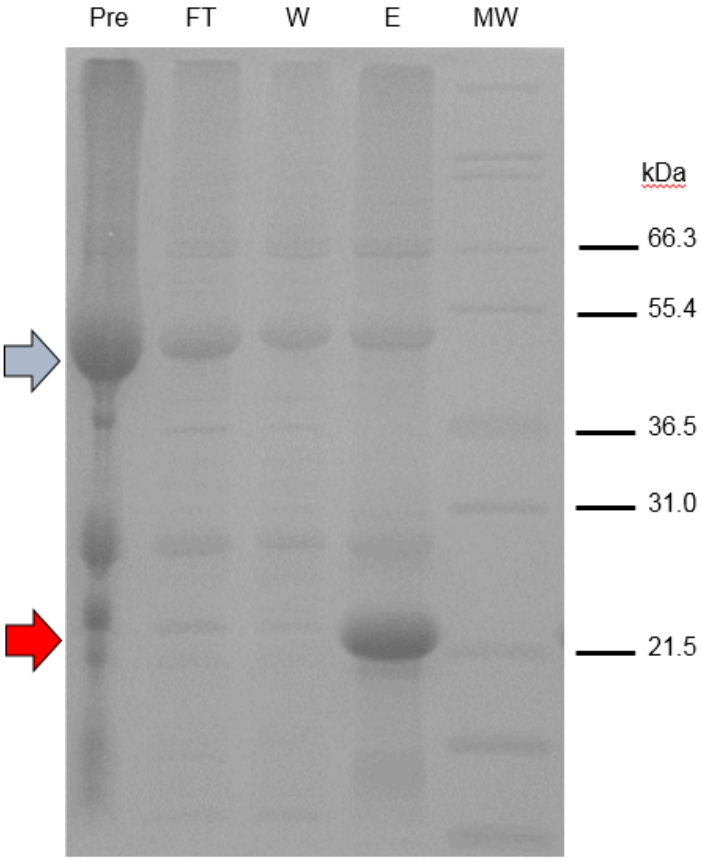


Figure 3.22: SDS-PAGE gel of FapB-NTH:FapE-GST pull down experiment. The fractions shown are the mixture before Ni-purification (Pre), the initial flow-through (FT), the wash fraction (W) and the final elution (E). FapB-NTH band molecular weight is indicated with a red arrow, FapE-GST bands are indicated with a blue arrow. FapB-NTH is purified to a high concentration as expected by the Ni-purification and appears to pull-down some FapE-GST, however the FapE-GST appears to continually be leaching off the column during the wash so the binding to FapB may be weak and/or an artifact of mixing two disordered proteins at such a high concentration.

3.2.7.2 Co-refolding for Co-elution by Gel Filtration

The proteins, (FapA, FapB, FapD and FapE), were purified individually under denaturing conditions as described previously (**3.2.3**). The samples were then combined and refolded by dialysing out the Urea as described in the methods (**2.3.8.3**). The solution was then concentrated to a volume less than 2.5 mL, filtered and applied to a S-200 gel filtration column, the resulting fractions were collected and run on an SDS-PAGE gel for analysis (**Figure 3.23**). FapD appears only in the void suggesting the protein has aggregated. Interestingly there are also some signs of association between FapB and FapE as they appear to co-elute on the column and the elution volume (~73 mL) is significantly higher than would be expected for the individual proteins. The peak (~73 mL) is non-symmetrical suggesting that the proteins are in polydispersed states. FapE is also at a very low concentration. The trailing peak (~88 mL) appears to contain a small quantity of FapA alone, indicating it doesn't interact with the other components.

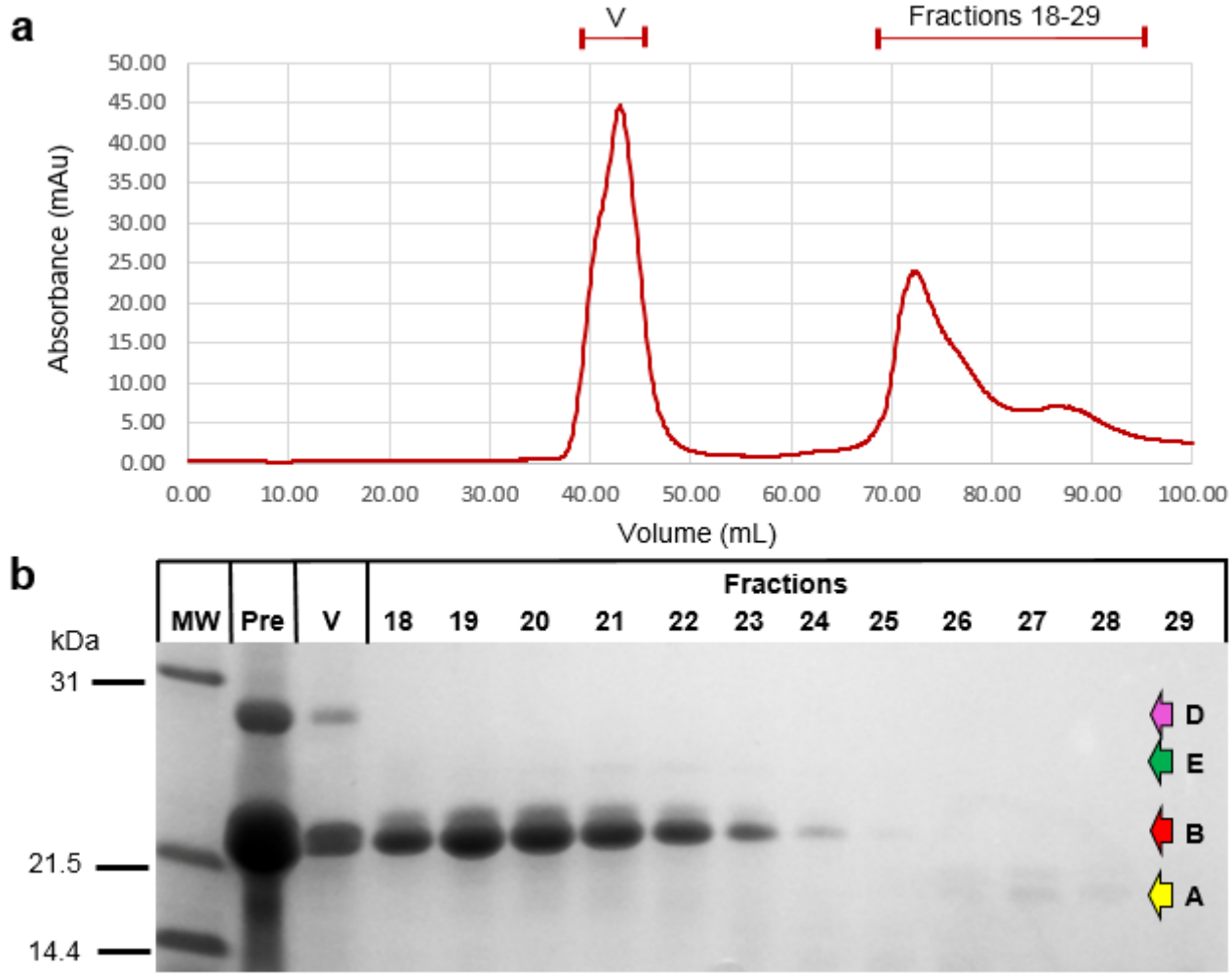


Figure 3.23: Co-refolding and gel filtration of FapA, FapB, FapD and FapE. a, Gel Filtration trace of CsgH purified by gel filtration on an S75-superdex column, fractions were collected and ran on an SDS-PAGE gel. **b**, SDS-PAGE gel showing the dialysed sample before gel filtration (PRE), the protein in the void volume (V) and fractions 18-29 collected from 70-94 mL. Bands corresponding to particular Alf components are indicated with coloured arrows with letters indicating the relevant protein, the bands corresponding to FapA and FapE are particularly faint.

3.2.8 Pull-down experiments with OmpASS constructs

3.2.8.1 Construct Design

An alternative to searching for the interactions entirely *in vitro* was to look for interactions between the components of Alf within *E. coli*. To achieve this a series of new constructs were designed with an OmpA signal sequence for export to the periplasm followed by a histidine tag for purification; these constructs were designed for the *Pseudomonas* UK4 strain, this was to allow for the proteins produced to interact with the proteins produced by the pMMB190Ap:UK4fapA-F full operon plasmid [**92**] provided by Professor Otzen. Co-expression of the tagged proteins with the full operon should allow us to pull-down untagged proteins from the working Alf system, to this end the plasmid used for the periplasmic constructs was selected to have a different origin of replication from the pMMB190 plasmid. The FapF constructs designed in this system were also useful to allow extraction of the β-barrel protein directly from the membrane which will be discussed later (**3.4.5.3**). The necessary plasmid was constructed successfully using pRSF-1B plasmid with an inserted OmpA signal sequence (MKKTAIAIAVALAGFATVAQA) followed by a hexa-histidine tag followed by a site for the protein of interest. Several constructs were designed and made in this plasmid using the primers in **Table 2.4** (**2.2.1.2**) including each of the Alf genes FapA-F from *Pseudomonas* UK4.

3.2.8.2 Pull down results

Pull-down experiments were conducted for FapA-E by co-expression of the individual proteins with the full operon followed by lysis and native purification. Distinguishing individual interacting proteins on the SDS-PAGE gels was complicated by the large number of different proteins present in the elution (**Data Not Shown**). It is possible that this is a result of the large scale expression of the amyloid system producing large amounts of agglutinative protein. Attempts were made to improve the purity of the elution by increasing the amount of imidazole in the wash steps but there was no improvement in purity (**Figure 3.24**). These results could be interpreted as indicating that many components of the Alf system are pulled down during the purification, including amyloid components in an oligomeric state but it is also possible that there is a very high level of background binding.

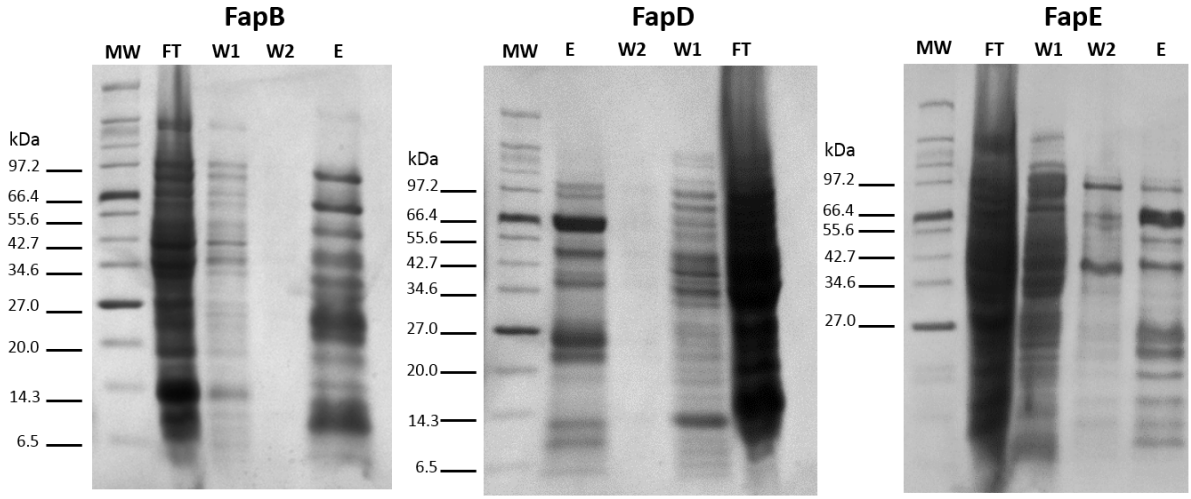


Figure 3.24: Examples of pull-down experiments conducted with the periplasmic constructs and the full operon. SDS-PAGE gels labelled with the initial flow-through (FT), the wash fraction 10 mM imidazole (W1), the second wash fraction 20 mM imidazole (W2) and the final elution (E). Specific bands are enriched during the pull-down but the diverse range of different molecular weights make it difficult to identify the bands and suggesting a high degree of non-specific binding.

3.2.9 Summary

A broad range of different constructs were designed and produced for the Alf components from several *Pseudomonas* strains. Protein expression and refolding was optimised allowing some biophysical experiments to be conducted on all the Alf components. CD and 1D NMR experiments indicated that FapB, FapE and FapA were all unstructured proteins in keeping with the bioinformatics predictions. The function of the Alf components was probed with ThT assays on FapB and FapC samples. Both proteins could form amyloid on their own *in vitro* which is consistent with the idea that FapC and FapB are analogous to CsgA and CsgB. Interactions of the Alf components were probed using several different approaches but although there was some indication of an interaction between FapB and FapE, no convincing proof of interactions was obtained. Nevertheless a broad range of constructs from several homologs have been produced which can be used for future study of these proteins.

Although soluble FapD could not be produced in these initial experiments, the protein was successfully expressed and since it was expected to be a structured protein further optimisation was conducted (**3.3**). FapF was successfully refolded to produce soluble, structured protein which was used for further structural study (**3**.**4**).

3.3 FapD

3.3.1 FapD PAO1: Refolding Experiments & Construct Redesign

Attempts to refold FapD PAO1 by pulse refolding or dialysis both produced large amounts of insoluble precipitation with no soluble protein produced (**data not shown**). Matrix assisted refolding also produced no soluble FapD product (**Figure 3.7C**). Bioinformatics, using TMHMM [**175**], suggested that FapD may contain a transmembrane helices near the N-terminus (**Figure 3.25**). A truncated version of FapD PAO1 with this N-terminal sequence deleted to produce a $FapD^{45\text{-}226}$ construct was tested for soluble expression. The protein remained insoluble over the range of temperatures tested (**Figure 3.26A**) but could be purified under denaturing conditions (**Figure 3.26B**). In an attempt to refold FapD the protein was pulse refolded into various buffers at pH 9, 8 and 5, LDAO was tested as an additive as was the simultaneous addition of FapF (**Figure 3.26C**). Some FapD could be recovered but only in the absence of detergent, these samples precipitated gradually overnight with the amount of precipitation seemingly increasing with pH. The refolding experiment was then attempted with a pH 4 acetate buffer but precipitated immediately. Since the pH 5 sample appeared slightly more stable than the others a large scale purification and refolding of FapD was conducted at this pH producing a pure sample of the truncated FapD construct despite considerable precipitation during and after refolding (**Figure 3.17D**).

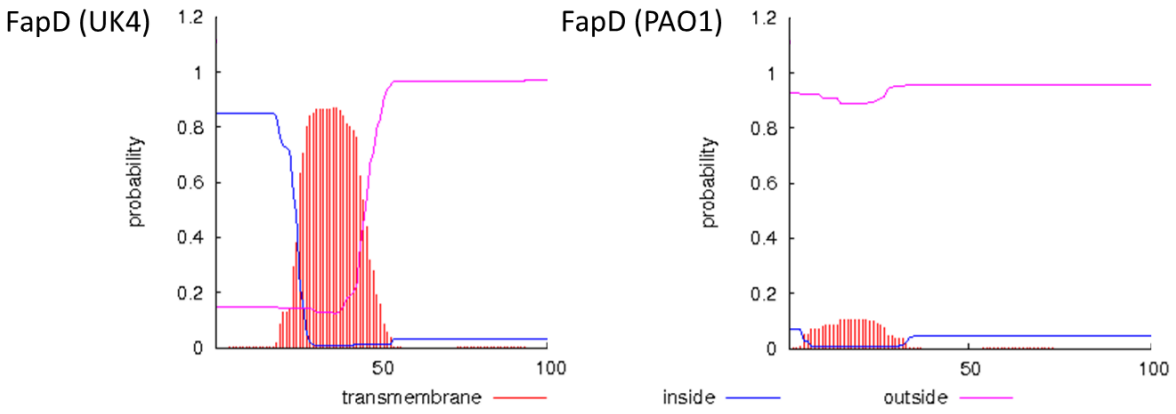


Figure 3.25: Transmembrane Helices Predictions for FapD PAO1 and UK4. Although the TMHMM [**175**] prediction for a transmembrane helices for FapD PAO1 suggests that a transmembrane helix is unlikely the UK4 homolog of the protein is predicted to contain a transmembrane helix. This suggests either that the proteins may contain a transmembrane helix rather than an N-terminal signal sequence or that the annotation of the UK4 construct is not reliable. The vertical red lines indicate the probability of the residue being in a transmembrane helix, the blue line indicates the probability of the residue being on the inner side of the inner membrane and the pink line indicates the probability of the residue being on the outer side of the inner membrane.

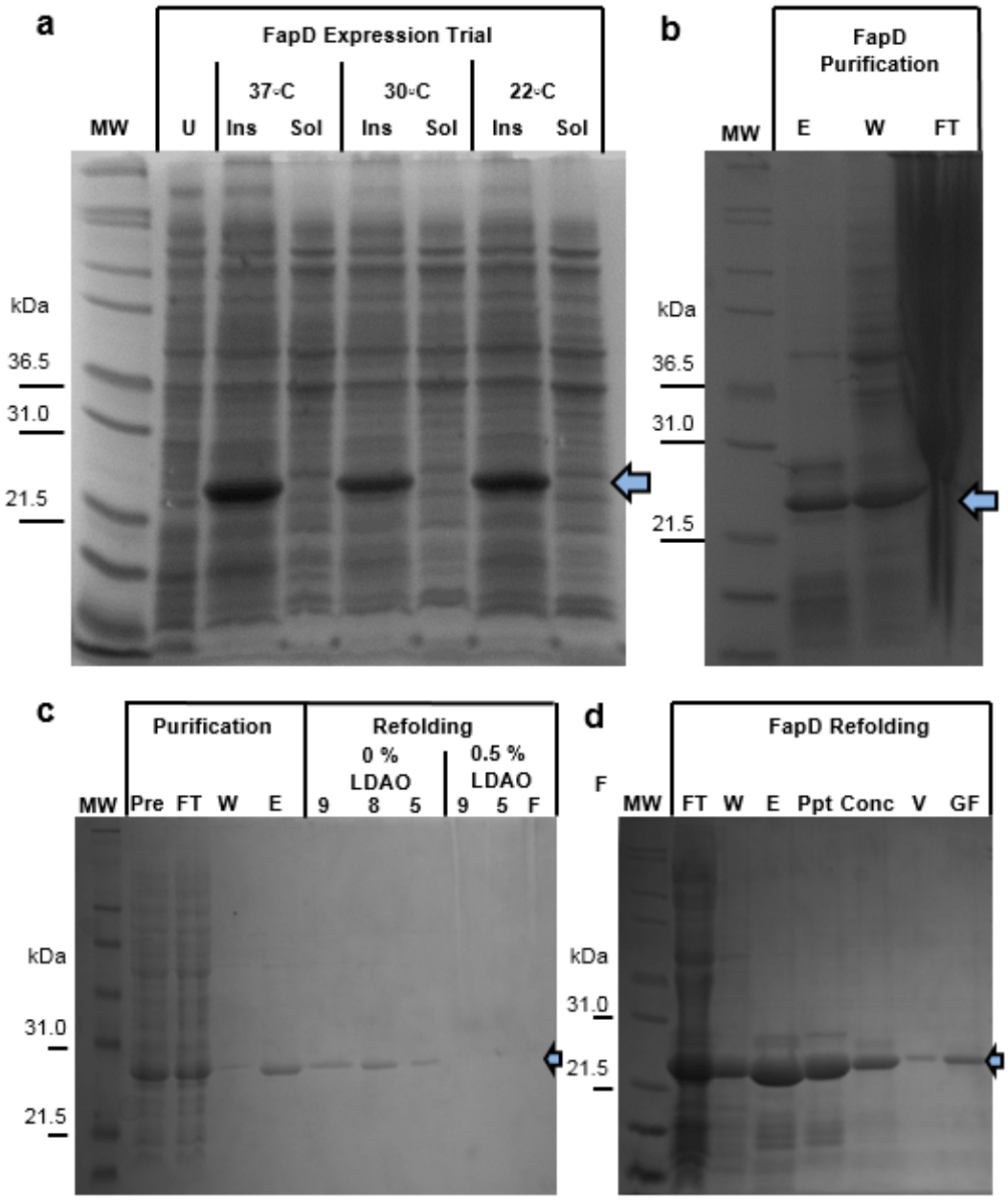


Figure 3.26: Expression, Purification and Refolding of FapD$^{45-226}$. a, SDS-PAGE gel of FapD expression trial showing the soluble (Sol) and insoluble (Ins) expression of FapD across a range of temperatures compared to the uninduced whole cell lysate (U), FapD is expressed insolubly at all three temperatures. **b**, SDS-PAGE gel of FapD denaturing purification showing Flow-through of Ni-purification (FT), the wash fraction (W) and the elution fraction (E), the protein is purified although a considerable amount of protein is lost in the wash, however this may be due to overloading of protein onto the column as overloading has also blurred the Flow-through. **c**, SDS-PAGE gel of FapD small scale refolding trials in 300 mM NaCl, at pH 9, 8, 5 and with LDAO at pH 9 and pH 5 as well as in LDAO with FapF present. The gel also shows the denaturing purification with the Flow-through of Ni-purification (FT), the wash fraction (W) and the elution fraction (E) showing FapD could be purified to a very good standard by Ni-IMAC but is unstable in the presence of LDAO. **d**, SDS-PAGE gel of FapD large scale refolding including the purification showing Flow-through of Ni-purification (FT), the wash fraction (W) and the elution fraction (E), a sample of the precipitant created during the refolding (Ppt), the concentrated protein after refolding (Conc) and the void (V) and the purified protein (GF) from gel filtration. Molecular Weight Ladder (MW) with the bands of relevant size labelled and blue arrows indicating the relevant protein bands are included on all the gels.

3.3.2 Expression Trials of Homologs

Homologs of FapD were identified in *Pseudomonas* PA7, *Pseudomonas* UK4, *Burkholderia* and in the Thermophilic Rice Compost Metagenome (TRCM). The genes were cloned into a pET46 vector, including some truncated versions of the PA7 and UK4 genes, based on bioinformatics predictions of disorder and domain boundaries. Expression trials showed that although several of the constructs including both the *Burkholderia* and TRCM FapD proteins expressed insolubly (Data Not Shown) was soluble expression of some truncations of both the PA7 and the UK4 versions of the genes, the constructs which expressed in the soluble fraction were successfully purified by Ni-IMAC (**Figure 3.27**). A periplasm targeted protein based on the OmpA signal sequence and pRSF1b was also produced and was shown to produce soluble protein (**Figure 3.27C**). The yield of FapD from the purifications was good for most of the constructs and the protein was relatively pure following gel filtration. The main exception to this was the periplasmic FapD construct which showed significantly lower levels of expression (**3.27C**), although the purity was good. It is not uncommon for proteins expressed to the periplasm to exhibit lower expression levels than the cytoplasm-targeted versions, although it is not always the case [**176**]. The soluble proteins were expressed and purified on a large scale, although the protein samples were somewhat unstable at room temperature precipitating gradually over time, 1D spectra were collected for the proteins showing they were at least partially folded (**Figure 3.28**).

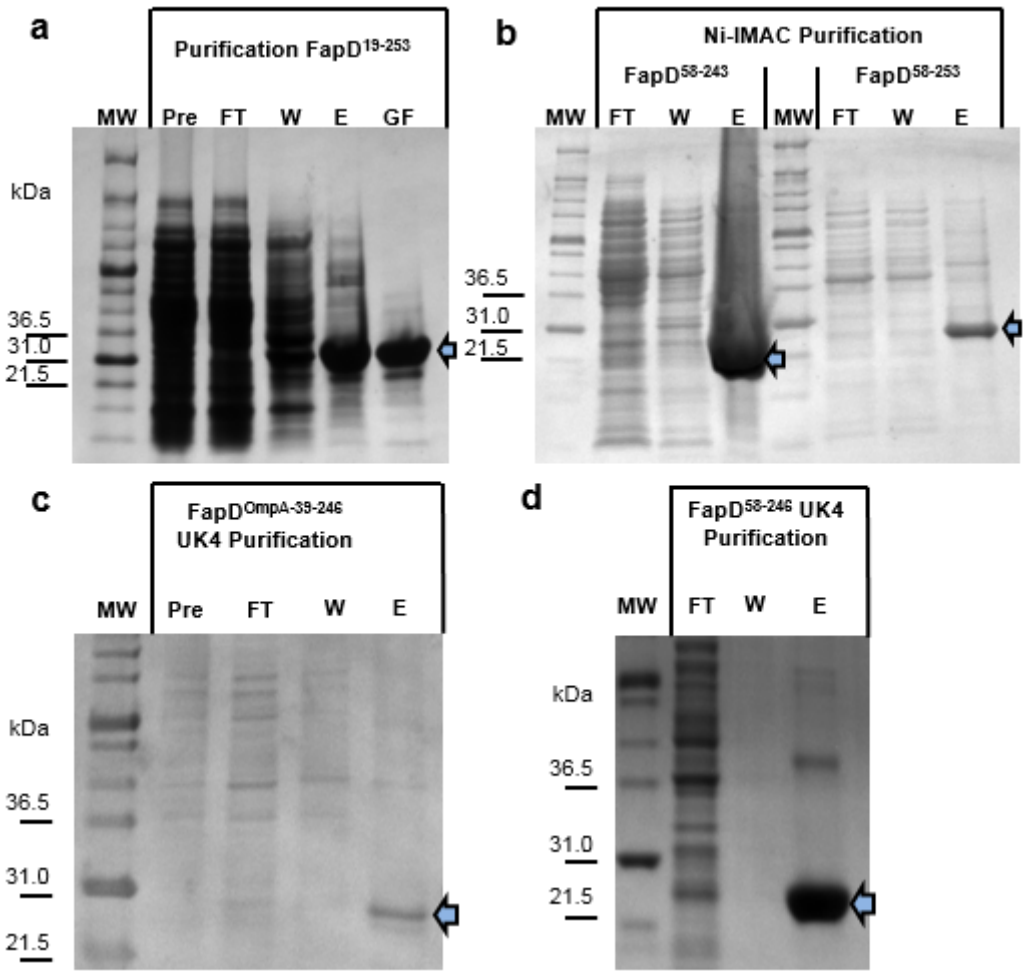


Figure 3.27: Example Purifications of FapD UK4 & PA7 Constructs. a, SDS-PAGE gel of PA7 FapD$^{19-253}$ native purification showing Flow-through (FT), the wash fraction (W), and the elution fraction (E) of Ni-purification as well as the gel filtered sample (GF). **b**, SDS-PAGE gel of PA7 FapD$^{58-243}$ and PA7 FapD$^{58-253}$ native purifications showing Flow-through of Ni-purification (FT), the wash fraction (W), the elution fraction (E) showing both proteins could be purified from the soluble fraction. **c**, SDS-PAGE gel of UK4 FapD$^{OmpAS-39-246}$ native purification showing Flow-through of Ni-purification (FT), the wash fraction (W), the elution fraction (E) showing that periplasmic FapD could be purified from the soluble fraction. **d**, SDS-PAGE gel of UK4 FapD$^{58-246}$ native purification showing the Flow-through (FT), the wash (W) and the elution fractions from Ni-purification. The proteins generally were produced with good yield, (apart from FapD$^{OmpA39-246}$. and purity sufficient for NMR however there was some precipitation overnight. Molecular Weight Ladders (MW) with the bands of relevant size labelled and blue arrows indicating the relevant protein bands are included on all the gels.

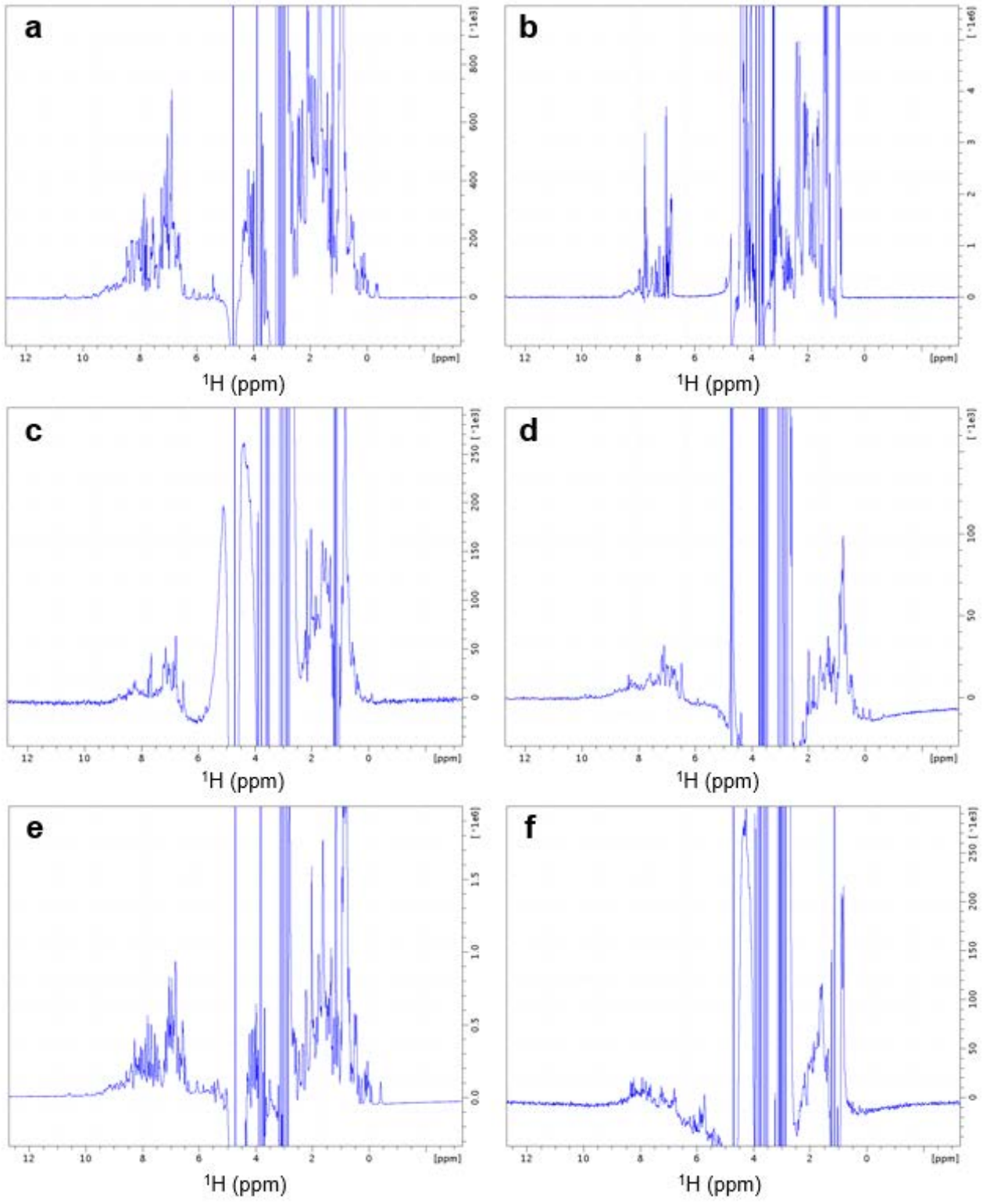
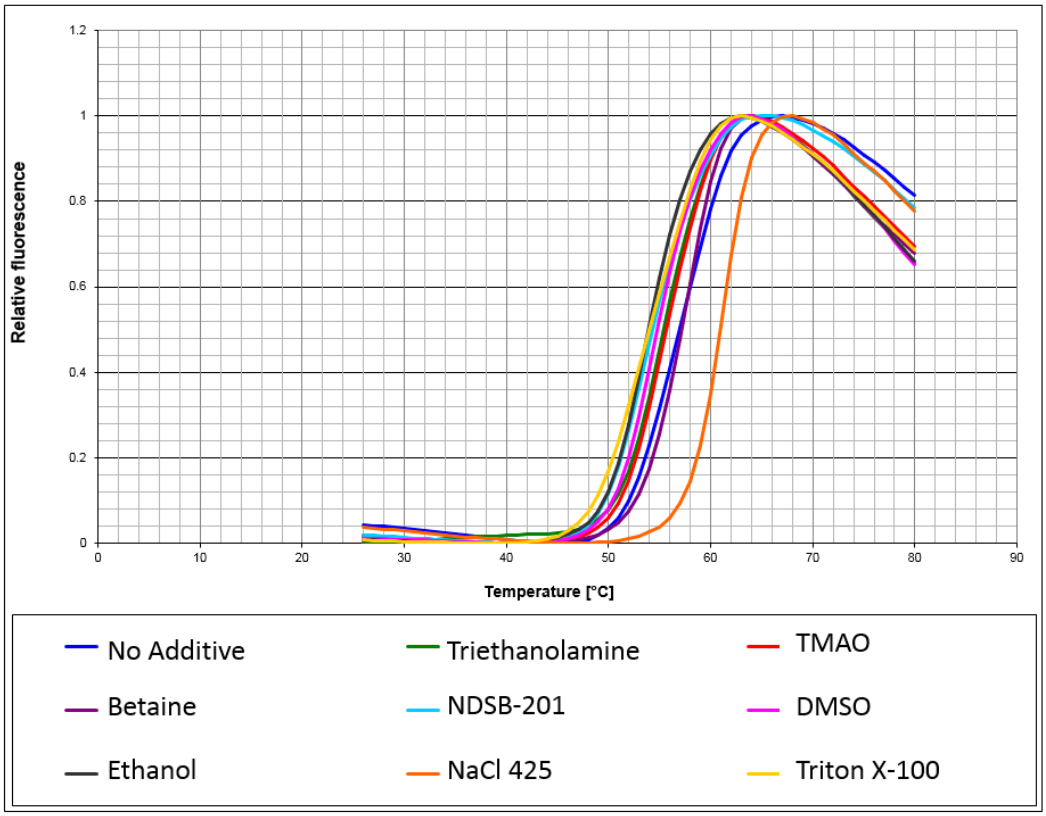


Figure 3.28: 1D Spectra of Soluble FapD Homologs. a, FapD$^{19\text{-}253}$ PA7 **b**, FapD$^{58\text{-}236}$ UK4 **c**, FapD$^{58\text{-}246}$ UK4 **d**, FapD$^{39\text{-}236}$ UK4 **e**, FapD$^{58\text{-}253}$ PA7 **f**, FapD$^{58\text{-}243}$ PA7. The spectra all show the protein to be at least partially folded, apart from FapD$^{58\text{-}236}$ UK4 and FapD$^{58\text{-}243}$ PA7 (both containing C-terminal truncations) which appears unfolded, with the best peak dispersion and number of peaks present for FapD$^{58\text{-}253}$ PA7 construct. All six constructs were in 250 mM NaCl, 20 mM HEPES pH 7.5.

3.3.3 Optimisation of FapD PA7 Stability

For NMR experiments it is necessary for a protein to be stable for a minimum of 3 days and preferably a couple of weeks to allow time for data collection. Since FapD$^{58-253}$ PA7 showed good expression, was the most stable and produced one of the better NMR spectra it was used for optimisation experiments. Additives were screened to test their effect on the protein stability, a broad range were initially added during the purification on a small scale and the amount of soluble protein measured qualitatively by SDS-PAGE (**Data Not Shown**). Additives were then selected from these results and DSF was used to test the quantitative effect on protein stability (**Figure 3.29**). The DSF results indicated that although the additives may improve the soluble yield of FapD many did not improve the thermal stability of the protein and several had a detrimental effect; however high concentrations of Glycerol (10 % *w/v*), Potassium Phosphate (100 mM) or Sodium Chloride (425 mM) did improve the thermal stability.



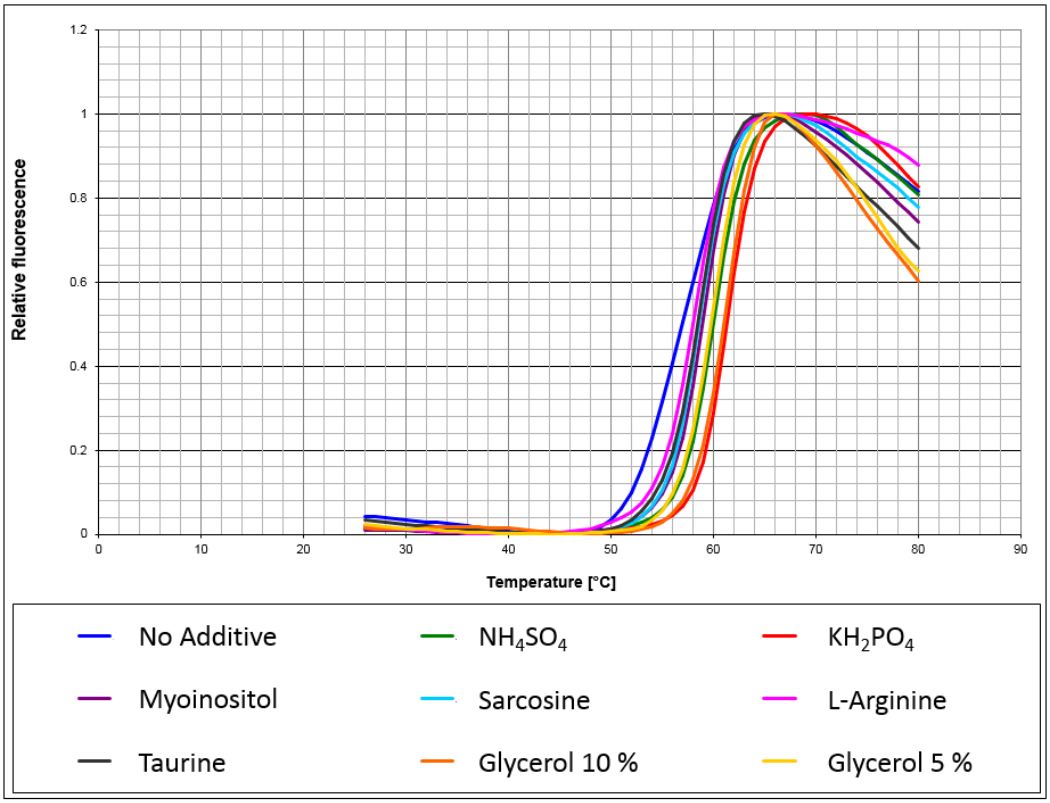


Figure 3.29: FapD Thermal Stability – Differential Scanning Fluroimetry (DSF). Graphs showing the normalised thermal denaturation curves of FapD in the presence and absence of a selection of additives. The results are separated into two separate graphs are included for clarity, despite the experiments being carried out concurrently. Notably Triton X-100 and Ethanol have a detrimental effect on the Stability of FapD as measure by the melting temperature (Tm) values (temperature at 0.5 relative fluorescence). While similarly Glycerol, Potassium Phosphate and a higher concentration of NaCl increased the Tm of the protein.

Since the best additives identified in the DSF screen were additional NaCl, potassium phosphate and glycerol. FapD was purified in the presence of these additives and tested for long term stability by 1D NMR. The potassium phosphate and increased NaCl samples precipitated before data could be collected but 1D data for glycerol was collected (**Figure 3.30A**). Glycerol appears to stabilise the protein over the period of several days, however it has an adverse effect on the spectra due to relaxation. Some alternative NMR buffers such as a Glutamine-Arginine buffer [**177**] were also tested (**Figure 3.30B**) but none produced a beneficial effect on stability.

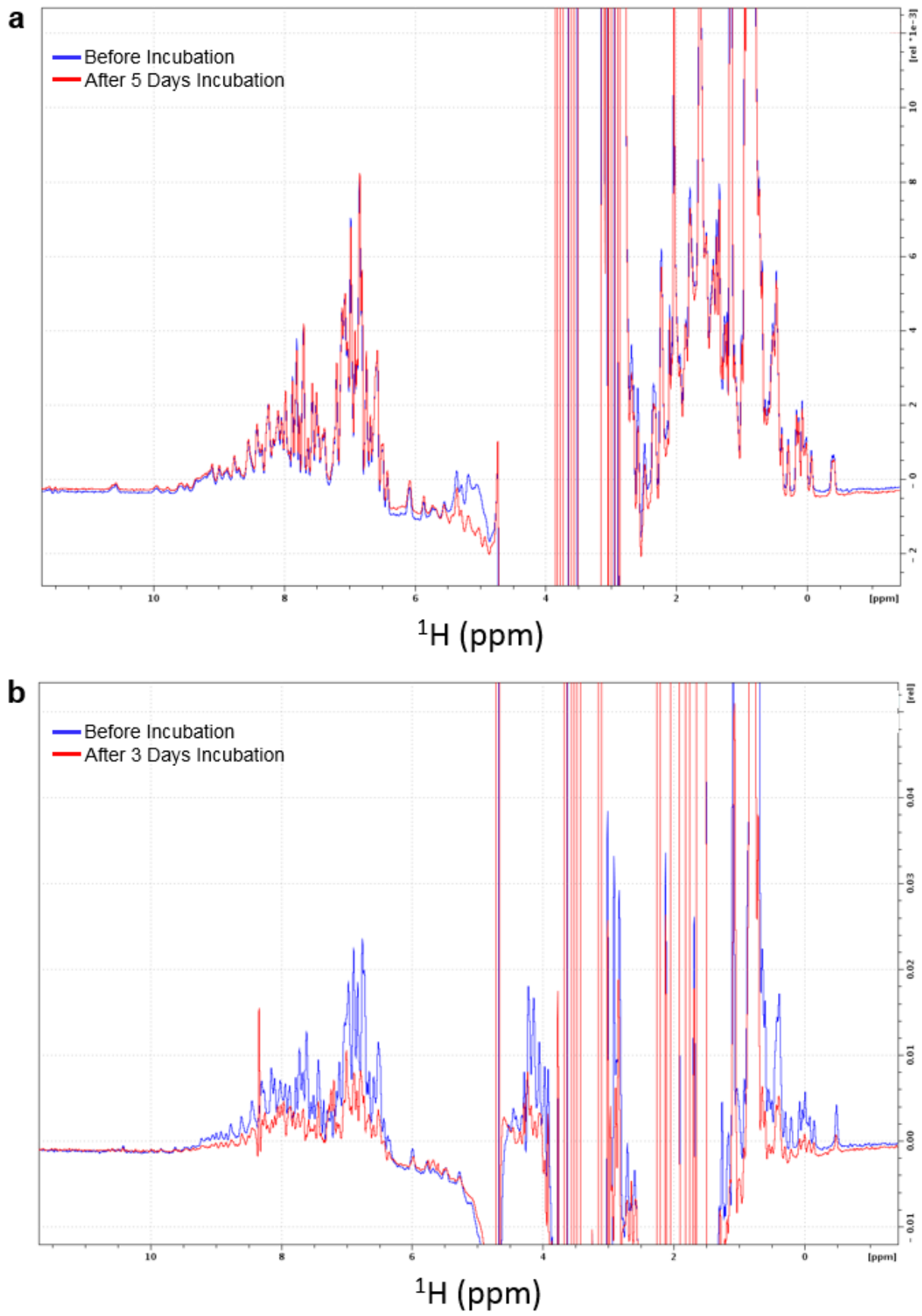


Figure 3.30: Stability Trials of FapD. a, Overlay of 1D spectra of FapD$^{58-253}$ PA7 in 10 % glycerol 250 mM NaCl, 20 mM HEPES pH 7.5 before (Blue) and after (Red) 5 day incubation at room temperature, showing no significant changes in signal. **b**, Overlay of 1D spectra of FapD$^{58-253}$ PA7 in Glutamine-Arginine NMR Buffer [**177**] before (Blue) and after (Red) 3 day incubation at room temperature showing a significant loss of signal due to sample precipitation.

To improve the NMR spectra but retain the stabilising effect glycerol was compared at 10% (*v/v*) (**Figure 3.30A**) 5 % (*v/v*) (**Figure 3.31A**) and 2 % (*v/v*) (**Figure 3.31B**) and the effects on stability compared, 2 % (*v/v*) glycerol was insufficient to stabilise the protein for a 5 day period but both 10 % (*v/v*) and 5 % (*v/v*) effectively stabilised the protein and produced well dispersed ^{15}N ^{1}H 2D spectra (**Figure 3.31C).**

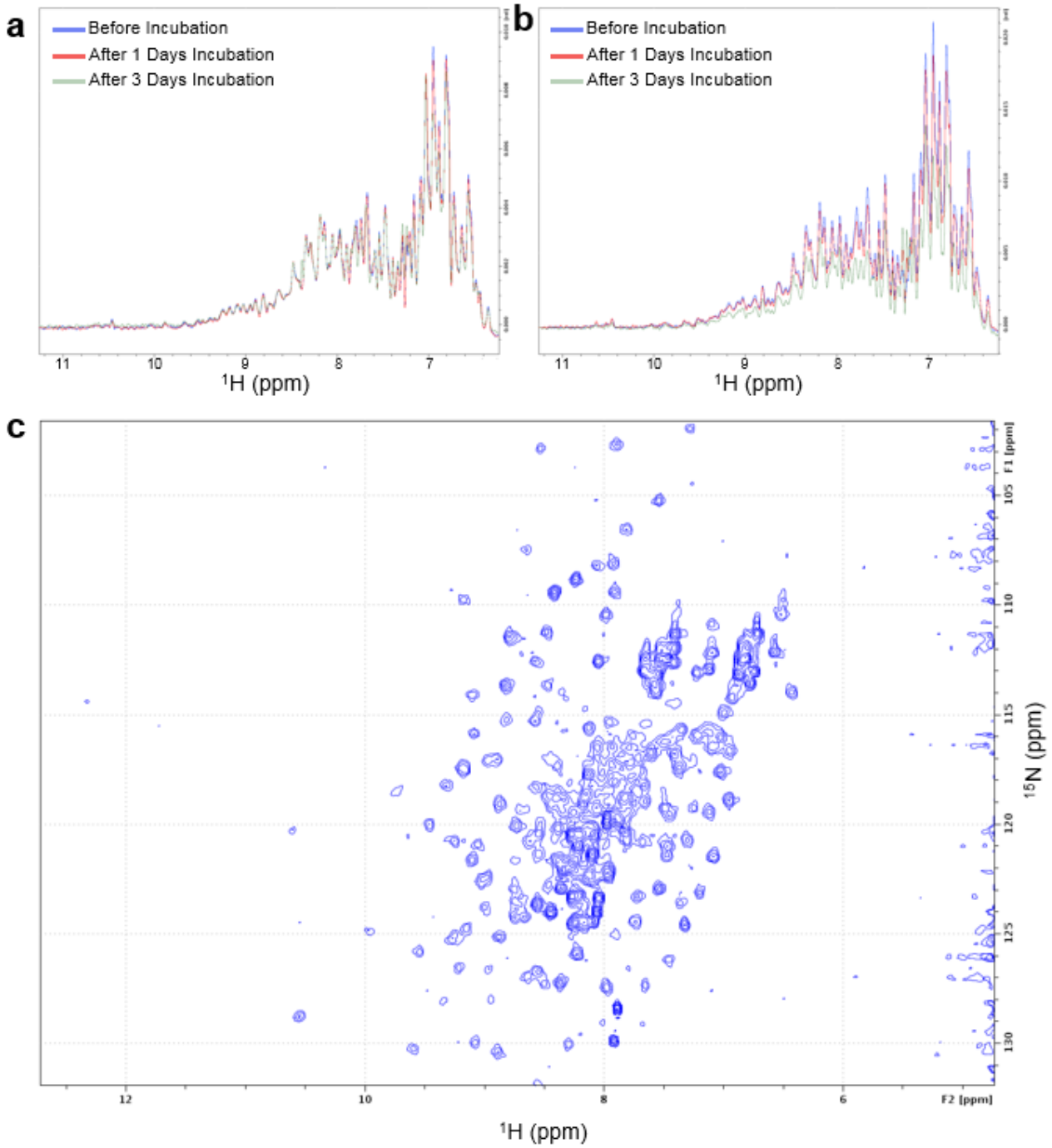


Figure 3.31: Stability of FapD in glycerol. a, Overlay of the amide proton region of 1D spectra of FapD$^{58\text{-}253}$ PA7 in 250 mM NaCl, 20 mM HEPES pH 7.5, 5 % (*v/v*) glycerol before (Blue), after 1 day (Red) and after 3 days (Green) incubation at room temperature, showing no significant changes in signal. **b**, Overlay of the amide proton region of 1D spectra of FapD$^{58\text{-}253}$ PA7 in 250 mM NaCl, 20 mM HEPES pH 7.5, 2 % (*v/v*) glycerol before (Blue), after 1 day (Red) and after 3 days (Green) incubation at room temperature showing a perceptible (~30 %) loss of signal due to sample precipitation. **c**, ^{15}N ^{1}H HSQC 2D spectra of FapD$^{58\text{-}253}$ PA7 in 250 mM NaCl, 20 mM HEPES pH 7.5, 5 % (*v/v*) glycerol showing the protein is stable and the spectra relatively well dispersed.

FapD contains a conserved cysteine residue in the active site which should be exposed and may affect the stability of the protein. Two approaches were adopted for dealing with this, the reducing agent DTT was used to reduce the cysteine and prevent potential intermolecular disulphide bridges forming and a C67A mutant version of the gene was made to eliminate any potential protease activity (although the precipitant showed no evidence of degradation (**Figure 3.26D**). 1D NMR was used to test the effect of the DTT and C67A mutant on the stability of FapD, some minor improvement was seen with the addition of DTT to glycerol, although DTT alone was insufficient to stabilise the protein for NMR (precipitated in hours) in the presence of 10 % (*v/v*) glycerol the protein was stable over a week at room temperature (**Figure 3.32A**). The FapD C67A mutant was also stable in the presence of 10 % (*v/v*) glycerol for at least 5 days (**Figure 3.32B**). The FapD C67A did not show any improvement in stability as the protein was still unstable over a time course of days in the absence of glycerol (**Figure 3.33A**). Comparison of the spectra of FapD C67A and FapD (**Figure 3.33B**) indicated that although there were some small differences the proteins displayed a similar degree of structure and probably retained the same fold. Since FapD C67A behaved similarly to the wild type protein this suggests the protease activity, if present, had little effect on FapD stability.

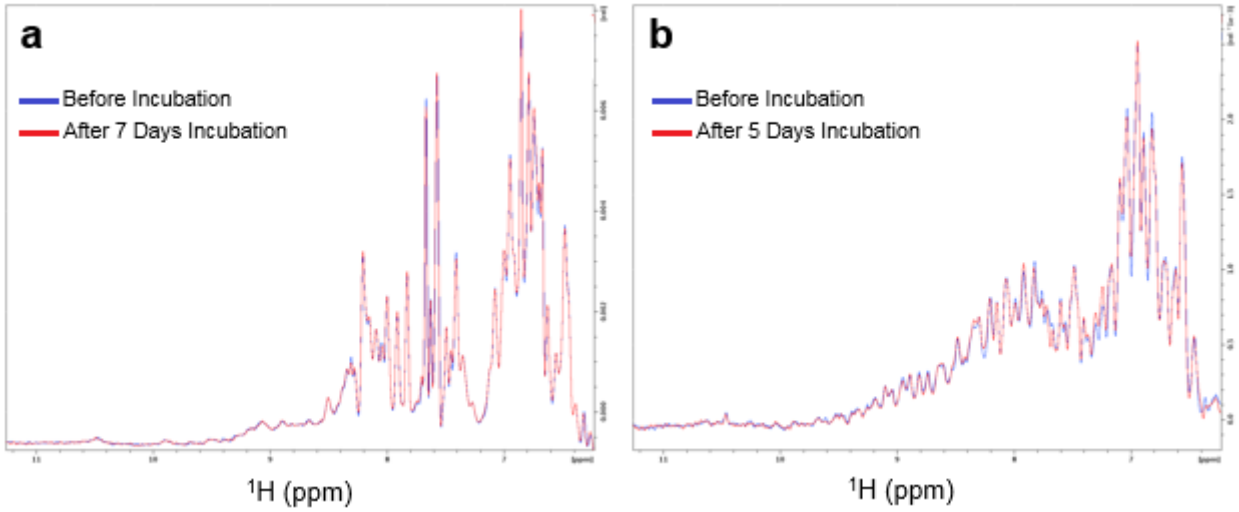


Figure 3.32: Stability Trials of FapD$^{58-253}$ PA7 with DTT and C67A substitution. a, Overlaid 1D ^{1}H NMR spectra of FapD$^{58-253}$ PA7 250 mM NaCl, 20 mM HEPES pH 7.5, with 5 mM DTT and 10 % (*v/v*) glycerol showing the spectra before (blue) and after (red) one week incubated at room temperature (~20 ◦C). **b**, Overlaid 1D ^{1}H NMR spectra of FapD$^{58-253}$ PA7 with C67A substitution in 250 mM NaCl, 20 mM HEPES pH 7.5, 10 % (*v/v*) glycerol showing the spectra before (blue) and after (red) five days incubated at room temperature (~20 ◦C). The spectra both show almost no change in the peak intensities or positions indicating the proteins remain stable over the period of the incubation.

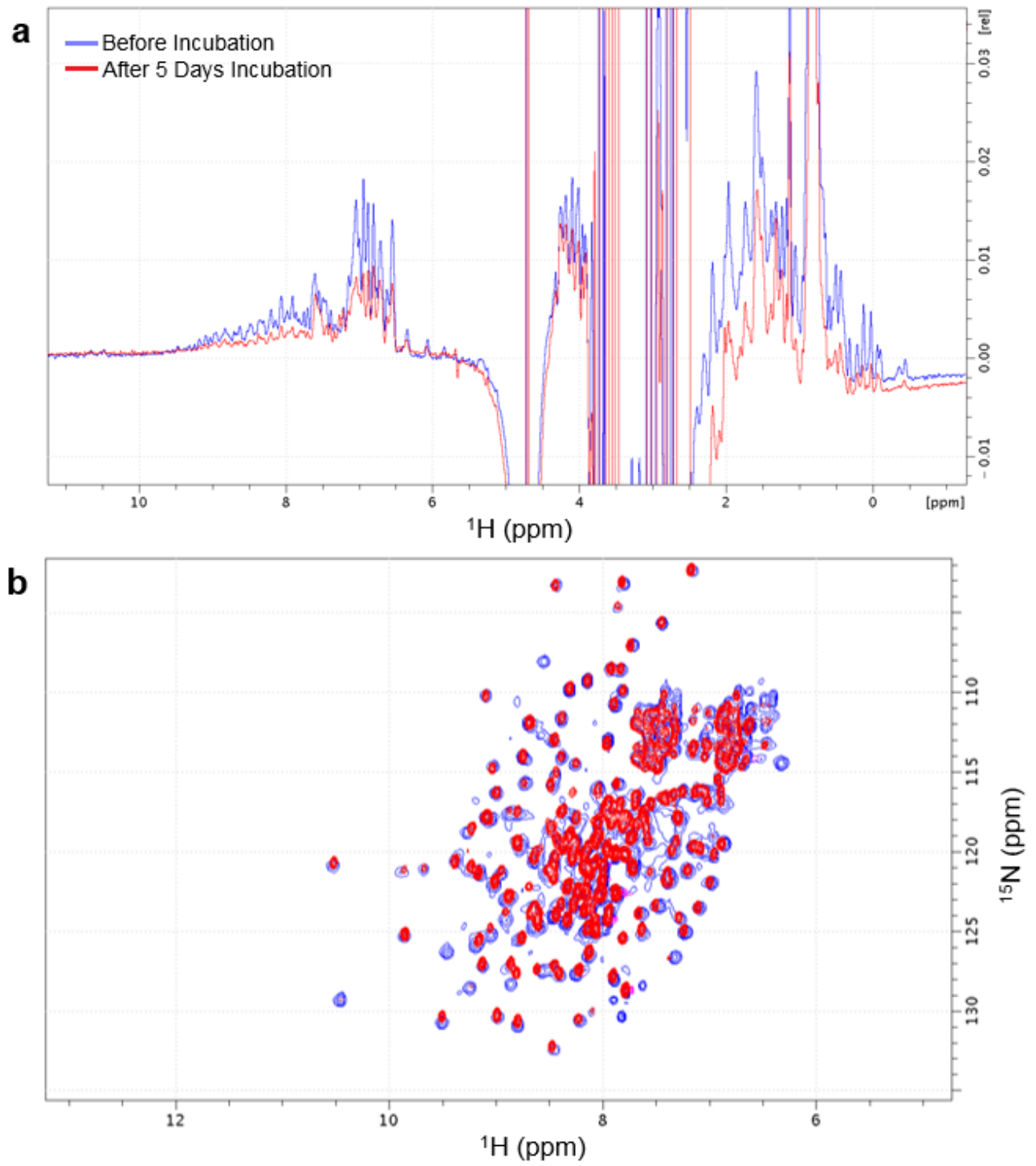


Figure 3.33: NMR of FapD C67A. a, Overlay of 1D spectra of FapD$^{58\text{-}253}$ C67A in 250 mM NaCl, 20 mM HEPES pH 7.5, 0 % (*v/v*) glycerol before (Blue) and after (Red) 1 day incubation at room temperature showing a significant loss of signal, probably due to precipitation **b**, Overlay of FapD$^{58\text{-}253}$ ^{15}N (Blue) and FapD$^{58\text{-}253}$ C67A (Red) ^{15}N ^{13}C ^{2}H, ^{1}H ^{15}N Transverse Relaxation optimised spectroscopy (TROSY) 2D spectra in 250 mM NaCl, 20 mM HEPES pH 7.5, 10 % (*v/v*) glycerol showing the proteins possess similar structure.

3.3.4 Backbone experiments for FapD PA7

Since FapD was stabilised by glycerol and DTT and produced a reasonably dispersed spectrum with a peak number close to that expected for FapD, triple labelled FapD was prepared, exchanged back into water and used to collect a TROSY spectra and some backbone experiments were conducted to see if assignment was feasible (**Figure 3.34**). Although the 2D was good enough for assignment (**Figure 3.34A**) the backbone data was incomplete with some strips lacking carbonyl peaks which will make the protein backbone difficult to assign (**Figure 3.34B**).

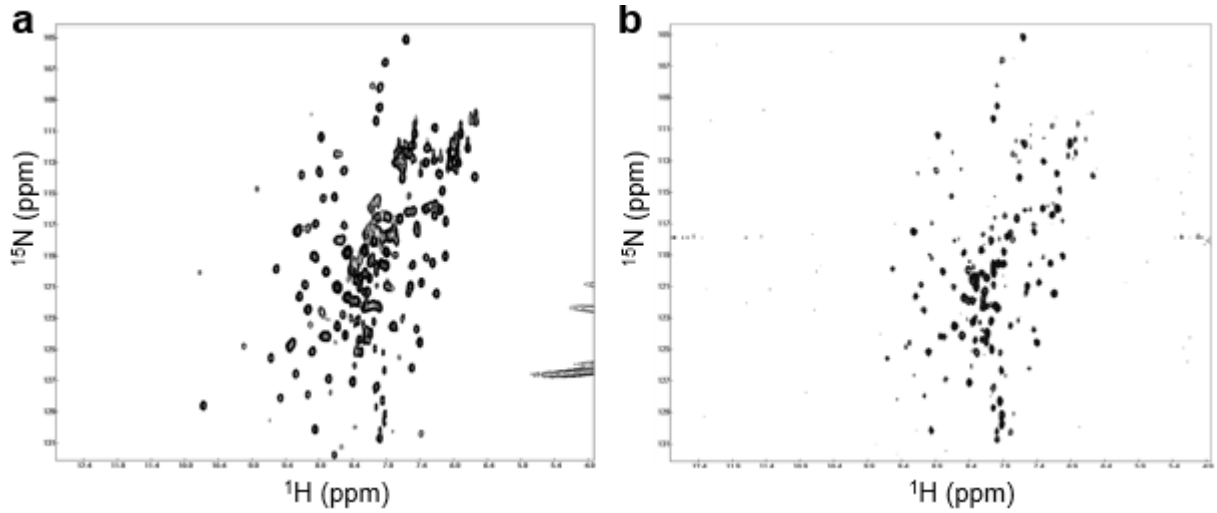


Figure 3.34: Backbone experiment data for FapD C67A. a, ^{1}H ^{15}N TROSY 2D spectra of FapD C67A in 250 mM NaCl, 20 mM HEPES pH 7.5, 10 % (*v/v*) glycerol, the protein is structured but with some large unstructured regions indicated by the clustering of peaks near the centre of the spectrum. **b**, 2D Projection of the HNCO data showing the low peak intensity and absence of several peaks which would be expected to be present at the same position as ^{15}N ^{1}H amide peaks in the HSQC (Figure 3.34A).

3.3.5 Summary

Various approaches were made in an attempt to produce refolded soluble FapD (PAO1): adjustment of the construct boundaries, removal of potential N-terminal transmembrane helix and optimisation of refolding conditions. Soluble FapD could be produced using an N-terminally truncated construct refolded at pH 5 but was highly unstable, precipitating rapidly over time. Constructs of homologs of FapD from *Pseudomonas* UK4 and PA7 strains were designed, incorporating the truncations used in PAO1, and produced successfully. These FapD homologs were successfully purified under native conditions and were shown to be folded using 1D NMR. The protein was still observed to be unstable and so the construct which produced the best NMR spectra ($FapD^{58-253}$ PA7) was used for buffer optimisation using a combination of DSF and NMR. Glycerol was identified as the best additive for stabilising FapD for NMR experiments at a concentration between 5 and 10 % (*v/v*). A point mutation was also made to remove the predicted active cysteine this was shown to have no effect on stability. HNCO experiments were conducted using the mutated, truncated PA7 construct with 10 % (*v/v*) Glycerol, but the relaxation in such a high concentration of glycerol would limit our ability to collect data from less sensitive experiments making backbone assignment difficult.

3.4 FapF

3.4.1 Refolding and Purification of FapF

The pNIC-NTH-FapF construct with the signal sequence removed could be expressed in large quantities in BL21 *E. coli* cells where the protein was found to be present in the insoluble fraction (**Figure 3.3**). This protein could be purified by lysing cells expressing FapF, pelleting the insoluble material, dissolving the protein in 8 M urea and then purifying by denaturing Ni-IMAC (**Figure 3.35**).

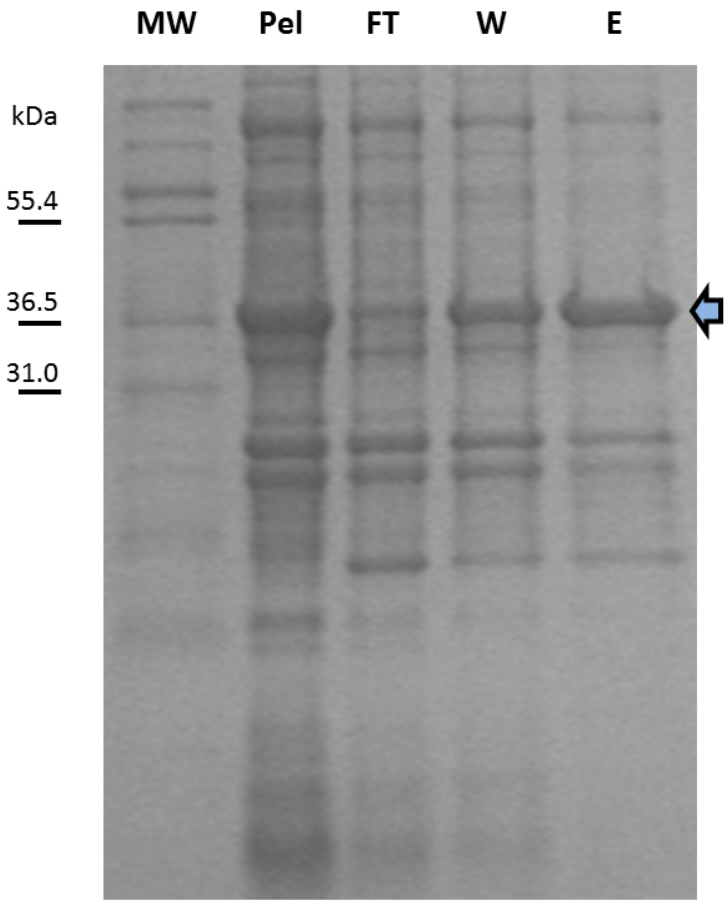


Figure 3.35: Denaturing Purification of FapF PAO1. SDS-PAGE gel of FapF Ni-IMAC purification conducted under denaturing conditions. The band of purified FapF in the elution (E) is indicated with the blue arrow, some protein is lost in the wash step (W) and the initial flow through from the column (FT), however the final fraction from the elution is purer than the pellet fraction (Pel).

Following the denaturing purification FapF was sufficiently pure to attempt to refold by pulse refolding. The purified fraction of FapF in 8 M urea was dripped slowly into 20x volume of 5 % (*v/v*) LDAO, 300 mM NaCl, 20 mM TrisHCl pH 8. This diluted the urea below the concentration where it is likely to denature a given protein and provided a large concentration of detergent to form micelles containing FapF. This refolding solution was dialysed to reduce the concentration of LDAO and remove the excess imidazole and urea. The sample was then concentrated using Ni-IMAC and the resulting elution gel filtered to remove any remaining impurities (**Figure 3.36**). This sample could be checked for folding by 1D ^{1}H NMR and could be seen to be well folded (**Figure 3.37**).

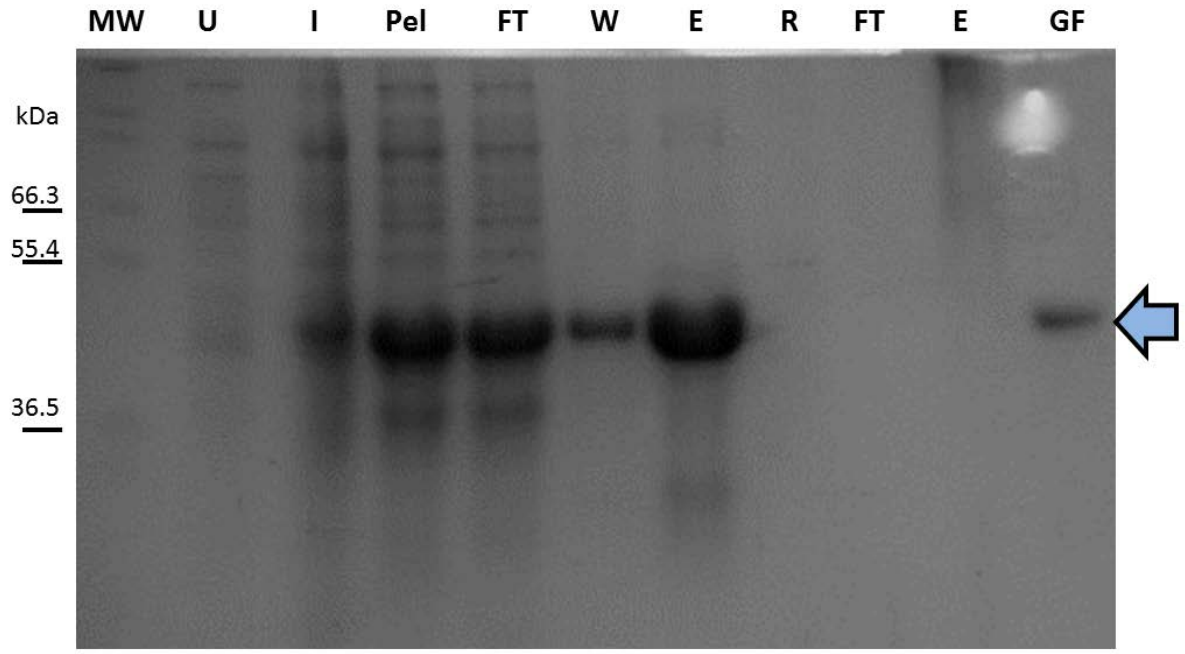


Figure 3.36: Purification and Refolding of FapF. SDS-PAGE gel of FapF refolding. Showing Molecular Weight Ladder (MW), Uninduced Cells (U), Induced Cells after overnight expression (I), Insoluble Pellet from Fractionation (Pel), Flow through of Ni-NTA column (FT), Wash (W), Elution (E), Refolding Solution (R) blank due to high dilution, FT of second Ni-NTA and then elution from second Ni-NTA purification (E), blurred due to high LDAO concentration and finally gel filtration fraction (GF). The molecular weight corresponding to FapF is indicated with blue arrow.

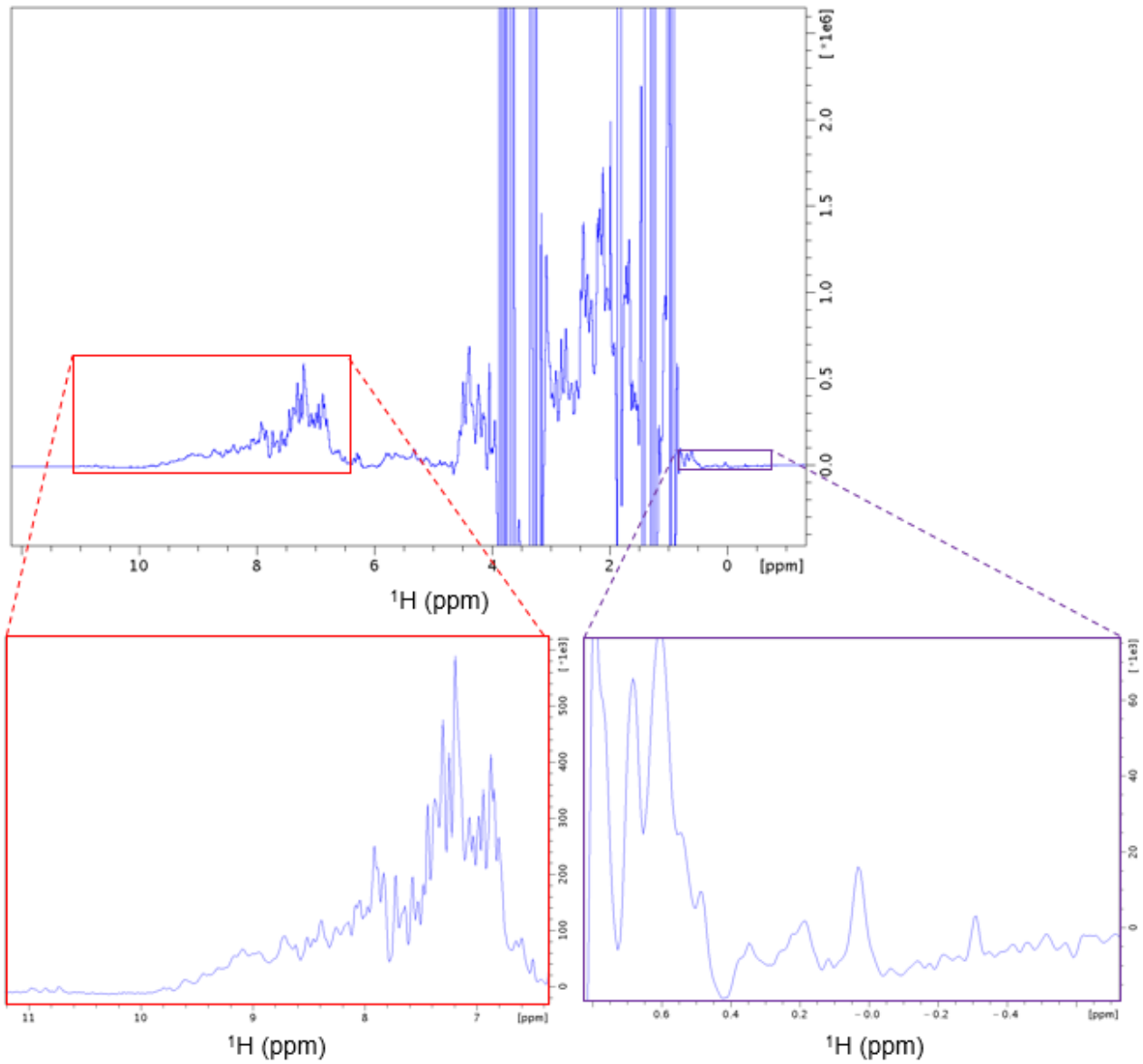


Figure 3.37: 1D ^{1}H NMR spectra of FapF. The spectra shows that the protein is folded as there are peaks above 9 ppm and below 0 ppm, as illustrated by the red and purple insets respectively, however due to the large size of the protein-micelle complex, (expected to be around 77 kDa with the 45 kDa protein in an LDAO micelle which is ~22 kDa in isolation), the peaks are broader than would be seen for a smaller protein. Spectra was collected at 310˚K in 0.1% (*v/v*) LDAO, 250 mM NaCl, 20 mM TrisHCl pH 8 on the Avance II 800 MHz.

3.4.2 SEC-MALS Analysis of FapF

FapF was transferred into three different detergent buffers LDAO, β-Octyl-Glucoside βOG and Tetraethylene Glycol Monooctyl Ether (C8E4). The samples were then analysed by Size Exclusion Chromatography – Multi-Angle Light Scattering Analysis (SEC-MALS) to determine the oligomeric state and to estimate the amount of detergent bound to the protein. MALS is an analytical technique that allows the determination of absolute molar mass and size of macromolecules based on the scattering of light by a protein solution with a known concentration. The oligomeric state observed for the protein appeared to vary with the detergent used with LDAO appearing to be dimeric, βOG trimeric and C8E4 tetrameric with bound detergent varying only from around 60 kDa to 75 kDa (**Figure 3.38, 3.39, 3.40**). Out of the three samples the FapF in LDAO appeared to be the best sample with a cleaner trace and the most accurately determined molecular weight. It is possible that this difference may reflect incomplete exchange of detergent when transferring from the FapF solubilised in LDAO to the other detergents rather than purely the quality of the sample in those detergents. To avoid this solubilisation of FapF directly into other detergents was attempted during the purification but was not efficient resulting in very poor yields (data not shown). Based on this data LDAO was used as the primary detergent for both solubilisation and crystal studies for most of the experiments involving FapF.

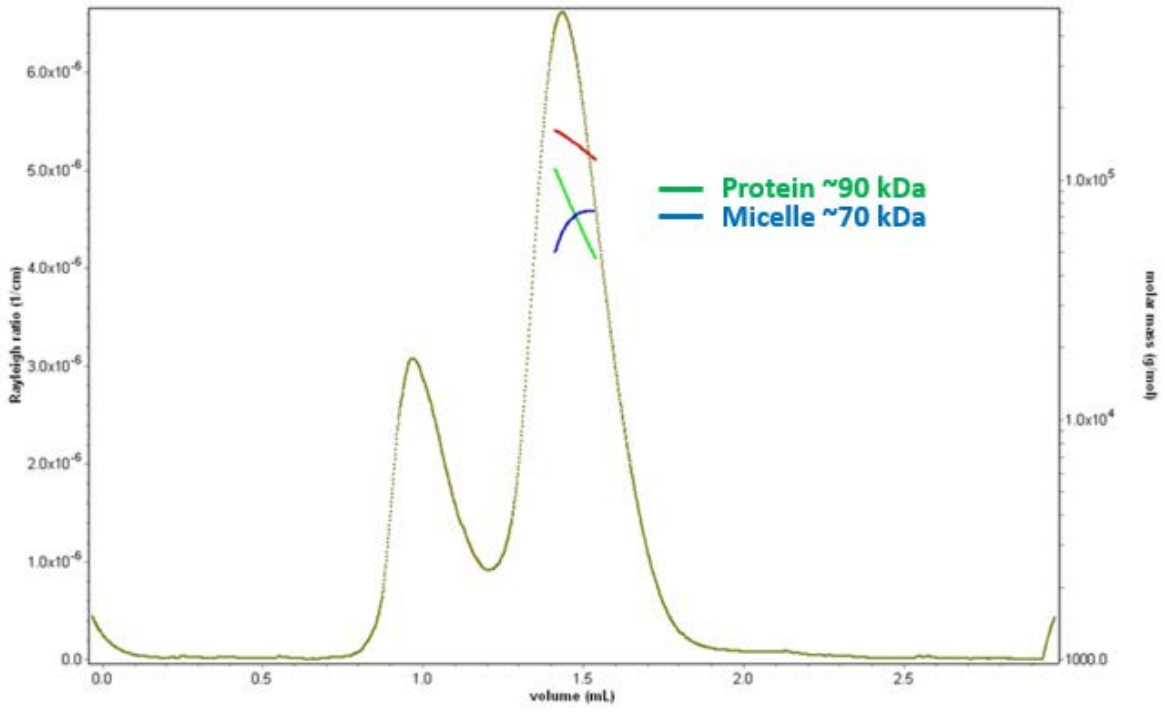


Figure 3.38: SEC-MALS data for FapF in LDAO. The light scattering is shown in olive green, with the estimated overall mass, protein mass and micelle mass shown in red, green and blue respectively. The data indicates that FapF is a dimer in LDAO (250 mM NaCl, 20 mM TrisHCl pH 8, 0.1 % (*w/v*) LDAO) which is bound by a similar mass of detergent. The sample appears reasonably mono-dispersed.

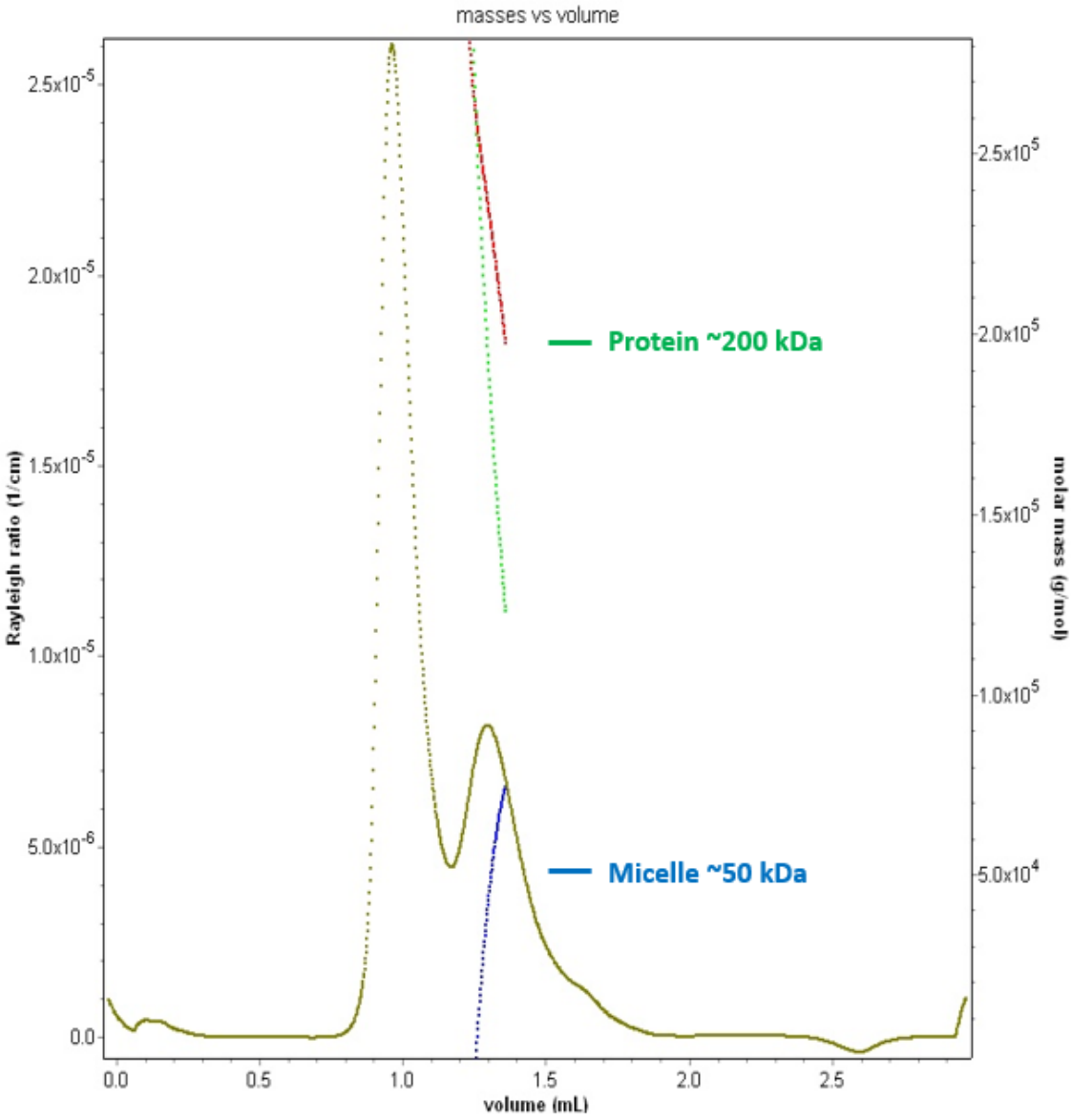


Figure 3.39: SEC-MALS data for FapF in C8E4. The light scattering is shown in olive green, with the estimated overall mass, protein mass and micelle mass shown in red, green and blue respectively. The data indicates that FapF may be a tetramer in C8E4 (250 mM NaCl, 20 mM TrisHCl pH 8, 16 mM C8E4), which is bound by around 50 kDa detergent however there is considerable overlap with a large peak in the void and a smaller molecular weight shoulder, indicating the sample is polydispersed. The large void may be due to instability in C8E4 and the smaller shoulder may indicated the retention of some LDAO bound protein both of which may add to the polydispersity.

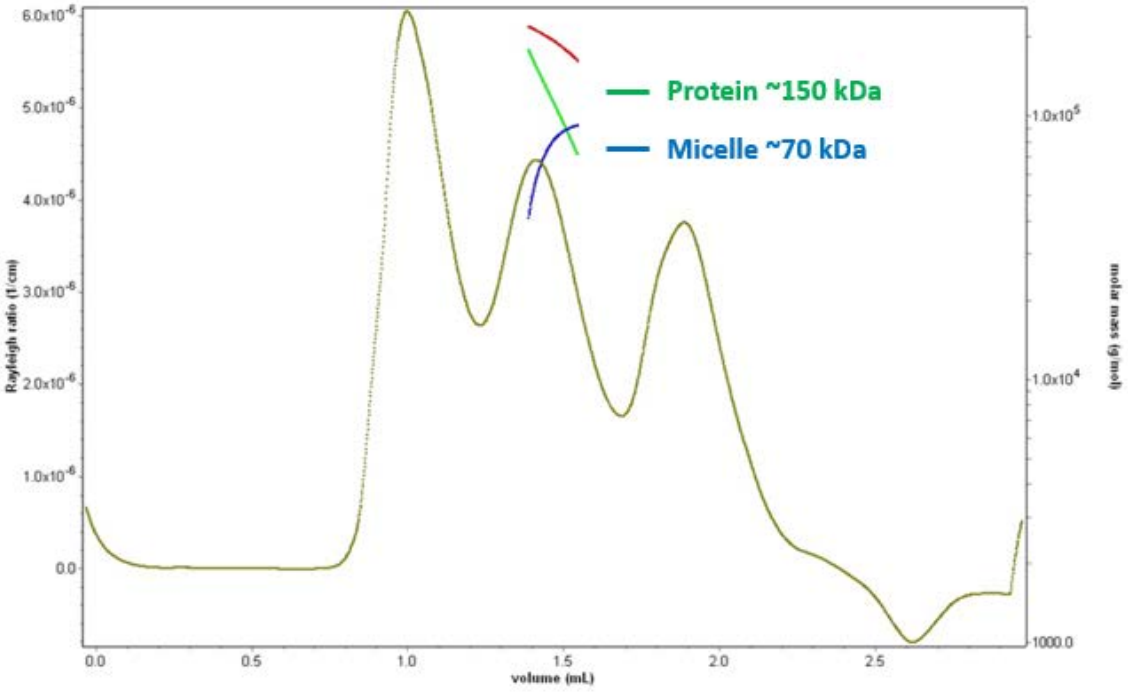


Figure 3.40: SEC-MALS data for FapF in βOG. The light scattering is shown in olive green, with the estimated overall mass, protein mass and micelle mass shown in red, green and blue respectively. The data indicates that FapF may be a trimer in βOG (250 mM NaCl, 20 mM TrisHCl pH 8, 50 mM βOG) which is bound by around 70 kDa detergent however there is overlap with a large peak in the void which may be increasing the apparent scattering and so the protein may still be dimeric as in LDAO. The third peak shows no UV absorbance in the original data and so is likely to be the isolated βOG micelle.

3.4.3 NMR of FapF

The stability of FapF was tested at a range of temperatures using the unlabelled protein and the results indicated the protein was stable at to 37˚C under the conditions tested (**Figure 3.41**). ^{15}N labelled FapF was purified and refolded using the same method as the unlabelled protein and was used to conduct a 2D ^{1}H ^{15}N TROSY experiment. The Spectra showed that the protein was folded with well dispersed peaks (**Figure 3.42**), however the number of clearly defined peaks is lower than expected for a protein of FapF's size and the clustering of many peaks near the centre of the spectra indicates that there is a considerable disordered region in the protein. Many β-barrels contain long N-terminal plug domains which may be disordered in solution **[178, 179, 180]**, it is possible that this is the large disordered region in the protein.

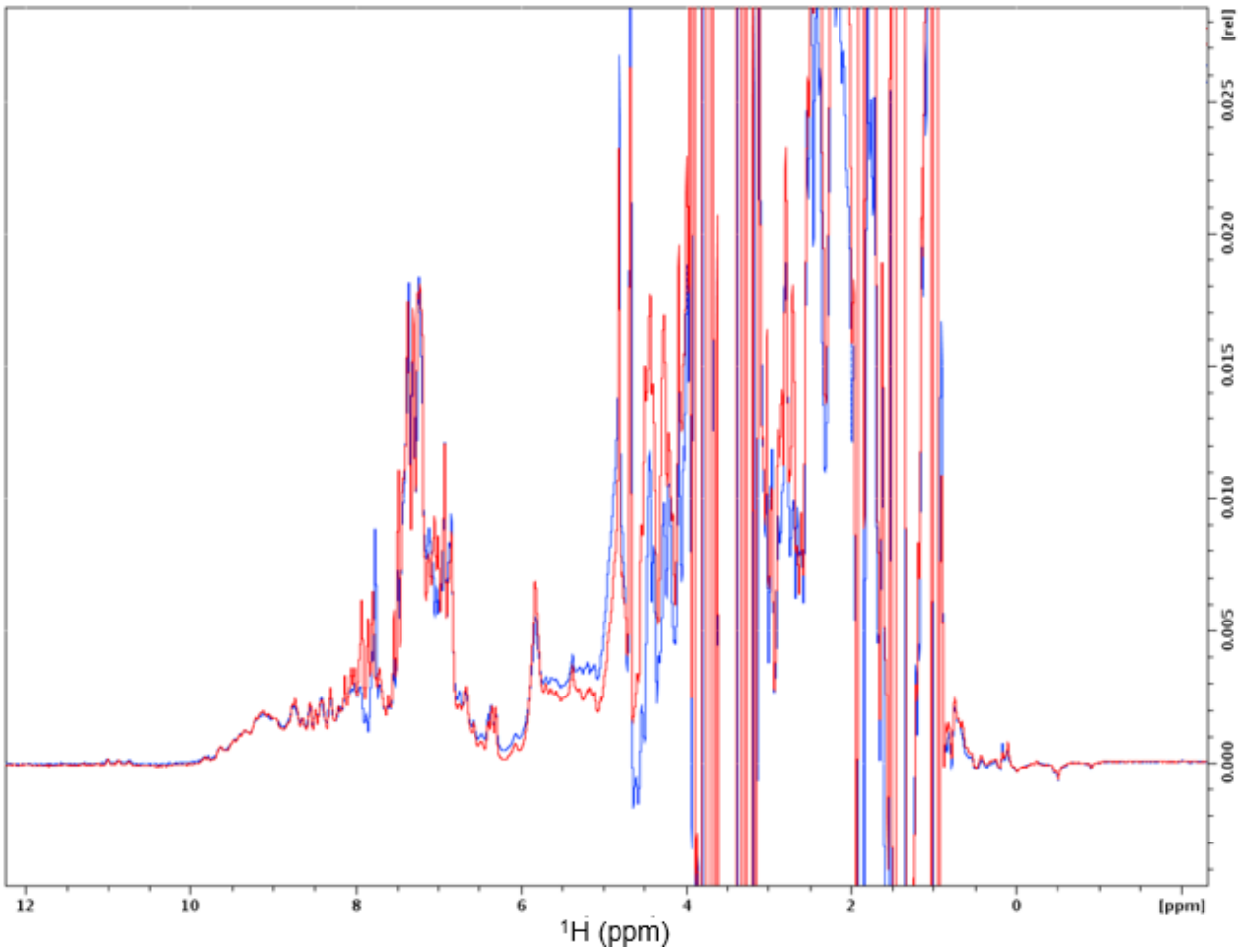


Figure 3.41: Overlay of 1D NMR spectra showing stability of FapF at 310 ˚ K. FapF before (Blue) and after (Red) overnight incubation at 310˚K. The protein appears to be stable as there is no decrease in peak intensity or change in peak position. Spectra were collected at 310˚K in 0.1% (*v/v*) LDAO, 250 mM NaCl, 20 mM TrisHCl pH 8 on the Avance II 800 MHz.

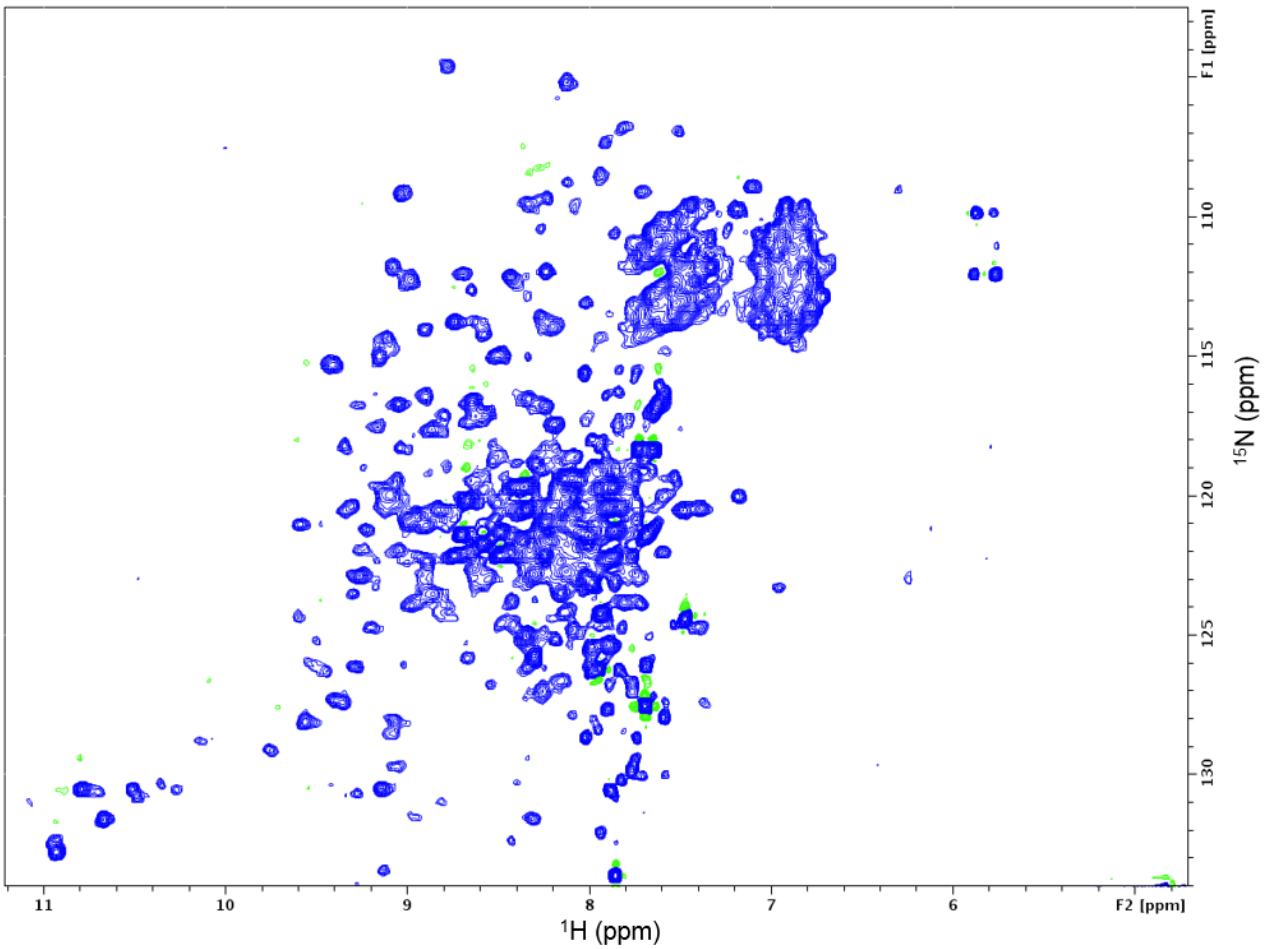


Figure 3.42: 2D ^{15}N ^{1}H TROSY spectra of FapF. The distribution of peaks indicates that there is a folded domain in the sample but there is a large disordered region, perhaps corresponding to the predicted long disordered N-terminal extension. Due to the peak clustering in the centre it is hard to count the peaks exactly but it is likely that there are less peaks than expected for the 45 kDa protein. Spectra were collected at 310˚K in 0.1% (*v/v*) LDAO, 250 mM NaCl, 20 mM TrisHCl pH 8 on the Avance II 800 MHz.

3.4.4 Crystal Trials FapF

FapF was expressed, refolded and purified as previously before transferring into a low salt buffer for crystallisation (50 mM NaCl, 10 mM TrisHCl, 0.1 % (*v/v*) LDAO, pH 8.0) by gel filtration. The sample was concentrated to ~13 mg/ml using a 50 kDa MWCO concentrator. Crystal trials were conducted using a series of commercial screens at 13 mg/ml and 6.5 mg/ml concentration in 200 nL to 200 nL drops. The trays were incubated at 20˚C and microcrystals were identified after one week in a single condition (0.02 M TrisHCl pH 7.5, 22 % (*w/v*) PEG 550 MME) at 6.5 mg/ml (**Figure 3.43**). The crystals were tested for diffraction at diamond using the microfocus beamline but only diffracted weakly to around 20 Å (Data not shown).

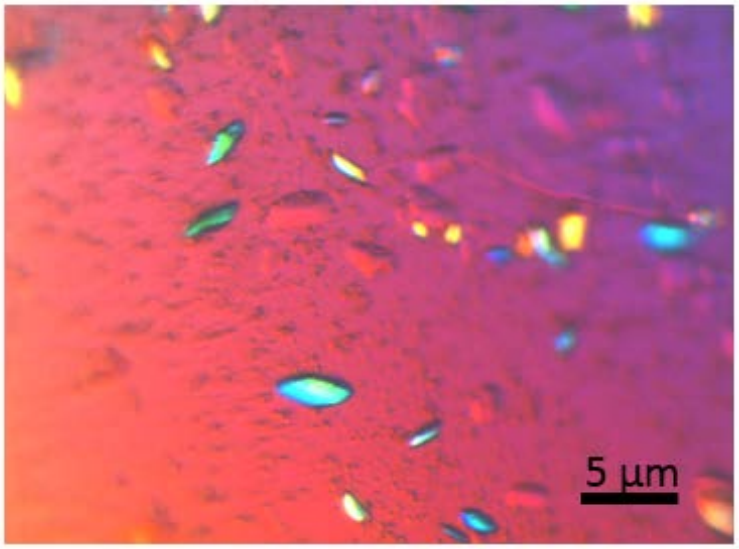


Figure 3.43: Microcrystals of FapF PAO1 viewed under polarised light. FapF formed crystals in 1 week at 20˚C, with the drop conditions, (including the protein buffer solution) 0.02 M TrisHCl pH 7.5, 22 % (*w/v*) PEG 550 MME, 0.1 % (*v/v*) LDAO, 5 mM TrisHCl pH 8.0, 50 mM NaCl. The crystals were very small, most around 1 µm in size.

3.4.5 Optimisation of FapF Constructs

To improve the properties of FapF for structural studies limited proteolysis was used to identify a stable fragment of the protein, chymotrypsin, trypsin and subtilisin were used and the results indicated that stable fragments of FapF could be produced of around 35 and 31 kDa (**Figure 3.44**). Bioinformatics predictions indicated that the N-terminus has a disordered N-terminal domain of more than 100 amino acids (**Figure 3.7**) and the NMR data also showed evidence of a disordered region, perhaps the plug (**Figure 3.42**). Based on these data new FapF constructs were designed with N-terminal truncations to remove the disordered region and produce a more structured protein for study. Two of these constructs were designed based on the predicted chymotrypsin and trypsin cleavage sites (**Figure 3.44C**) corresponding to the stable fragments produced by limited proteolysis and the other was based on bioinformatics predictions of the β-strands (**3.1.6**). Another approach to find an improved construct that we used was to clone a series of homologs for study, for these purposes we made constructs for the *Pseudomonas* PA7, *Pseudomonas* UK4, *Burkholderia* and a version of the gene identified in the Thermophilic Rice Compost Metagenome (TRCM). An additional construct was also made for FapF retaining the native signal sequence and containing a C-terminal hexahistidine to allow membrane extraction of FapF.

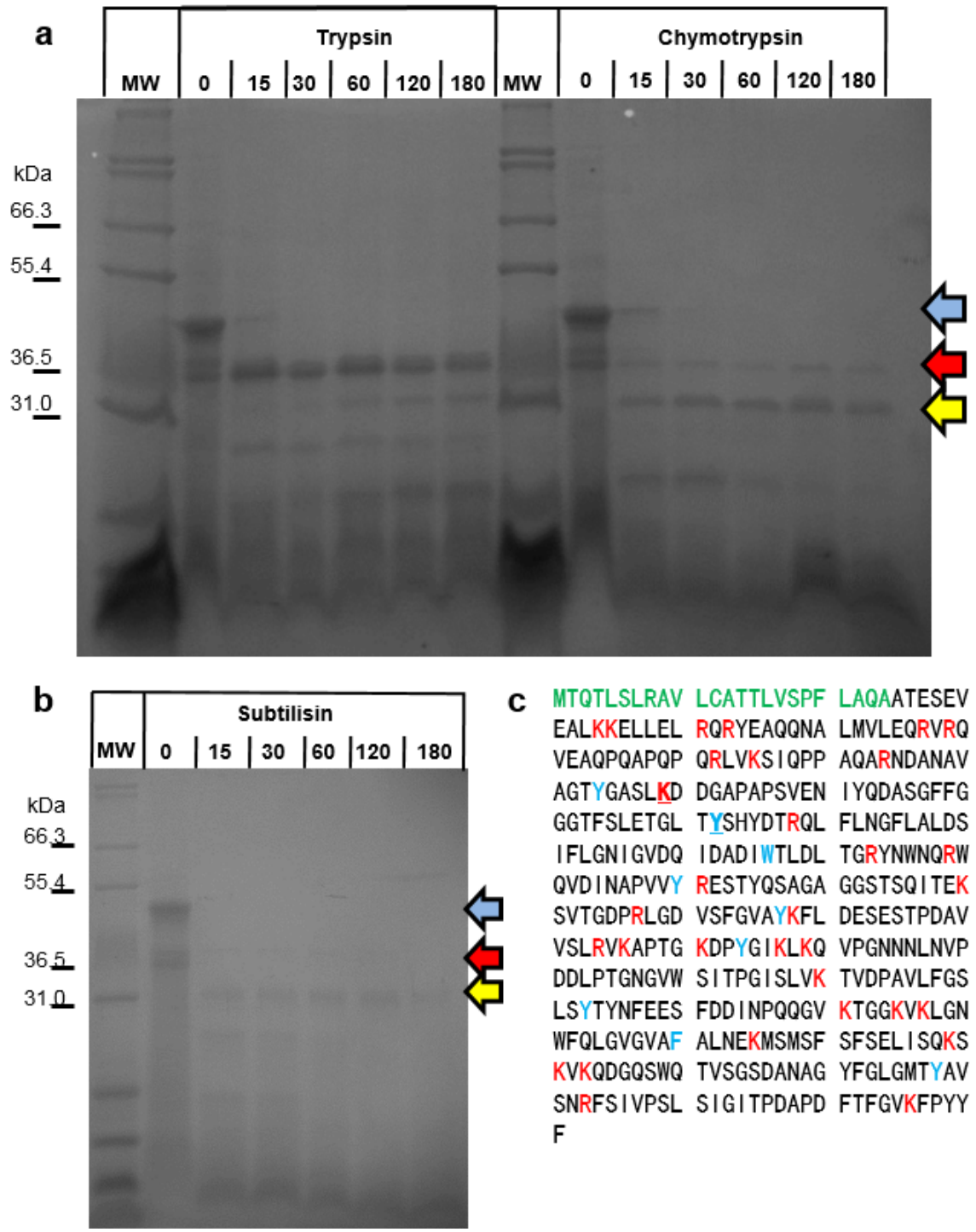


Figure 3.44: Limited Proteolysis of FapF. a, SDS-PAGE gel showing samples of FapF digested over a range of time periods with Trypsin and Chymotrypsin with samples taken periodically over the course of 180 minutes as indicated. **b**, SDS-PAGE gel showing samples of FapF digested over a range of time periods in minutes with Subtilisin. The protein is processed to produce stable fragments by all three proteases. The original protein molecular weight is indicated with the blue arrow and the protein appears to be digested to produce fragments around 35 kDa (Red Arrow) or 30 kDa (Yellow Arrow). Trypsin largely produces the larger fragment with a smaller quantity of the smaller one slowly being created. Chymotrypsin and Subtilisin both produce the smaller fragment, both are more rapid at digesting FapF than trypsin with the latter digesting the protein down to the smaller fragment within 15 min. **c**, Sequence of FapF (PAO1) with predicted cleavage sites for chymotrypsin and trypsin, (PeptideCutter [**181**]), coloured for trypsin (Red) and chymotrypsin (Blue); the signal peptide not present in the construct is coloured in green; suggested cleavage sites corresponding to the stable fragment produced by each enzyme are underlined and in bold.

3.4.5.1 Truncations of FapF PAO1

3.4.5.1.1 Purification and Refolding of Truncated FapF PAO1

Each of the FapF truncated constructs were expressed and purified using the same method as previously (**3.4.1**). FapF$^{99-421}$ and FapF$^{130-421}$ both expressed and refolded successfully (**Figure 3.45)** however the FapF$^{130-321}$ truncation appeared to be less stable with lower yields and a lower molecular weight contaminant which may have been a significant degradation product. The truncation trimming the N-terminus to the start of the β-barrel did not refold successfully indicating that the truncation may have affected the structure of the barrel (**data not shown**).

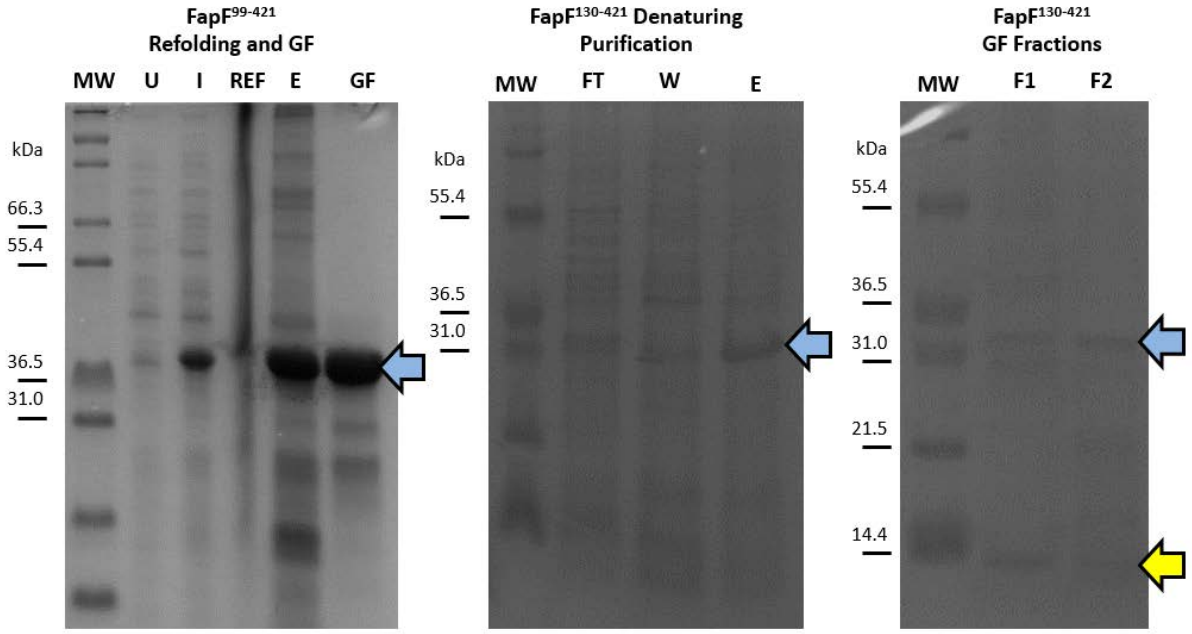


Figure 3.45: Purification and Refolding of FapF Truncations. SDS-PAGE gels of FapF purifications and gel filtered samples. Showing Molecular Weight Ladder (MW), Uninduced Cells (U), Induced Cells after overnight expression (I), Refolded protein (blurred due to high concentration of detergent), Elution of FapF from Ni-resin (E), Gel filtered FapF (GF) and the individual gel filtration fractions from the FapF$^{130-421}$ (F1 and F2). The protein of interest on each gel is indicated with a blue arrow, while the lower molecular weight contaminant is indicated with the yellow arrow.

3.4.5.1.2 Crystallisation trials of FapF truncations

3.4.5.1.2.1 Micelles

FapF$^{99-421}$ samples were expressed, refolded and purified as previously before transferring into a low salt buffer for crystallisation (50 mM NaCl, 10 mM TrisHCl, 0.1 % (*v/v*) LDAO, pH 8.0) by gel filtration. The samples were concentrated to ~10 mg/ml using a 30 kDa MWCO concentrator. Crystal trials were conducted using a series of commercial screens at 10 mg/ml and 5 mg/ml concentration in 200 nL to 200 nL and 200 nL to 100 nL drops (Protein:Preciptant). The trays were incubated at 20˚C and microcrystals could be seen growing after one day, these were allowed to grow for a couple of weeks before images were taken (**Figure 3.46**). Although the protein appears to crystallise more readily and in a broader range of conditions the size of the crystals remains very small. To test whether changing the detergent will improve the crystal size and diffraction crystal trays were also set up using protein samples prepared in βOG (50 mM NaCl, 10 mM TrisHCl, 50 mM βOG, pH 8.0) and Decyl Maltoside (DM) (50 mM NaCl, 10 mM TrisHCl, 3.6 mM DM, pH 8.0). Some crystals formed in both DM (**Figure 4.46E**) and βOG (**Figure 3.46F**) but they were similar in size to the crystals formed in LDAO detergent.

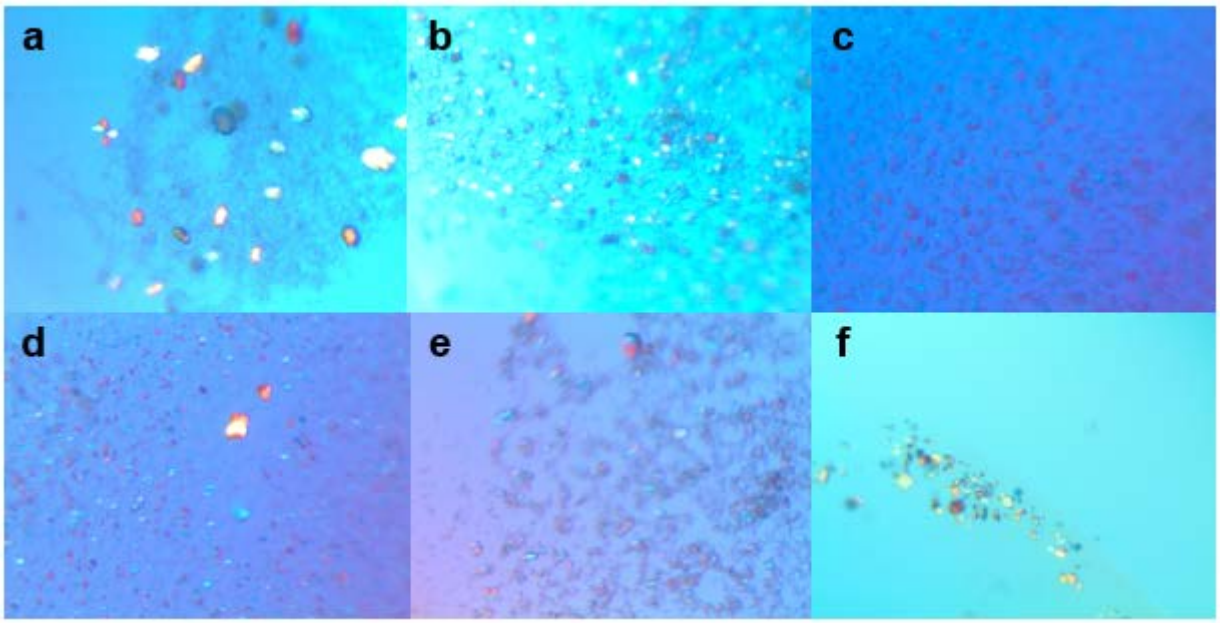


Figure 3.46: Microcrystals of FapF$^{99-421}$ viewed under polarised light. a, 0.1 M NaCl, 0.1 M MES pH 6.5 30 % (*v/v*) PEG 400, **b**, 0.1 M MgCl2, 0.1 M Na Acetate pH 4.6, 30 % (*v/v*) PEG 400, **c**,0.05 M NaCl, 0.02 M MgCl2, 0.1 M Sodium Citrate pH 6, 22 % v/v PEG 400, **d**, 0.2 M Ammonium Sulphate, 0,1 M Na Acetate pH 4.6, 12 % (*w/v*) PEG 4000, **e**, 0.05 M NaSO4 0.05 M LiSO4, 0.05 M Tris PH8.5, 35 % PEG 400, **f**, I0.2 M MgCl2, 0.1 M Tris pH 8.5. 25 % w/v PEG 4000.

3.4.5.1.2.2 Bicelles and Lipidic Cubic Phase (LCP)

When crystallising membrane proteins it is important to consider the environment that the protein is embedded in, so far we have been using a micelle of detergent but other options such as the use of Bicelles or Lipidic Cubic Phase (LCP) are available and we investigated both.

Bicelles can be considered as lipid bilayer disks formed from the combining of an amphiphile (e.g. CHAPSO) and a long chain detergent (e.g. DMPC) under aqueous conditions, the ratio of the components determining the dimensions of the discs [**182**]. Bicelles have successfully been used in membrane protein purification and the change in protein contacts can alter crystallisation [**183, 184**]. Purified $FapF^{99-421}$ was combined with DMPC and CHAPSO (2.8:1) bicelles to produce a solution of 8 % bicelles 0.05 % (*w/v*) LDAO as described in methods (**2.5.2.1.4**). This sample was used in crystal trials were conducted using a series of commercial screens at 10 mg/ml, crystals formed readily in a wide range of conditions, however control plates of the buffer alone were found to form crystals in exactly the same conditions, suggesting that the bicelles themselves were crystallising (data not shown). This is supported by the absence of diffraction from the crystals (data not shown).

LCP is another option for embedding the membrane protein of interest for crystallisation. The cubic phase consists of a single lipid bilayer, usually composed of the lipid monoolein, organized into a three dimensional lattice. Protein embedded in LCP is free to diffuse through the bilayer where it can form crystals. LCP has been used to solve the structure of several membrane proteins including G-protein coupled receptors [**185**]. Purified $FapF^{99-421}$ was used in LCP crystal trials as described in the materials and methods (**2.5.2.1.3**), but no crystals were observed to form in these screens (data not shown).

3.4.5.2 FapF Homologs

Homologous versions of the *fap*F gene were identified in *Burkholderia* and TRCM and ordered from Invitrogen with codon optimisation for *E. coli*. The genes were cloned into pRSF-1B vector and transformed into BL21, the proteins expressed successfully showing a similar pattern of insoluble expression to FapF PAO1 (data not shown). The proteins were then purified and refolded using the same method as FapF PAO1 pNICNTH with similar results, producing a sample pure enough for crystal trials (**Figure 3.47**).

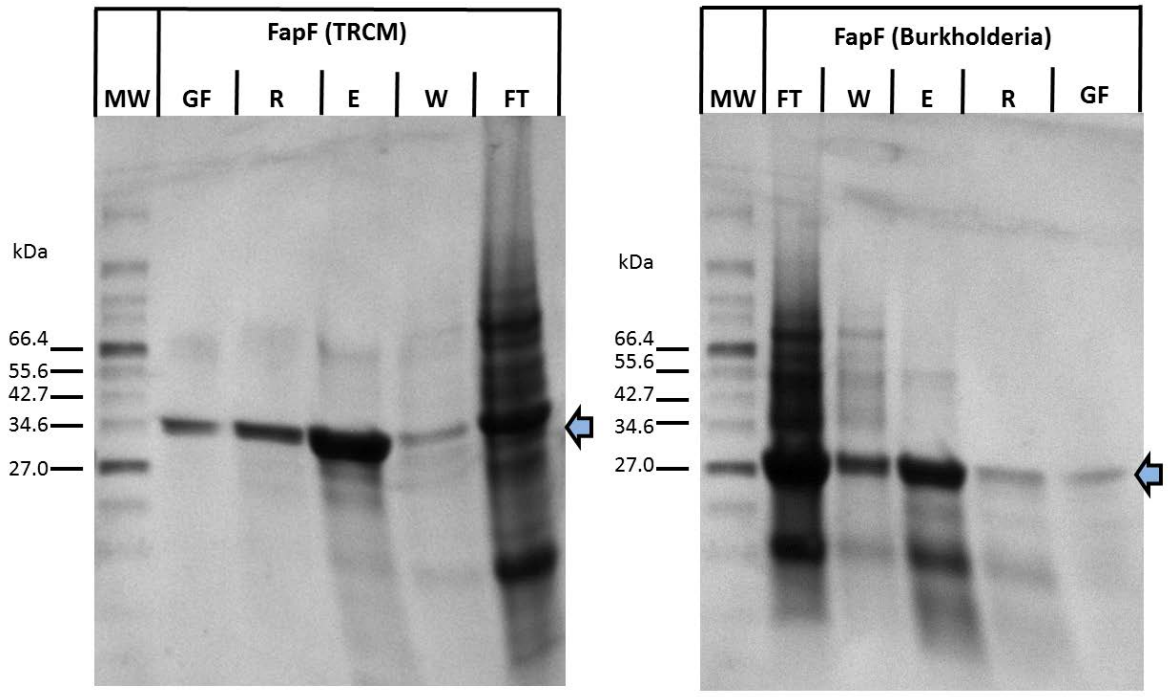


Figure 3.47: Purification and Refolding of FapF Homologs. Homologs of FapF from *Burkholderia* and the Thermophilic Rice Compost Metagenome (TRCM) were expressed, purified and refolded. SDS-PAGE gels of FapF purifications and gel filtered samples. Showing Molecular Weight Ladder (MW), The flow-through from the Ni-Column (FT), the wash from the Ni-column (W), the elution from the Ni-Column (E), the elution from the Ni-column after refolding (R) and the Gel filtered sample of FapF (GF). The protein of interest on each gel is indicated with a blue arrow. Both proteins express well and emerge relatively pure after the denaturing purification and is further purified following the gel filtration step.

3.4.5.3 Periplasmic FapF UK4

To produce a higher quality sample of FapF for crystallisation and in order to conduct *in vivo* experiments involving the full operon additional constructs were designed based on the truncations of the FapF PAO1 construct but for the FapF UK4 homolog with the N-terminal signal sequence replaced with the leader sequence from the *E. coli* membrane protein OmpA followed by a hexahistidine tag after the cleavage site. In principle this should produce a membrane embedded FapF with an N-terminal his-tag for purification. The decision to switch to the OmpA signal sequence was based on the fact that it is known to allow for high efficiency translocation into the periplasmic space in *E. coli* which may not be the case for the native sequence. The tag was placed on the N-terminus due to the poor yield of FapF PAO1 CTH which may have resulted from the tag obstructing the potential C-terminal consensus sequence for membrane integration [**186, 187**]. The plasmid the construct was based on (pRSF-1b) was chosen based on its OriR and resistance markers being compatible with those of the *fap*ABCDEF operon plasmid. The base plasmid with the OmpA signal sequence and N-terminal hexa-histidine tag was also used to produce periplasm targeted versions of each of the Alf genes (**3.2.6.1**). The periplasmic constructs were given to Dr Sarah Rouse and Dr Sebastian Lambert for further study. FapF was expressed and purified successfully from the membrane under native conditions (**Figure 3.48**).

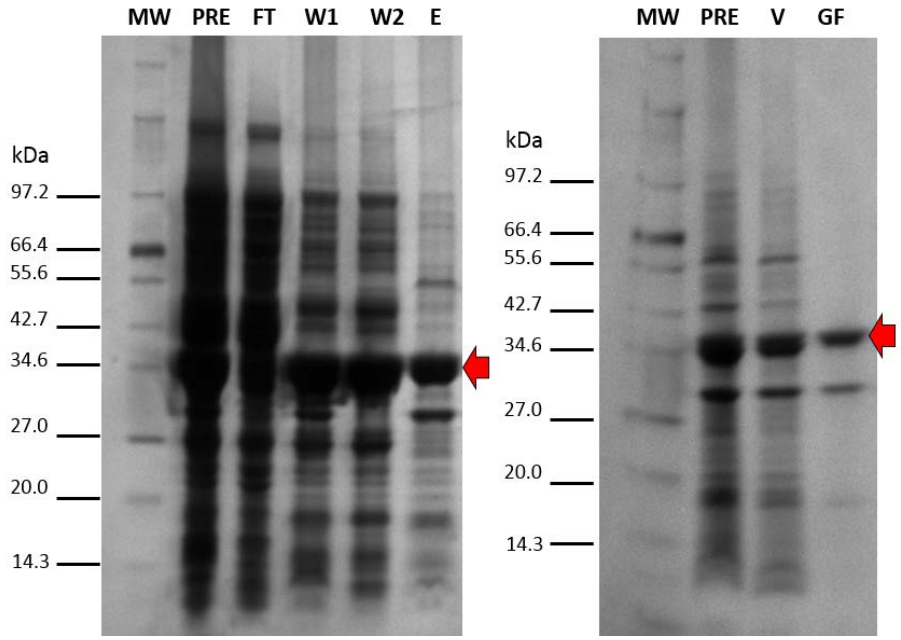


Figure 3.48: Purification of FapF UK4 from the membrane. SDS-PAGE gels of FapF purification and gel filtration. Showing Molecular Weight Ladder (MW), the flow-through from the Ni-Column (FT), the wash from the Ni-column (W1), the second wash (W2), the elution from the Ni-Column (E), samples before purification (PRE), the void fraction (V) and after gel filtration (GF). The band corresponding to FapF is indicated with a red arrow, the protein is purified to a high degree but there are some lower molecular weight contaminants which could be degradation products. These gels were provided by Dr Sarah Rouse who purified these samples.

3.4.6 Crystallisation of FapF UK4

FapF UK4 was purified from the membrane using a mixture of LDAO and C8E4 as the final detergent for crystallisation. The use of C8E4 as an additive was based on the solution of the structure of CsgG using a C8E4-LDAO detergent mixture and the discovery that it formed a β-barrel, as predicted for FapF [**105**]. Crystal trials were conducted by both Dr Sebastian Lambert and Dr Sarah Rouse using a series of commercial screens and successfully produced crystals in a range of conditions (**Figure 3.49**). The addition of C8E4 does appear to improve crystal quality as the same screens using LDAO alone did not produce as large crystals (**Figure 3.49**).

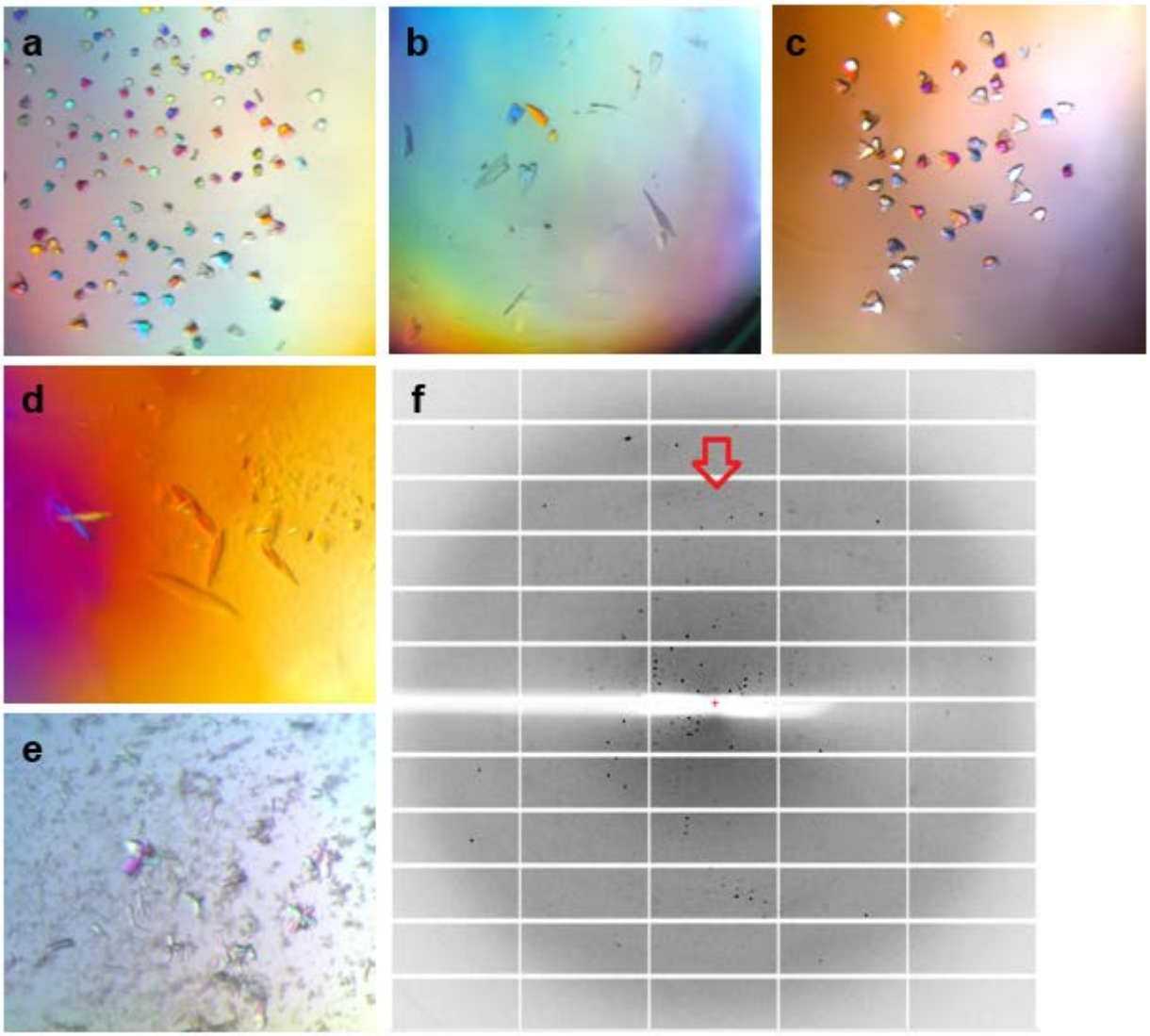


Figure 3.49: Crystallisation and Diffraction of Membrane Extracted FapF. A-E) Crystals of FapF UK4 viewed under polarised light. **a**, 0.1 M NaCl, 0.1 M Sodium citrate pH 5.5, 30 % (*v/v*) PEG 400; **b**, 0.1 M NaCl 0.1M Sodium citrate pH 5.5, 0.1 M Li_2SO_4, 30 % (*v/v*) PEG 400; **c**, 0.1 M NaCl, 0.1 M sodium citrate pH 5.6, 30 % PEG 400; **d**, 0.1 M NaCl, 0.1 M sodium citrate pH 5.5, 0.1 M Li_2SO_4, 30 % (*v/v*) PEG 400; **e**, 1.6 M $(NH_4)_2SO_4$, 0.1 M MES pH 6.5, 10 % (*v/v*) 1,4-dioxane; **f**, Example of FapF diffraction in this case extending to around 4 Å as indicated with the red arrow. Images courtesy of Dr Sebastian Lambert who purified and crystallised these samples.

3.4.7 Summary

FapF (*Pseudomonas* PAO1) was successfully expressed insolubly and refolded by pulse refolding. The refolded protein could be crystallised but the resulting microcrystals of refolded protein were not of sufficient quality for structure solution. The construct of FapF was tested by NMR and limited proteolysis in an attempt to optimise the construct. A truncation of the N-terminus, which evidence suggested was unfolded, produced some improvement in the crystallisation process but the resulting diffraction was still insufficient. Attempts were made to crystallise the protein using either LCP or bicelles but neither approach yielded protein crystals. Constructs based on FapF homologs from other species were designed, expressed and were shown to be refolded using the same method as the PAO1 version of the protein; these sample may prove useful for future work however work on producing a membrane expressed version of the *Pseudomonas* UK4 version of FapF was successful. This construct was used to produce promising crystals with diffraction sufficient for structure solution, a promising step towards structure solution. There were no sufficiently close structures in the PDB to phase the data by molecular replacement and so new crystals will need to be produced and further data collected to solve the phase problem by direct methods.

3.5 CsgH Structure

3.5.1 Selection of CsgH construct

3.5.1.1 Purification of CsgH$^{10\text{-}106}$NTH and CsgH$^{10\text{-}106}$CTH Constructs

Two CsgH constructs were provided by Sebastian Lambert, CsgH$^{10\text{-}106}$CTH (CsgH-CTH) and CsgH$^{10\text{-}106}$NTH (CsgH-NTH) for comparison with previous data collected by Dr Jonathan Taylor. Both proteins were purified under native conditions by Ni-affinity chromatography producing relatively pure samples (**Figure 3.50**) which were further purified by gel filtration (**Figure 3.51**).

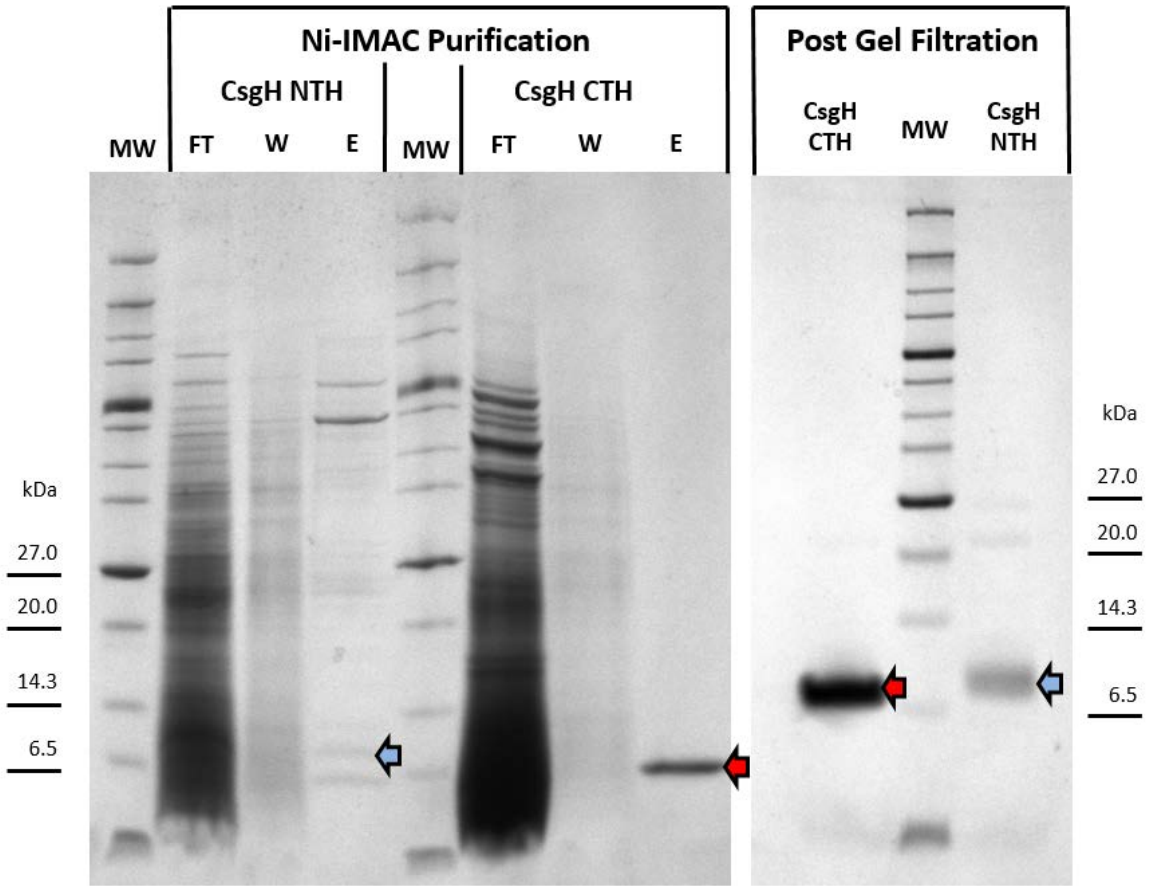


Figure 3.50: Purification of CsgH-NTH and CsgH-CTH constructs. SDS-PAGE gels showing samples from Ni-IMAC purification of the CsgH-NTH and CsgH-CTH constructs. The bands corresponding to CsgH-NTH are indicated with blue arrows and CsgH-CTH with red arrows. The yield and purity of CsgH-CTH was significantly better than the CsgH-NTH construct.

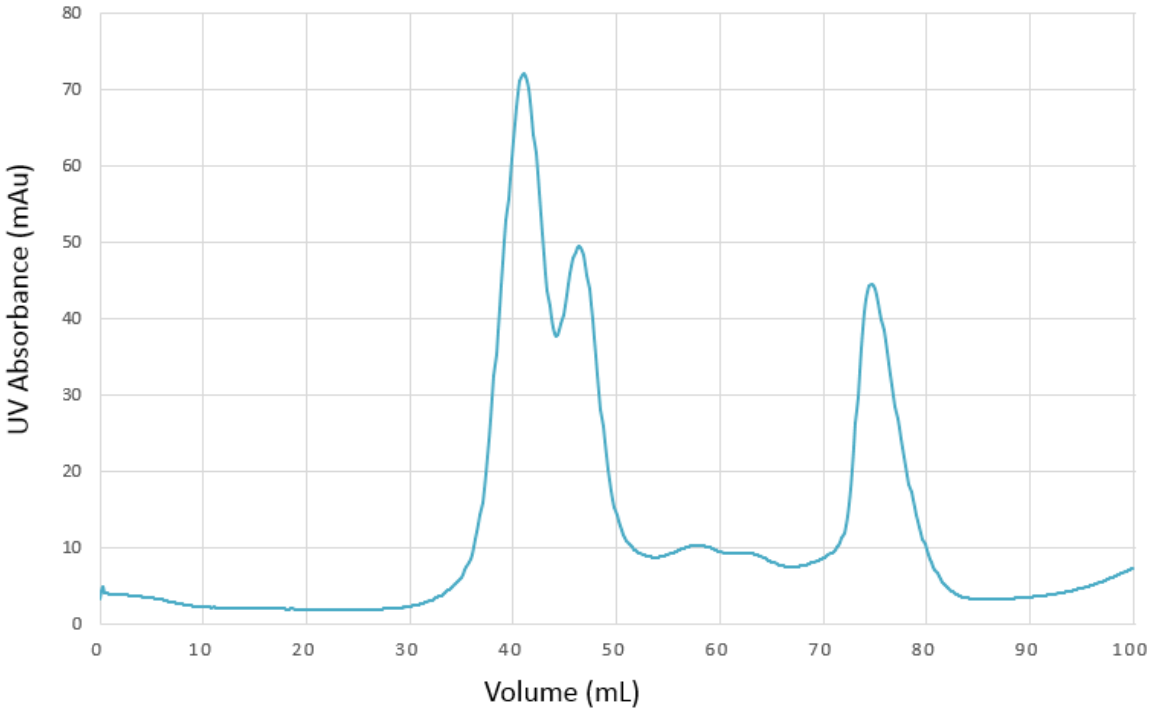


Figure 3.51: Gel Filtration of CsgH. Example Gel Filtration trace of CsgH in 100 mM NaCl, 20 mM HEPES pH 7.5, purified by gel filtration on an S75-superdex column. The peak centred at around 77 mL contains our protein of interest and the corresponding fractions were pooled.

3.5.1.2 Comparison of spectral quality

^{15}N labelled samples were purified and 2D ^{15}N HSQC spectra collected for both CsgH constructs in 100 mM NaCl, 20 mM HEPES pH 7.5 and compared to previous data collected by Dr Jonathan Taylor (**Figure 3.52A**). Notably the truncated constructs both produced better spectra than the longer construct previously studied, despite having fewer amino acids they both possess more peaks. Out of the three constructs arguably CsgH-CTH produced the best spectra, with the peaks in the spectra well dispersed, well defined and close in number to what would be expected for the protein. Although the spectra of the N-terminally tagged construct was similar in quality to the C-terminally tagged one, the yield of the protein is worse, a potential problem for structural solution where higher protein concentration is more favourable. The thermal stability of CsgH-CTH was then tested by 1D NMR and showed the protein was stable at 303˚K for 5 days with no loss of signal or change of peak frequency (**Figure 3.52B**). Due to the high quality of the ^{1}H ^{15}N HSQC spectra and the good thermal stability of CsgH-CTH the protein was selected for structure determination.

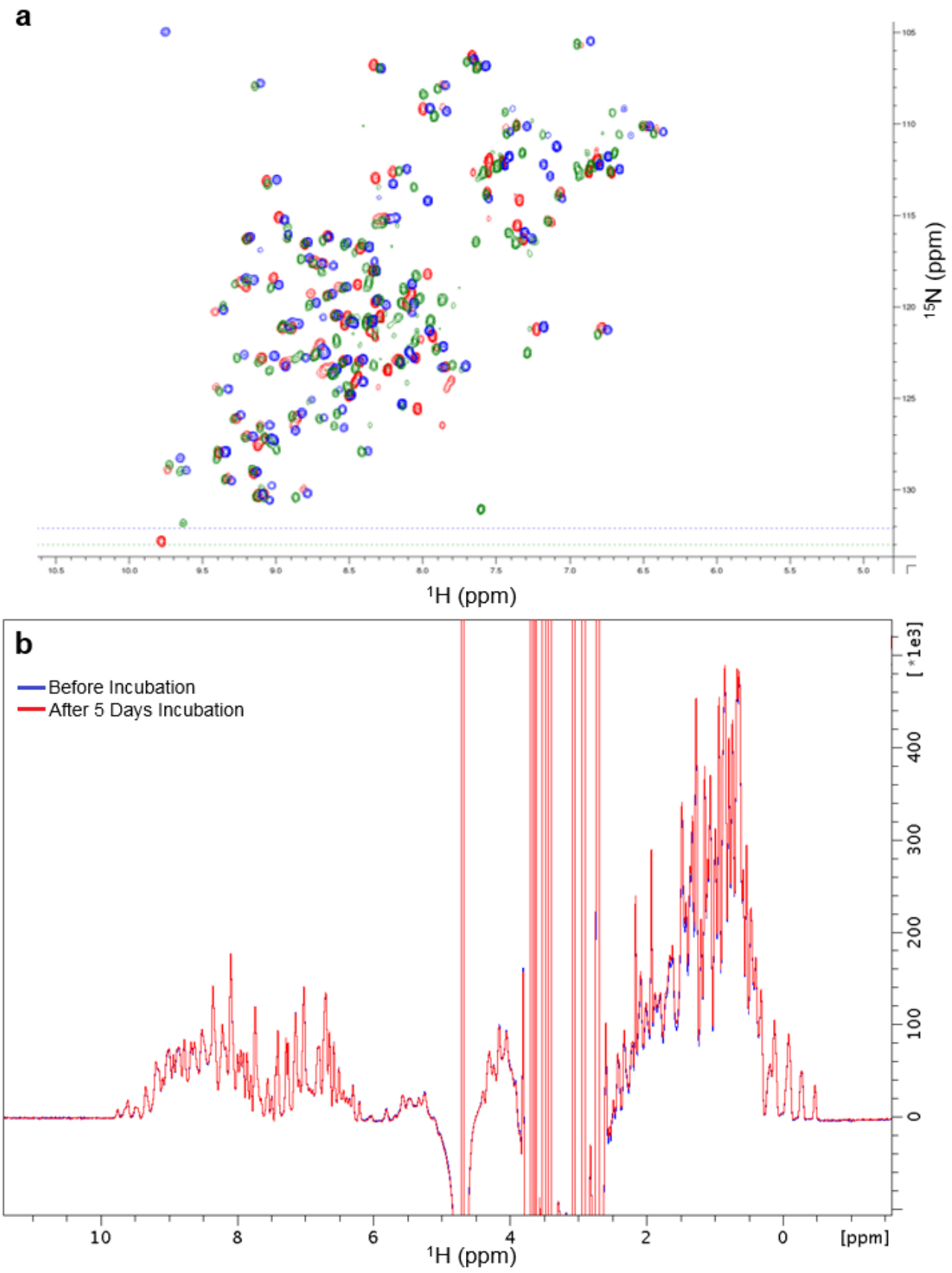


Figure 3.52: Selection of construct for structural study by NMR. a, Overlay of ^{1}H ^{15}N HSQC spectra of CsgH$^{10\text{-}106}$-CTH (Blue), CsgH$^{10\text{-}106}$NTH (Green) and data previously collected by Dr Jonathan Taylor for CsgH$^{1\text{-}106}$CTH (Red) showing the dispersion, number of peaks and peak sharpness of the spectra for comparison. **b**, Comparison of 1D spectra of CsgH-CTH before (Blue) and after (Red) 5 days at 303 ℃, showing the spectra to be essentially identical. The NMR buffer used for the experiments was 100 mM NaCl, 20 mM HEPES pH 7.5.

3.5.2 Backbone Assignment CsgH-CTH

A double labelled ^{15}N ^{13}C sample was expressed and purified by the same methodology as the ^{15}N labelled samples. The purified sample (400 μM) was used to record 3D HNCACB, CBCA, HNCO and HN(CA)CO spectra at 292℃ required for backbone assignment. The spectra were processed using NMRPipe (**156**) and rendered using CcpNmr analysis 2.4.0 (**153**). The chemical shifts of the HN and N atoms of the backbone are provided by a ^{15}N ^{1}H HSQC experiment, 113 peaks were initially selected and inspection later identified 10 peaks overlapping these for a total of 123 peaks. The HNCACB experiment provides chemical shift values of Cα and Cβ atoms associated with each peak in the HSQC (where present), importantly it also includes the Cα and Cβ atoms of the preceding residue. Separating the carbon peaks of the current residue from those of the preceding residue is aided by the CBCA experiment which only shows the Cβ and Cα atoms of the preceding residue. Similarly the HNCO experiment provides the chemical shifts of the backbone carbonyl carbon atom associated with a particular HN and N in the backbone as well as the atom from the preceding residue. Together, the chemical shifts and sequential information provided by these experiments, combined with our knowledge of the protein sequence and the probable chemical shifts of a given residue, (From BMRB database **[144]**), it becomes possible to assign the chemical shifts to specific residues. The chemical shifts were selected using CcpNmr analysis 2.4.0 [**188**]. The program MARS [**145**] was used to automatically assign the chemical shifts to specific residues based on the pattern of Cα and Cβ chemical shifts, the data on the preceding residue's chemical shifts and the protein sequence. MARS was run with 0.25 ppm tolerance for C' and 0.5 ppm tolerance for Cα and Cβ residues. The resulting assignments could be displayed and inspected in CcpNmr (**Figure 3.53**). The data was generally of good quality although there were some heavily overlapped amide peaks and some peaks were absent from the 3D experiments, particularly in the HNCACO spectra.

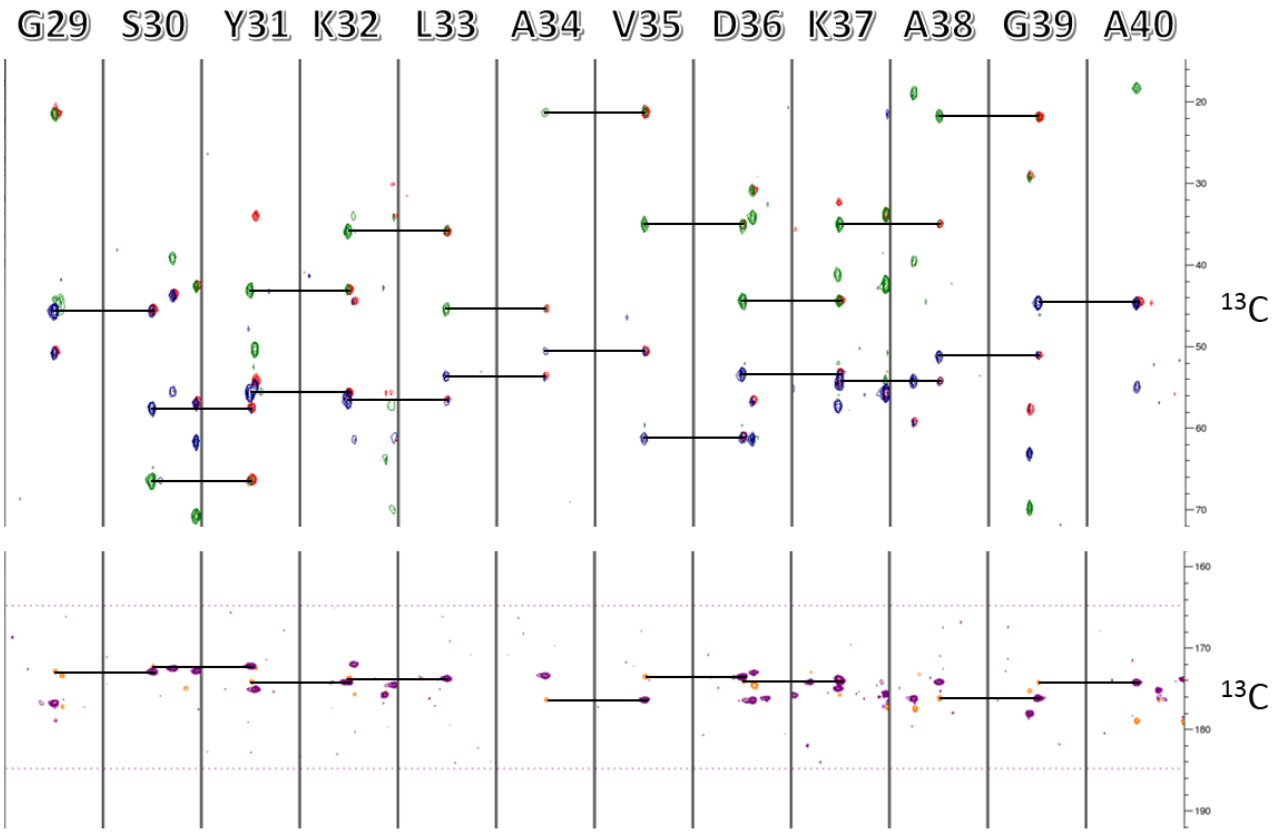


Figure 3.53: Backbone Assignment Strips. Strips from NMR spectra used for backbone assignment showing the data from the HNCACB (Positive Blue, Negative Green), CACB (Red), HNCO (Purple) and HNCACO (Orange). Peaks linking the strips are illustrated with black lines, the data is of good quality but there are still some gaps such as between alanine 34 and leucine 33 in the carbonyl region.

While the bulk of the protein could be assigned quickly (~70 %) some peaks were more difficult to assign, generally due to peak overlap, inaccurate chemical shifts or sequence similarity causing some degree of ambiguity. Through cycles of inspection, addition of new peaks, adjustments to the chemical assignments and automated assignment by MARS the assignments were gradually improved, overlapping peaks distinguished and the backbone assigned as completely as possible. The final backbone assignment (**Figure 3.54**) was ~99 % complete for the protein backbone excluding proline residues, the linker and the poly-histidine tag (95 % for entire protein). MARS uses the chemical shift values for the spin systems, the polypeptide sequence and the secondary structure prediction and attempts to assign the full backbone to satisfy the available data, the consistency with which a particular spin system is assigned to a particular residue is used as an estimate of confidence. The final MARS run assigned 97 residues with a high confidence, 4 with medium confidence (Glu100, His101, His102, His103) and 5 were unassigned (Met1, Pro94, His104, His105, His106). Notably most of these less confidently assigned and missing residues occur in the tag (Residues 99-106), it is likely that overlap makes it impossible to distinguish between the Histidine residues in the tag after His101 as they are likely to have similar chemical environments. Proline will not produce a peak due to its distinct structure and so it is an anticipated gap in the assignment. The N-terminal methionine residue is absent from the spectra because its -NH3+ group is in fast exchange with the solvent and so cannot be observed.

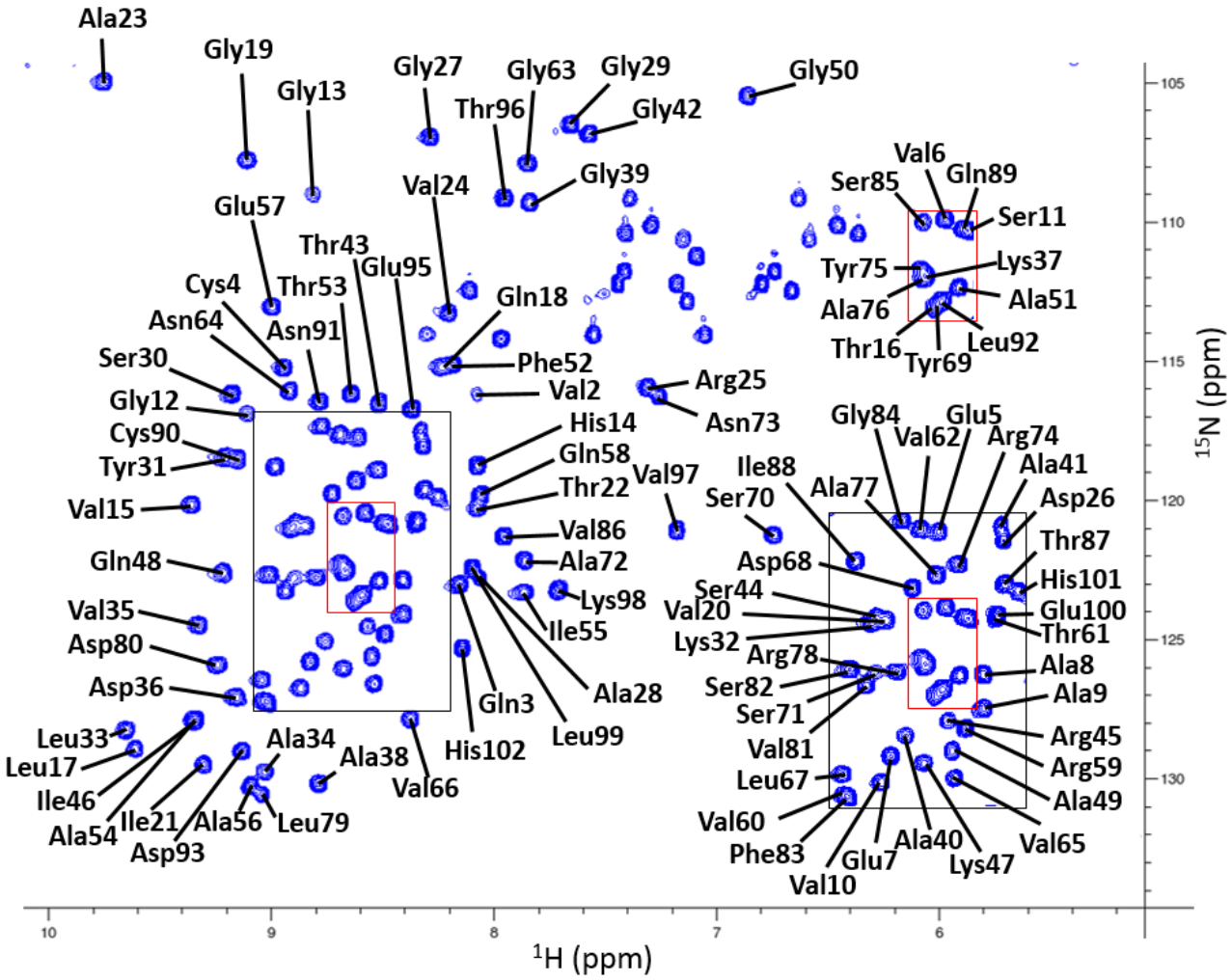


Figure 3.54: Backbone Assignments of ^{15}N ^{1}H HSQC. The 2D HSQC spectra of CsgH used to solve the protein structure is shown here with the assignments made using CcpNmr Analysis and MARS. The central regions of the spectra are shown labelled in the separate boxes on the right for clarity. The spectra is well dispersed and 95 % of the backbone amides has been assigned. The unassigned peaks are likely to be side chains or noise. Note that the ^{1}H ^{15}N peak for Alanine 23 is folded explaining its unusual chemical shifts.

The assigned chemical shift values of the ^{15}N, Hα, Cα, Cβ and C' of each residue can be used together with the sequence to predict the Phi and Psi angles together with the secondary structure of the protein. The secondary structure predicted from the assigned chemical shifts using DANGLE (**155**) compared favourably to the predicted structure of the protein (PSIPRED [**126**]) with strands positioned almost exactly as predicted based on the protein sequence (**Figure 3.55**). The consistency of the secondary structure prediction supports the accuracy of the assignments and the dihedral angles predicted were useful for the structural calculations, although the final structure used dihedral angles predicted by Talos+ [**189**].

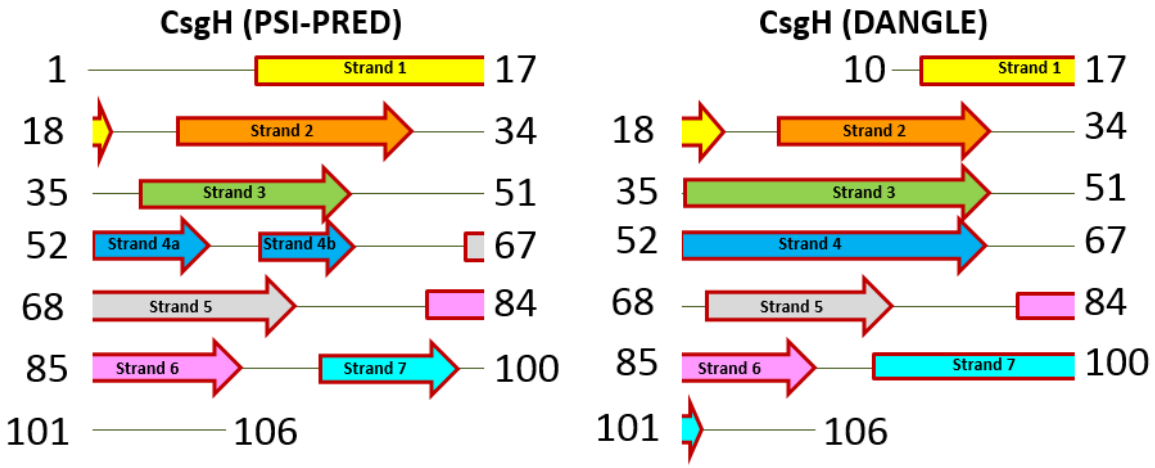


Figure 3.55: Comparison of secondary structure predictions. Cartoon figure comparing the secondary structure of CsgH based on sequence only (PSIPRED [**126**]) and the secondary structure predicted from the measured chemical shift data (DANGLE). The structures appear to be very similar with strands in approximately the same position. Note that the construct used for NMR lacks the presumably disordered N-terminal region and so this is not included in the prediction by DANGLE. Although the C-terminus of CsgH according to DANGLE has been shown as disordered but the prediction is more indeterminate with some evidence of both helical and coiled conformations.

3.5.3 Sidechain Assignment CsgH-CTH

To assign the sidechain atoms of the protein it was necessary to conduct another series of 3D NMR experiments using the same protein sample as before: HBHA(CO)NH, CC(CO)NH and HCCH-TOCSY. The CC(CO)NH experiment in provides the chemical shifts of all the carbon atoms, bound to a hydrogen in the preceding sidechain (**Figure 3.56**). The quality of the data from the CC(CO)NH experiment was generally good and provided the chemical shifts of most of the Cγ and Cδ carbon atoms.

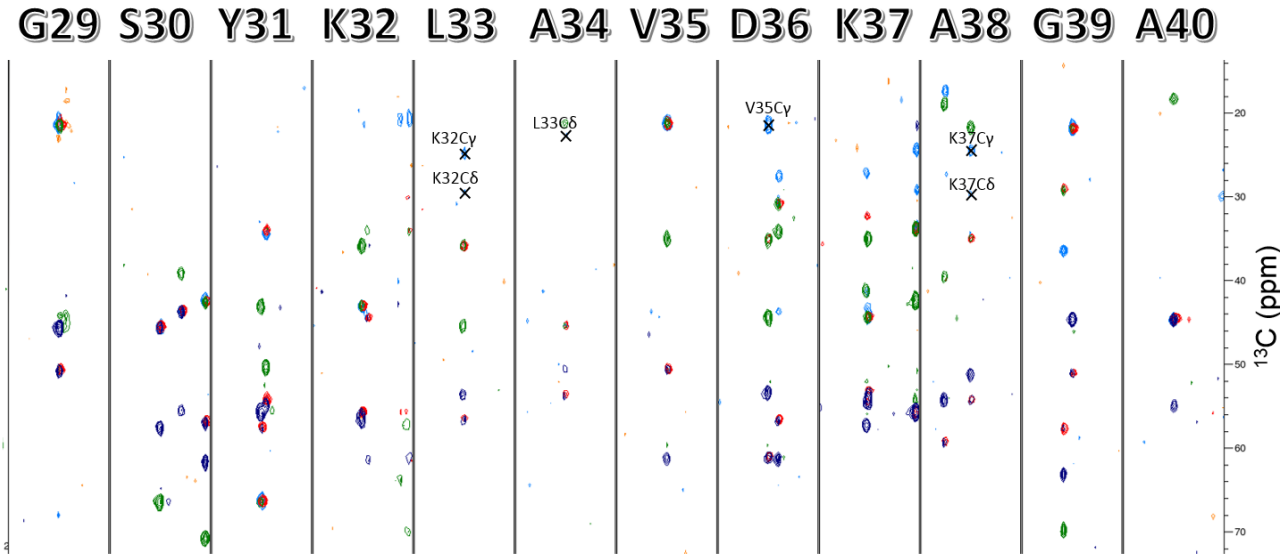


Figure 3.56: Strips for Sidechain Carbon Atom Assignments. Strips from NMR spectra used while working on sidechain assignment showing the data from the HNCACB (Positive Blue, Negative Green), CC(CO)NH (Light Blue). Sidechain peaks assigned using this spectra have been indicated with a cross and a description of the atom assigned. Note the peaks correspond to carbon atoms on the preceding residue.

The HBHA(CO)NH spectrum provides in turn the chemical shift values of the Hβ and Hα atoms of the preceding residue (the specific residue is known from the backbone assignment) (**Figure 3.57A**. The information from these experiments is useful but not essential for assigning the side chain atoms, the main experiment was the HCCH-TOCSY which provides all the carbon and carbon attached protons in the same sidechain as a particular CH atom. From the carbon and proton atoms already assigned the other side chain atoms can therefore be found in the HCCH-TOCSY (**Figure 3.57B**).

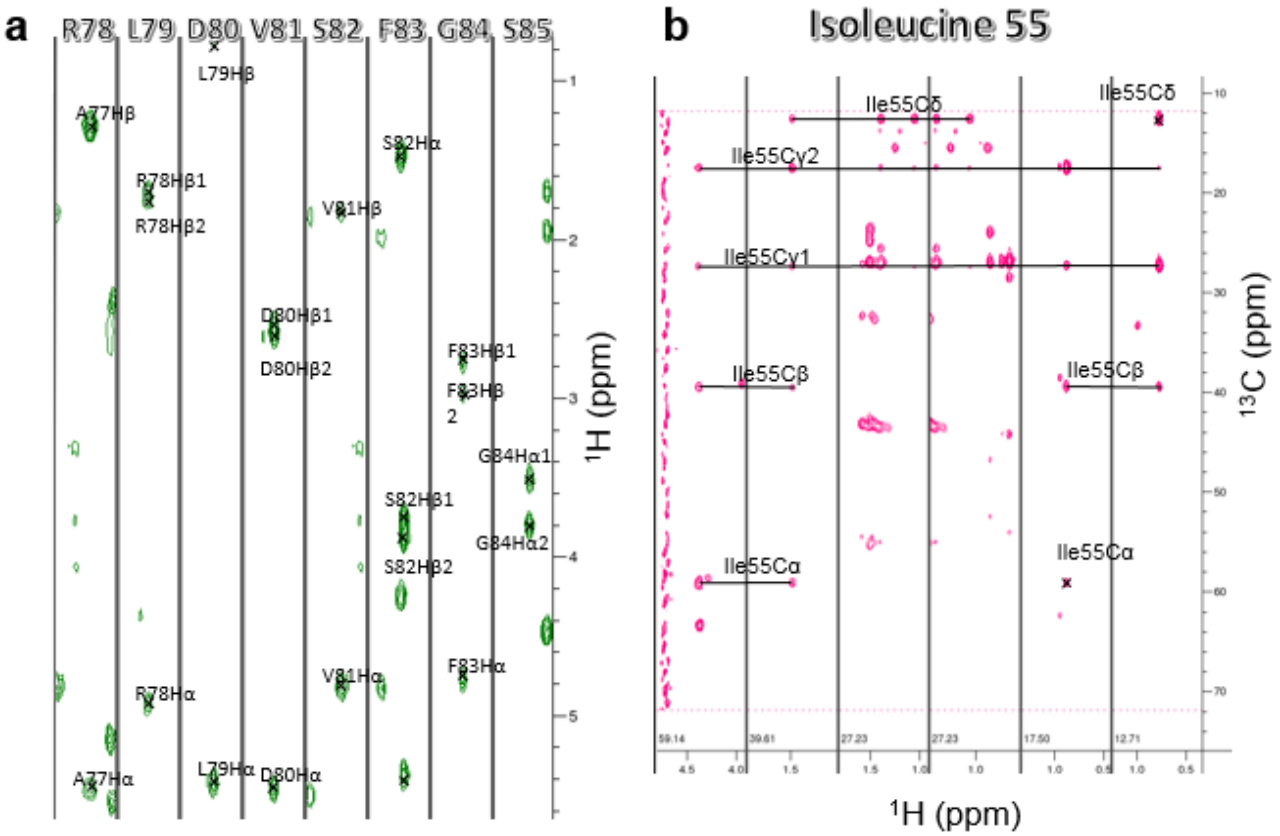


Figure 3.57: Strips for Sidechain Atom Assignments. a, Strips from HBHA(CO)NH spectra, labelled by the relevant ^{15}N ^{1}H resonance, illustrating data used to assign Hα and Hβ atoms for individual residues. The relevant peaks are marked with a cross and the exact atom labelled. **b**, Strips from HCCH-TOCSY spectra for assignment of Ile55 illustrating data used to assign the sidechain atoms of each residue. The relevant peaks are marked with a cross or linked by a line if present in adjacent strips, the carbon atom corresponding to the ^{13}C resonance on the vertical axis is labelled.

Assignment of the atoms in the aromatic residues required some additional experiments a ^{2}H-Heteronuclear multiple quantum coherence (HMQC) and an H-H NOESY which were duly conducted. Note that in order to improve the spectra for the H-H NOESY the protein was transferred into D_2O by dilution and concentration. Together these experiments allow us to cautiously assign the protons and carbon atoms in the aromatic rings from prior assignments, this is aided in this case by the presence of only five aromatic residues in CsgH. CcpNMR analysis was used for the sidechain assignments initially but they were then checked and some corrections made using scripts designed for sidechain assignments by Marchant *et al*., **[146]** in NMRview **[190]**. The majority of the ^{1}H, ^{13}C and ^{15}N atoms in CsgH were successfully assigned; 88.5 % not including the tag, 81.6 % including the tag.

3.5.4 Structure Calculation CsgH-CTH

Structure calculation was conducted using both the assignments as well as some structural constraints. To provide distance constraints a ^{15}N-NOESY-HSQC and a ^{13}C-NOESY-HSQC were conducted the former in H_2O and the latter in D_2O. Further distance constraints were available from the H-H NOESY used during the assignment process. These distance constraints consisted of picked peaks in the NOESY spectra with known chemical shifts and with the distance derived from the peak height. Dihedral angle constraints were also used during the structural calculation, these were derived by TALOS using the chemical shift data from the assignments. Furthermore since Cys4 and Cys90 were expected to form a disulphide bridge and their chemical shift assignments supported this arrangement an additional force constraint was added for a disulphide bridge between the two residues. The structure was calculated using ARIA, the resulting violations in the structures were checked and the NOESY spectra curated by inspection to eliminate spurious or ambiguous peaks, the calculation methods were also adjusted to produce the best results (minimal violations). The final calculation was carried out as described in materials and methods (**2.5.1.2**). The ensemble of the best 20 structures was produced (**Figure 3.58A, Figure 3.58B**) and the lowest energy structure was also selected to best illustrate the structure (**Figure 3.58C**). The statistics for the final ensemble of NMR structures for CsgH are shown in **Table 3.1**.

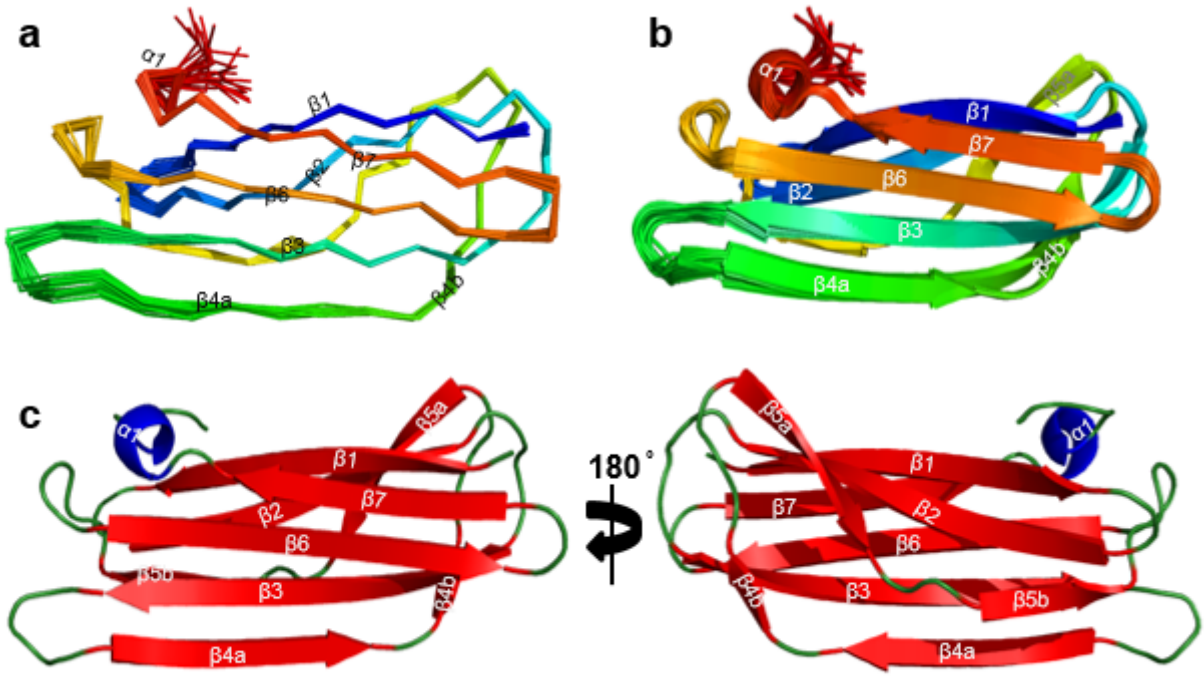


Figure 3.58: Calculated CsgH Structure. a, Ensemble of 20 calculated CsgH structures aligned over the entire structure, showing the protein backbone and coloured from Blue to Red from N-terminus to C-terminus. **b**, Ensemble of 20 calculated CsgH structures aligned over the entire structure, cartoon diagram showing the secondary structure and coloured from Blue to Red from N-terminus to C-terminus. **c**, Lowest Energy Structure of CsgH viewed from opposite sides, cartoon diagram of the secondary structure with sheets coloured in red, loops in green and helices in blue

Table 3.1: NMR and Refinement statistics for final CsgH structure ensemble.

NMR Distance and Dihedral Constraints	
Total Unambiguous NOE Constraints	1202
Intraresidue	575
Interresidue	627
Sequential (|i-j|)=1	200
Medium range (|i-j|)<4	54
Long range (|i-j|)>5	373
Ambiguous Distance Constraints	519
Talos+ Dihedral Angle Restraints	
ψ	85
φ	85
Structural Statistics Violations (Mean (SD))	
Distance Constraints (Å)	0.25 (0.55)
Dihedral Angle Constraints (˚)	0.5 (0.61)
Maximum Dihedral Angle Violation (˚)	4.93
Energies	
Mean Constraint Violation Energy	283 (14.1)
Mean Amber Energy	-2680 (43.5)
Maximum distance constraint violation (Å)	0.4871 (0.257)
Mean Deviations From Idealised Geometry	
Bond Length (Å)	0.00616 (0.00015)
Bond Angle (˚)	0.629 (0.0167)
Average Pairwise RMSD (Å) (Residues 1-98)	
Heavy	1.01 (0.1)
Backbone	0.42 (0.11)
Ramachandran Plot	
% in most favoured regions	94.5 % (1.3 %)
% in allowed regions	99.4 % (0.6 %)
% in disallowed regions	0.6 % (0.6 %)

3.5.5 Unusual Chemical Shifts & Ramachandran Outliers

3.5.5.1 Ramachandran Outliers

The Ramachandran plot was developed in 1963 [**191**], essentially a plot of ψ and φ angles from the backbone of the protein. The possible angles are restrained physically to a narrow range of conformation and study of many protein structures has shown that actual angles cluster to an even narrower range of favoured angles. The Ramachandran plot of the structural ensemble for CsgH is show in **Figure 3.59** and shows that the vast majority of peaks are in the favoured region but there are some outliers, these correspond to the C-terminus and the loop containing Glycine 12, these residues are likely to be artifacts, present in less structured regions with less constraints to define the structure.

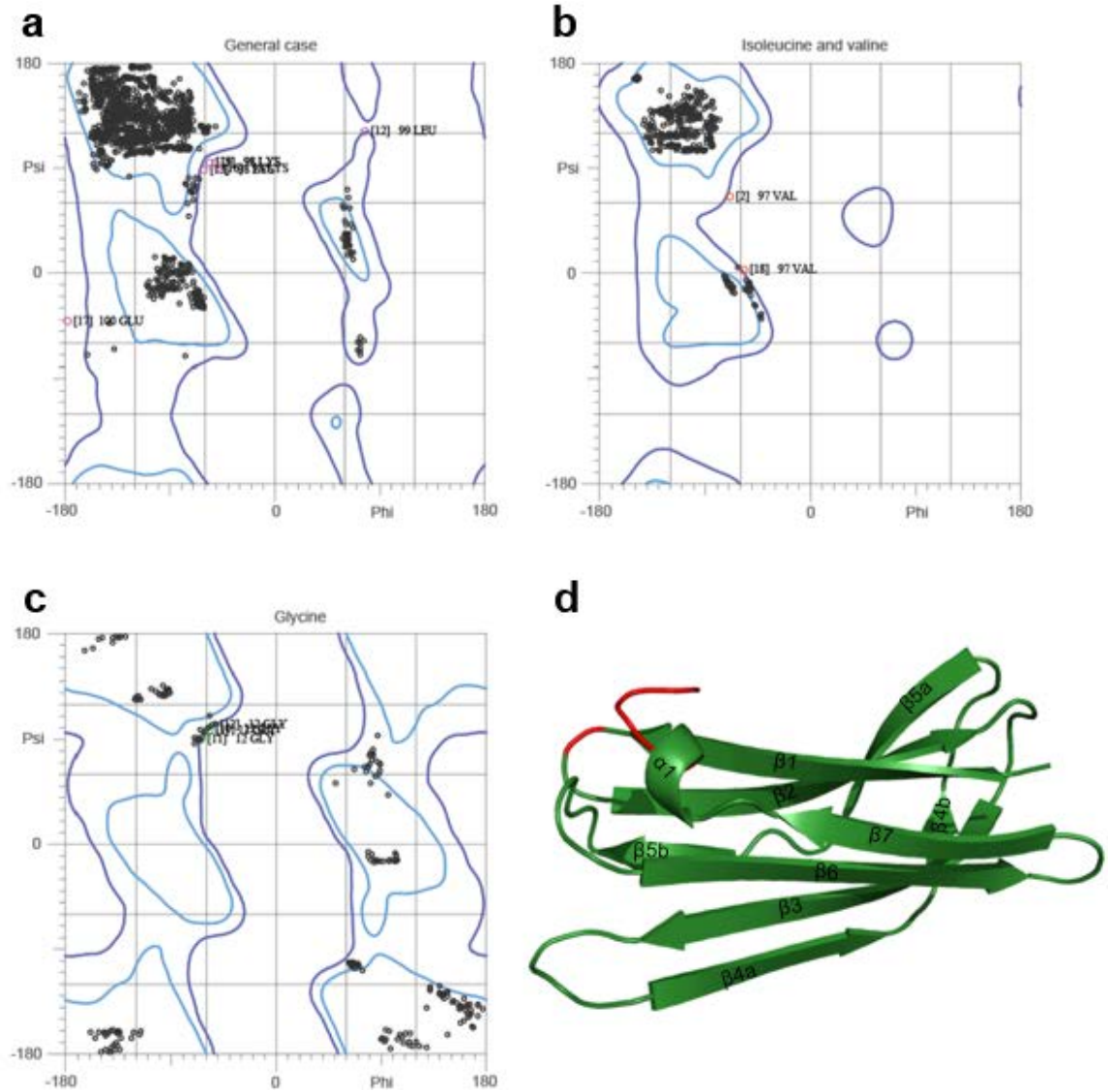


Figure 3.59: Ramachandran outliers. a-c, Ramachandran plots (Generated by Molprobity [**151**]) of selected residues from the Structure Ensemble of CsgH. The plot shows theoretically favoured regions in light blue and allowed regions in dark blue. The outliers are labelled in square brackets with the individual model within the ensemble (1-20) and then the specific residue in the structure. Note that for the sake of clarity the plots for pre-proline and proline residues have been excluded as they include no outliers. **d**, Cartoon Structure of CsgH showing residues which appear as outliers are highlighted in red (G12, V97, K98, L99, E100).

3.5.5.2 Unusual Chemical Shifts

The measured chemical shifts of CsgH and the co-ordinates of the lowest energy structure were analysed using the BMRB tool Coordinates versus Assigned Chemical Shifts Checker [**144**]. This software indicated no issues between the chemical shift file and the co-ordinates but did highlight that Ala23 N (104.959 ppm) and Val62 Hβ (0.339 ppm) possessed chemical shifts more than five standard deviations away from the average chemical shift. The average chemical shift for alanine N is 123.26 ppm with a standard deviation of 3.47 ppm and valine Hβ has an average chemical shift of 1.98 ppm with a standard deviation of 0.32 ppm The chemical shift of the Ala23 nitrogen can be easily explained as the shift is folded back onto the spectrum and so actually has a shift of 132 ppm. Val62 Hβ is unusual but may result from its close proximity to Tyr31 in the structure (**Figure 3.60B**) and may be being ring current shifted. Further analysis using ShiftX [**192**] to predict the chemical shifts from the structure indicated that most chemical shifts were close to those predicted from the structure, however Ser70 HN is in an unusual position on the HSQC at around 6.7 ppm (**Figure 3.54**). Ser70 lies in a loop near two tyrosine residues (**Figure 3.60A**), since the loop is flexible and relatively weakly constrained it is possible that the Serine residue is closer to one of the aromatic rings and a strong ring current shift is causing the unusual chemical shift.

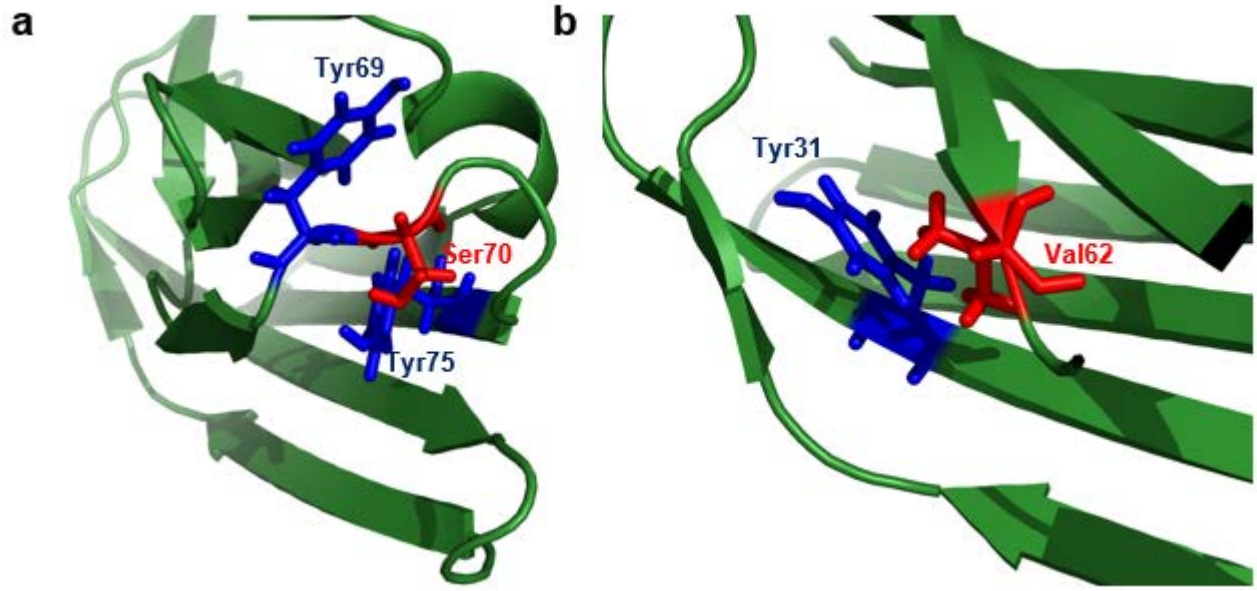


Figure 3.60: Unusual chemical shifts and proximal aromatic residues. a, Cartoon Structure of CsgH in green, with Ser70 shown as sticks in red and Tyr69 and Tyr75 shown as sticks in blue **b**, Cartoon Structure of CsgH green, withVal62 shown as sticks in red and Tyr31 shown as sticks in blue.

3.5.6 The Structure of CsgH

The final structure of CsgH is remarkably similar to that of CsgC as anticipated from the secondary structure prediction (**Figure 1.14**), adopting a compact immunoglobulin-like β-sandwich with the β-strands forming two sheets (**Figure 3.61**). The C-terminus of CsgH contains a short helical region not present in CsgC (**Figure 3.61**) but it is unclear whether this is a significant feature.

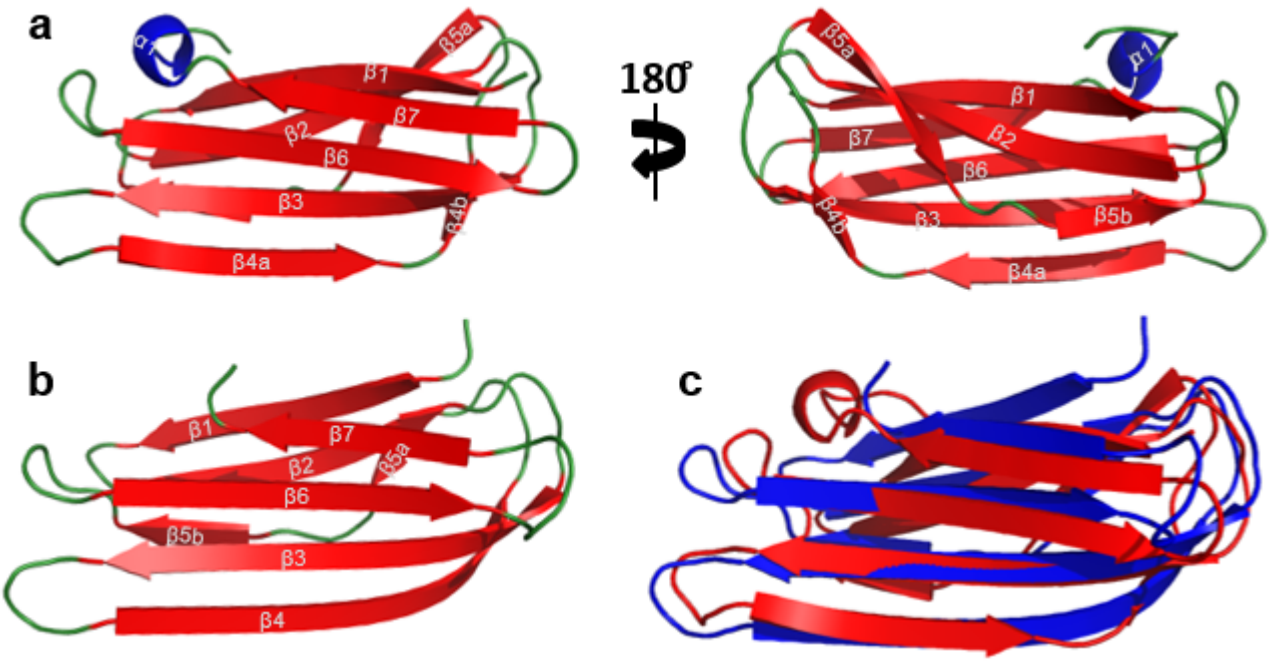


Figure 3.61: Comparison of the structures of CsgC and CsgH. a, Cartoon diagram illustrating the structure of CsgH in two orientations as indicated by the arrow, the secondary structure is highlighted by colour with the strands in red, loops in green and helices in blue. **b**, Cartoon diagram illustrating the Crystal Structure of CsgC 2Y2T as solved by Dr Taylor the secondary structure is highlighted by colour with the strands in red, loops in green and helices in blue showing the similarity in overall structure to CsgH. **c**, Cartoon diagrams of CsgH (Red) and CsgC (Blue) aligned using Pymol (RMSD ~2.4) illustrating the similarity between the two structures.

The topology of CsgH is illustrated in **Figure 3.62A**, showing the arrangement of what are effectively seven β-strands forming two sheets, these sheets are bound together by a disulphide bridge between the most N-terminal and most C-terminal strands. Although the topology of CsgC is the same the cysteines in CsgC do not connect these strands, instead forming a CxC motif which has been suggested to have some sort of functional role [**106**], the cysteines in CsgH are much further apart and are not present at the same location in the structure. The position of the conserved cysteines in CsgH suggest they perform an important structural role (**Figure 3.62**).

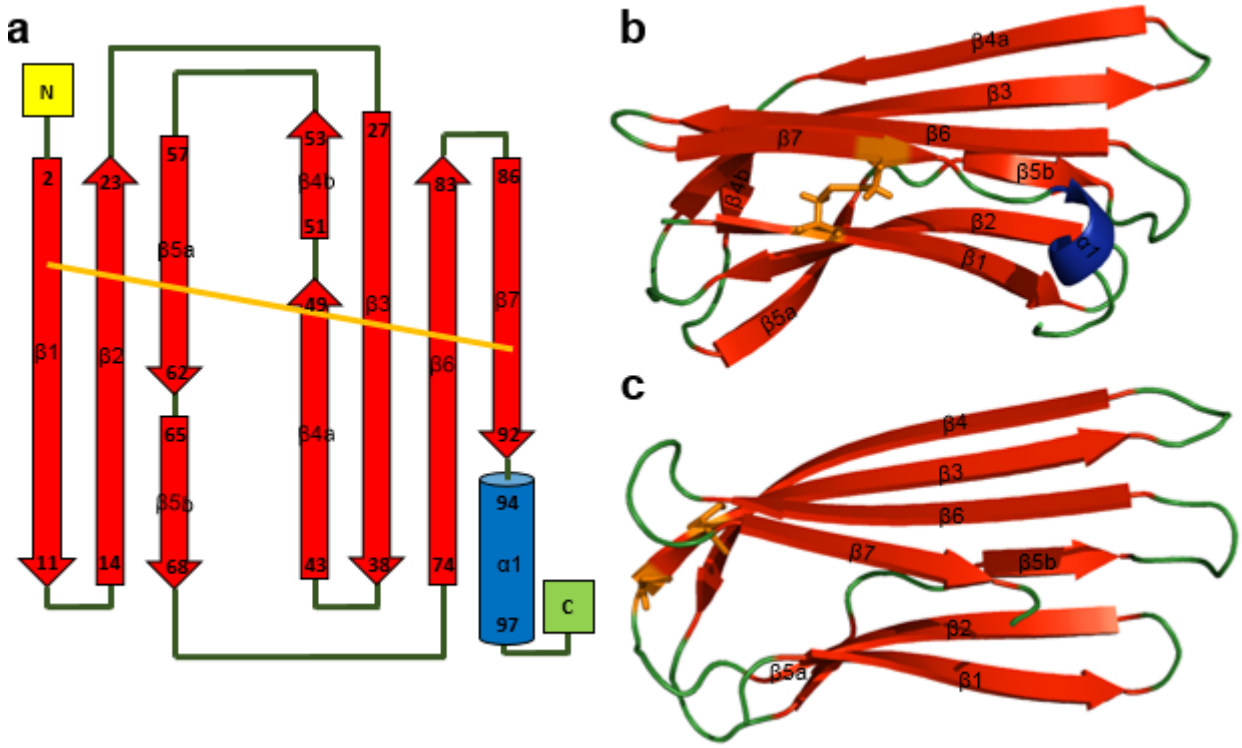


Figure 3.62: Topology and Disulphide bond position in CsgH. a, Topology of diagram of CsgH with B-strands shown as red arrows and helices as blue cylinders, with their starting and ending residue positions numbered; linkers are shown as green lines. The disulphide bond is indicated with an orange line; and the N-terminus and C-terminus are indicated with boxes labelled N and C respectively. **b**, Cartoon diagram of CsgH structure with the secondary structure highlighted by colour with the strands in red, loops in green and helices in blue, the disulphide bridge has been shown as sticks and coloured in orange. **c**, Cartoon diagram of CsgC structure with the secondary structure highlighted by colour with the strands in red, loops in green and helices in blue, the disulphide bridge has been shown as sticks and coloured in orange. The two structures are aligned and show that the position of the bridge is different and show how the disulphide bridge links the first and last strands of CsgH with a covalent bond.

Interestingly both proteins appear to have a similar pattern of surface charge with clear separation between a largely negative patch on one face of the protein and a positively charged patch on another face (**Figure 3.63**). The positions of these patches are largely similar with the positive patch most conserved in position. Both positive and negative patches appear particularly distinct on CsgH and may have a role in function.

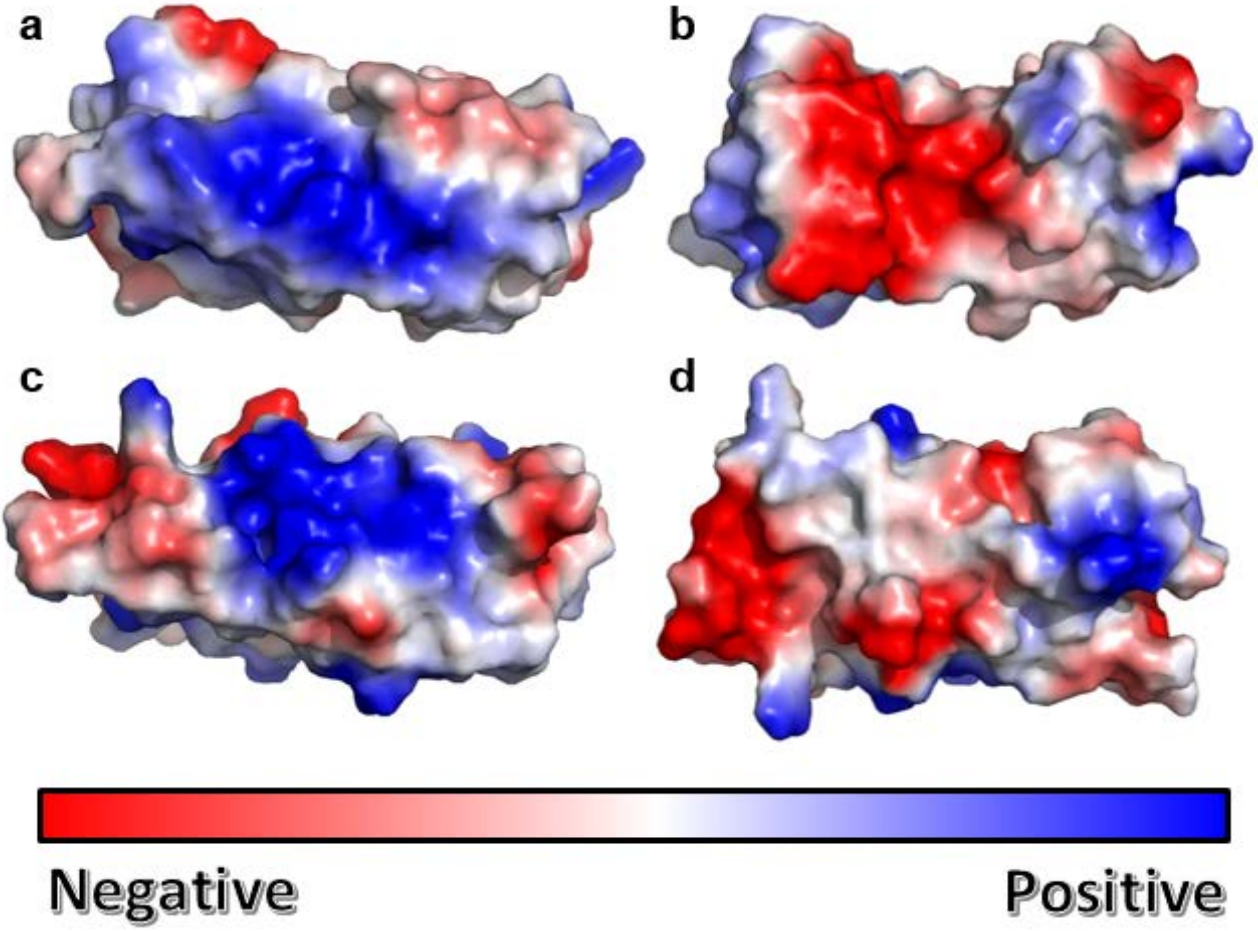


Figure 3.63: Surface electrostatics of CsgH and CsgC. a, Structure of CsgH showing the surface coloured by electrostatics illustrating the positively charged patch on the surface of CsgH. **b**, Structure of CsgH showing the surface coloured by electrostatics illustrating the negatively charged patch on the surface of CsgH. **c**, Structure of CsgC showing the surface coloured by electrostatics illustrating the positively charged patch on the surface of CsgC. **d**, Structure of CsgC showing the surface coloured by electrostatics illustrating the negatively charged patch on the surface of CsgC. Note that the figures are oriented so that the perspective of A and B correspond to C and D respectively. The red shaded regions are negatively charged and blue shaded regions are positively charged while the white regions are uncharged.

3.5.7 Bioinformatics analysis of CsgH structure

The structure of CsgH was used to search for structural homologs using Dali [**192**] (**Table 3.2**) and BioXGEM 3D-BLAST (**Table 3.3**) [**193, 194**], the closest structural homolog according to both proteins was CsgC but a broad range of other proteins were also identified reflecting the ubiquity of the IG fold. Many of the proteins are involved in Protein-Protein interactions (e.g. Interleukin), particularly on the cell surface (e.g. ICAM-1), no proteins known to be involved in amyloid inhibition were detected as homologous other than CsgC. The BioXGEM results included several β-barrels in the results but these results can be explained, depending on the case on subdomains of the barrel having IG-like folds or as artefacts resulting from the high propensity of β-strands in both proteins.

Table 3.2: Summary of Dali Results. Showing examples from each group of proteins identified by Dali [**192**] with a Z-score of 6 or higher. Displaying the protein name, the species the protein is from and the sequence identity to CsgH over the aligned sequence. Note that for many of these proteins CsgH aligns to an IG-like subdomain of a larger protein.

Protein	Source	Sequence Identity
CsgC	*Escherichia coli*	20 %
Transglutaminase	*Pagrus major*	9 %
Beta-mannosidase	*Bacteroides thetaiotaomicron*	7 %
Coagulation Factor XIII	*Homo sapiens*	7 %
Interferon Gamma	*Homo sapiens*	9 %
Beta-galactosidase	*Bacteroides fragilis*	7 %
Interleukin-22	*Homo sapiens*	9 %
ApaG	*Bordetella pertussis*	5 %
Integrin beta-4	*Homo sapiens*	8 %
CD2	*Homo sapiens*	8 %
IG new antigen receptor	*Ginglymostoma cirratum*	7 %
RbmA	*Vibrio cholerae*	11 %

Table 3.3: Summary of BioXGEN 3D BLAST Results. Showing examples from each group of proteins identified by BioXGEN 3D BLAST [**193, 194**] with an E-value below e^{-10}. Displaying the protein name, the species the protein is from and the structural identity to CsgH over the aligned regions. Strangely these results include several β-barrel proteins, there is no evidence that CsgH is a membrane protein, this is likely due to the high propensity of β-strands in both proteins.

Protein	Source	Structural Identity
CsgC	*Escherichia coli*	41 %
ICAM-1	*Homo sapiens*	38 %
SufD	*Escherichia coli*	40 %
PapC	*Escherichia coli*	34 %
OpmA	*Rhodopseudomonas blastica*	36 %
Alpha Adaptin-C	*Mus musculus*	35 %
NedA	*Micromonospora viridifaciens*	38 %
XynY	*Ruminiclostridium thermocellum*	43 %
DUF1425	*Shewanella loihica*	39 %
M-keima	*Montispora sp. 20*	42 %
PEBP2	*Homo sapiens*	37 %

3.5.8 Summary

Several CsgH constructs were expressed and compared using NMR to identify the best construct for further study. The C-terminally hexahistidine tagged version of CsgH with the first nine amino acids removed ($CsgH^{10\text{-}106}CTH$) was selected for structure solution. The solution structure of CsgH was solved using NMR producing a well-defined structure with minimal violations. The structure of CsgH is an IG-like β-sandwich similar to CsgC, with a backbone RMSD between the proteins of ~2.4 Å despite only 19 % sequence identity. Similarly there is a conservation in surface charge between the two proteins with both proteins possessing distinct negative and positive patches on opposite surfaces of the protein. The structure of CsgH was used to probe the PDB for structural homologs, CsgC was detected as a close homolog but otherwise no interesting structural homologs were identified.

3.6 CsgH Function

3.6.1 CsgH Inhibition of CsgA

CsgH shares a similar structure to CsgC and is present in strains where CsgC is absent, CsgC is known to inhibit amyloid formation and by homology CsgH is expected to have a similar function. To test whether CsgH can inhibit amyloid formation the ThT amyloid fibrillation assay was used to measure amyloid formation of CsgA with and without CsgH. CsgA was purified under denaturing conditions, transferred into native buffer, filtered and exchanged into phosphate buffer (**2.3.5**). Purified CsgH was added at a range of concentration ratios and the amyloid formation monitored over a couple of days. CsgH was found to inhibit amyloid fibrillation with a similar efficacy to CsgC (**Figure 3.64**). Both inhibitory proteins increase the apparently duration of the lag phase and decrease the overall maximum level of fluorescence.

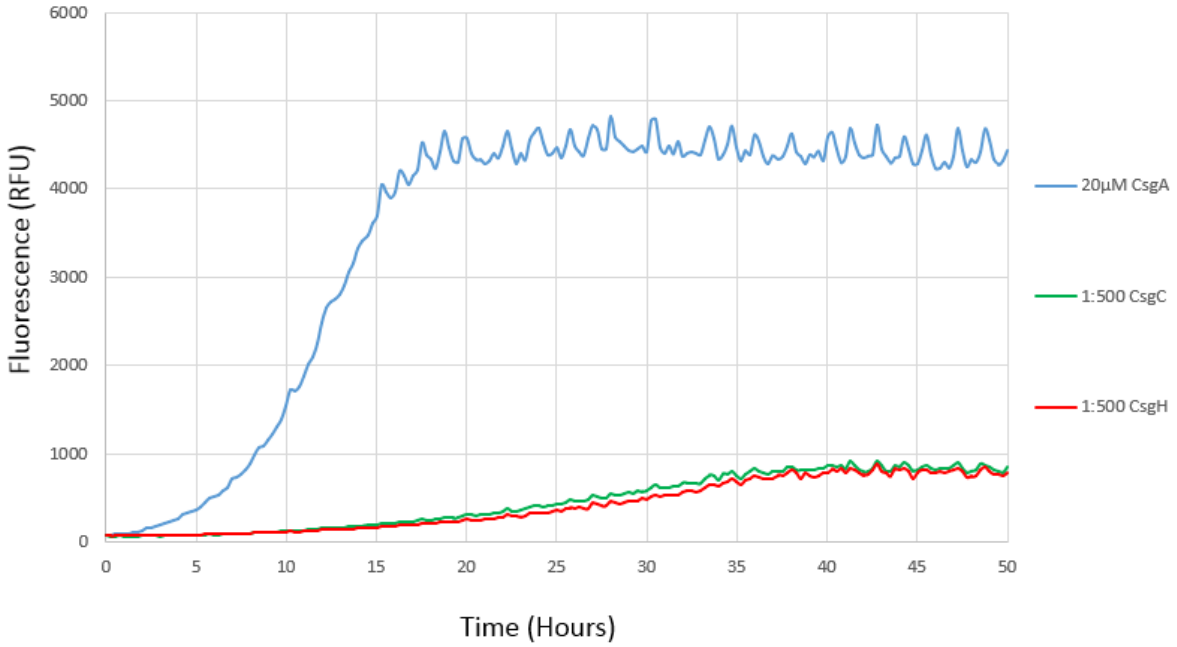


Figure 3.64: ThT Fluorescence Curves showing CsgH Inhibition of Amyloid Formation. CsgA alone (Blue) shows the characteristic amyloid formation curve observed for curli amyloid formation. Interestingly CsgH (Red) and CsgC (Green) show almost identical levels of inhibition with an elongated lag phase and a lower final level of ThT fluorescence.

3.6.2 ThT assay of CsgH Mutants

3.6.2.1 Initial CsgH Mutants

To probe the function of CsgH a series of mutations were made to the sequence to test the effect of point mutations. Research by Dr Jonathan Taylor had already identified several regions of CsgC which appeared to be important for function, analogous sites were identified in CsgH and these residues (**K32A, D36S, R45S, K47E**) were mutated (**Figure 3.65**), in CsgC multiple point mutations were needed to detect a noticeable change in activity (Personal Communication From Dr Jonathan Taylor) so a quad mutant of all four residues and double mutants with pairs of mutations on each strand were designed. These four residues are all localised to a large positively charged patch on the surface of CsgH. The mutant proteins were expressed and purified using the same method as CsgH wild type (CsgH-WT) and tested for stability by 1D NMR. The proteins appear to be folded retaining peak dispersion and ring-shifted methyl peaks and overall showing minimal changes from the wild type protein (**Figure 3.66**).

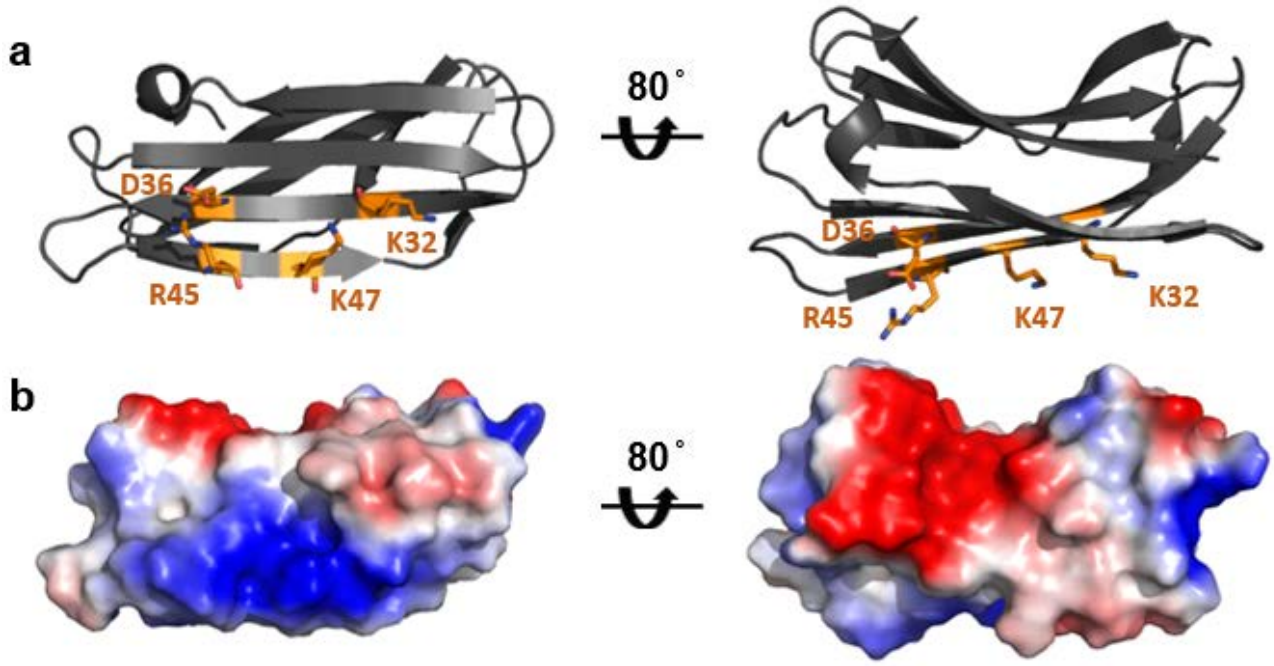


Figure 3.65: Initial CsgH Mutants. These mutants were selected based on CsgC mutants produced by Dr Jonathan Taylor. **a**, Cartoon diagrams with the mutated residues coloured with carbon atoms in orange, nitrogen atoms in blue and oxygen atoms in red and labelled with the residue number and amino acid type. **b**, Surface diagram of CsgH structure coloured by charge with negative patches in red, positive patches in blue and neutral patches in white and oriented to match the structures shown in section A.

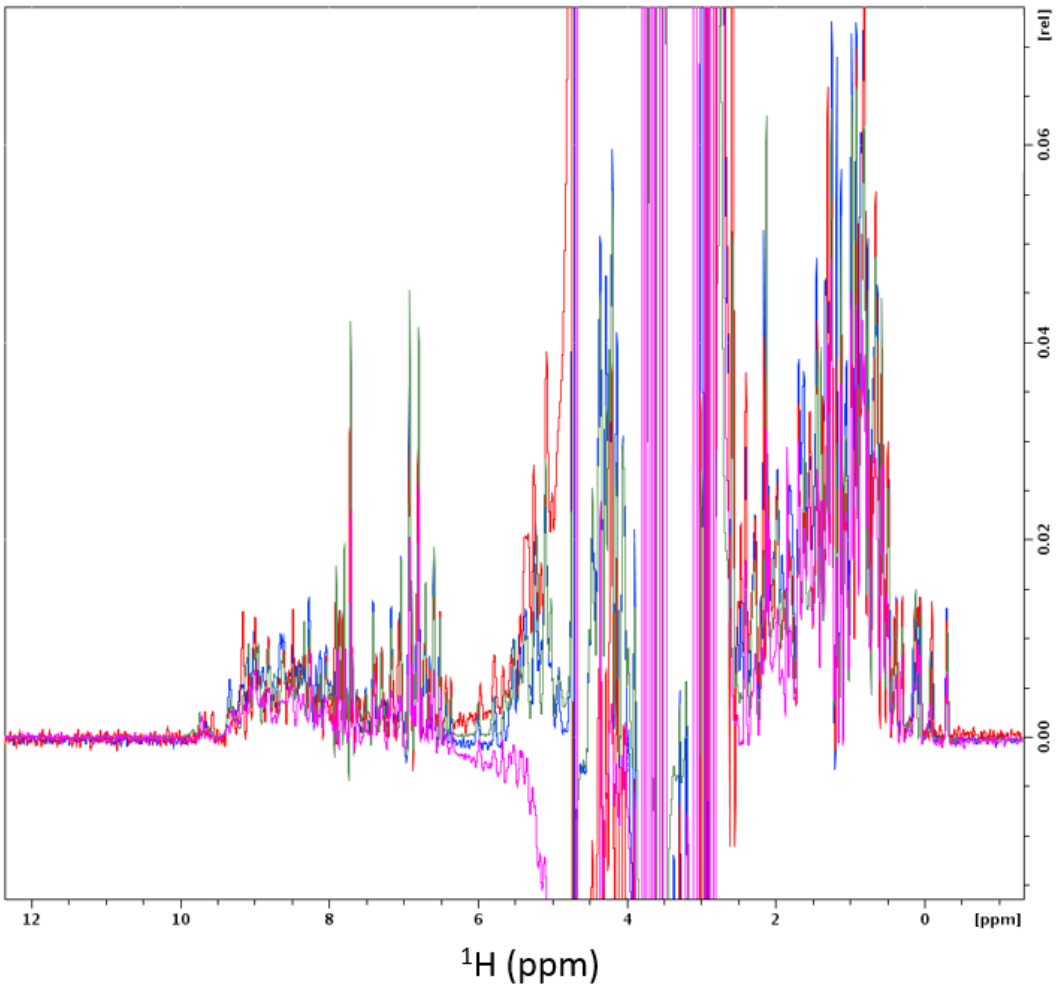


Figure 3.66: Overlay of 1D NMR spectra showing mutants of CsgH retain overall structure. The 1D spectrum of CsgH WT is shown in blue, the quad (K32A, D36S, R45S, K47E) mutant in Red, the double mutant R45S, K47E in green and the double mutant K32A, D36S in pink. Spectra were collected at 295˚K in 100 mM NaCl, 20 mM HEPES pH 7.5 on the Avance III 600 MHz.

The ability of the quad mutant and the double mutant (R45S, K47E**)** to inhibit amyloid formation was tested using the ThT assay for amyloid fibrillation. The mutant proteins were tested at a 1:500 ratio of CsgH:CsgA and compared to CsgA alone and the wild type protein at the same ratio. Both sets of mutations had a very strong effect on the ability of CsgH to inhibit amyloid formation (**Figure 3.67**). As the protein structure and stability does not appear significantly affected this indicates that these residues have an important role in the function of CsgH.

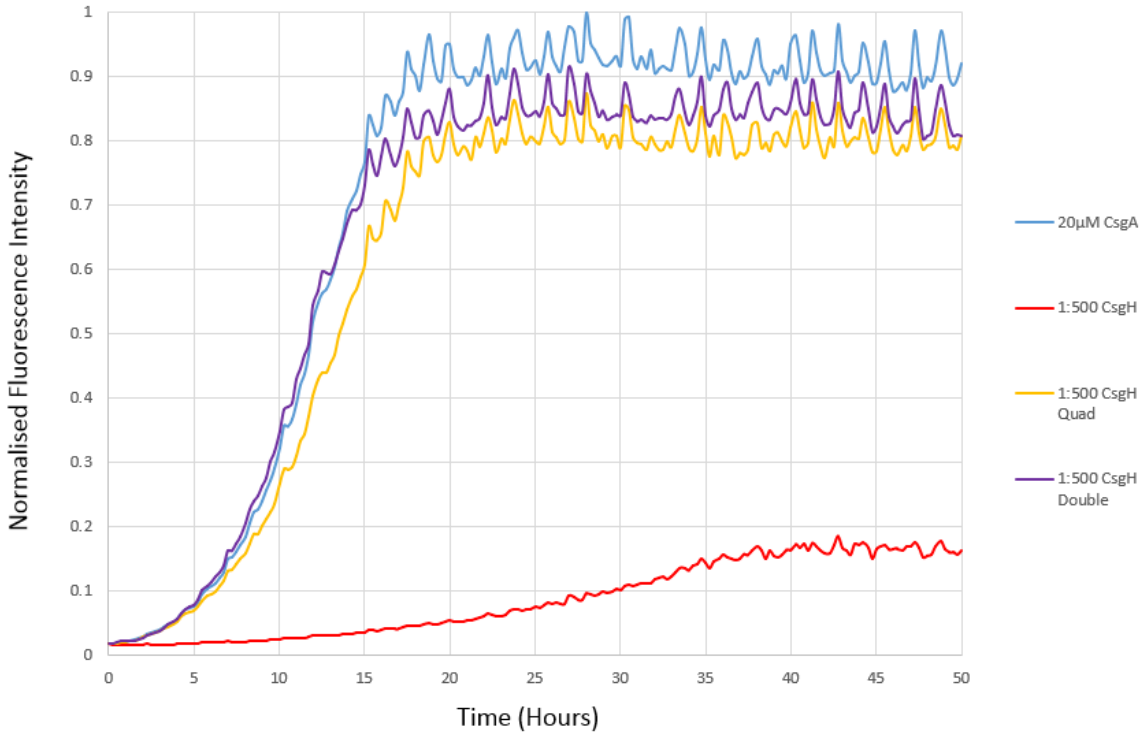


Figure 3.67: Normalised ThT Fluorescence Curves for initial CsgH Mutants. The fluorescence curves are normalised so that the maximum observed fluorescence is 1. Both the quad mutant (K32A, D36S, R45S, K47E) of CsgH (Orange) and a double mutant (R45S, K47E) (Purple) show a significant decrease in their ability to inhibit amyloid with similar lag phase and final fluorescence to CsgA alone (Blue). CsgH wild type is included for comparison (Red).

3.6.2.2 Further CsgH Mutants

Further point mutations were designed mutating individual residues based on mutants produced by Dr Taylor on CsgC and based on the NMR structure of CsgH as well as the initial chemical shift data (**Figure 3.70**). The residues R45, K47. R25. D93. T16, Q18 and V20 were all mutated in CsgH (**Figure 3.68**).

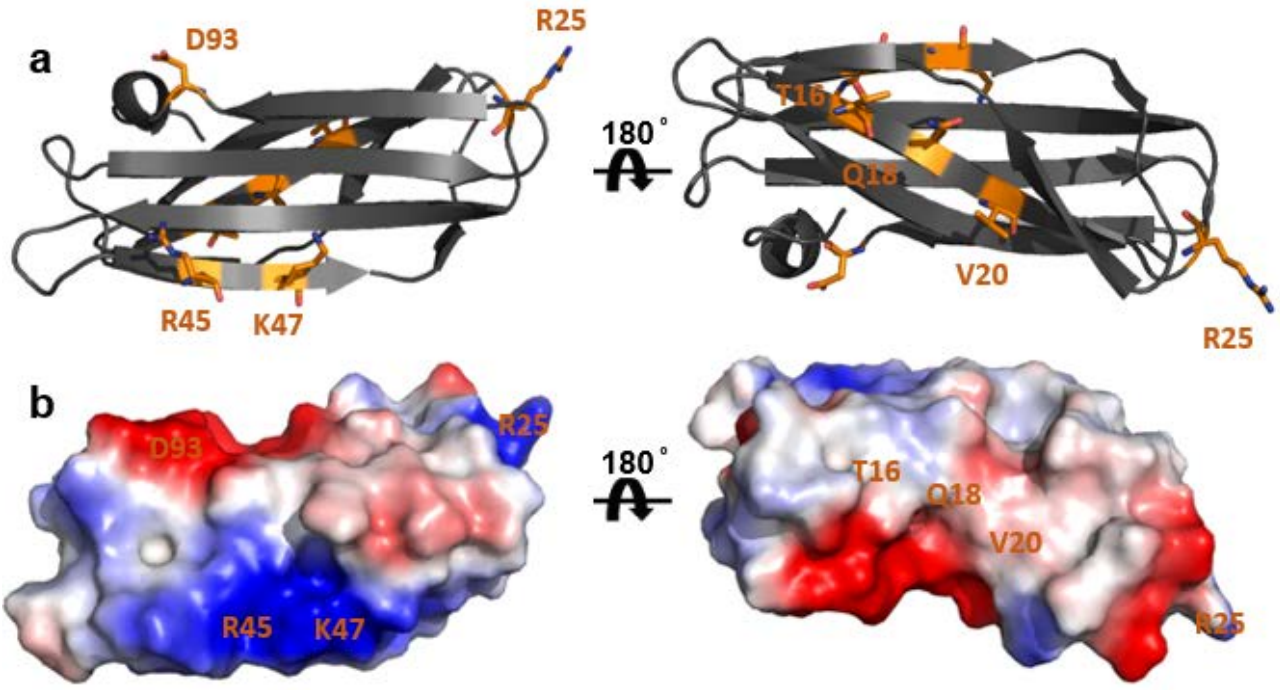


Figure 3.68: Further CsgH Mutants. Based on CsgC mutants produced by Dr Jonathan Taylor and using the structure of CsgH. **a**, Cartoon diagrams with the mutated residues coloured with carbon atoms in orange, nitrogen atoms in blue and oxygen atoms in red and labelled with the residue number and amino acid type. **b**, Surface diagram of CsgH structure coloured by charge with negative patches in red, positive patches in blue and neutral patches in white and oriented to match the structures shown in section A with labels showing the location of the mutant residues on the surface.

The mutant proteins were purified and expressed using the same method as the wild type protein and were tested for their ability to inhibit CsgA amyloid fibrillation using the ThT assay. The multiple point mutations had a significant effect on the protein functions as did the mutations to individual charged residues such as K47 and D93, interestingly mutations to R45 and R25 did not have a significant effect on amyloid inhibition. (**Figure 3.69**). These results suggest that specific charged residues and the regions near the charged patches are important for function but that the bulk charge of the protein alone is not sufficient for function.

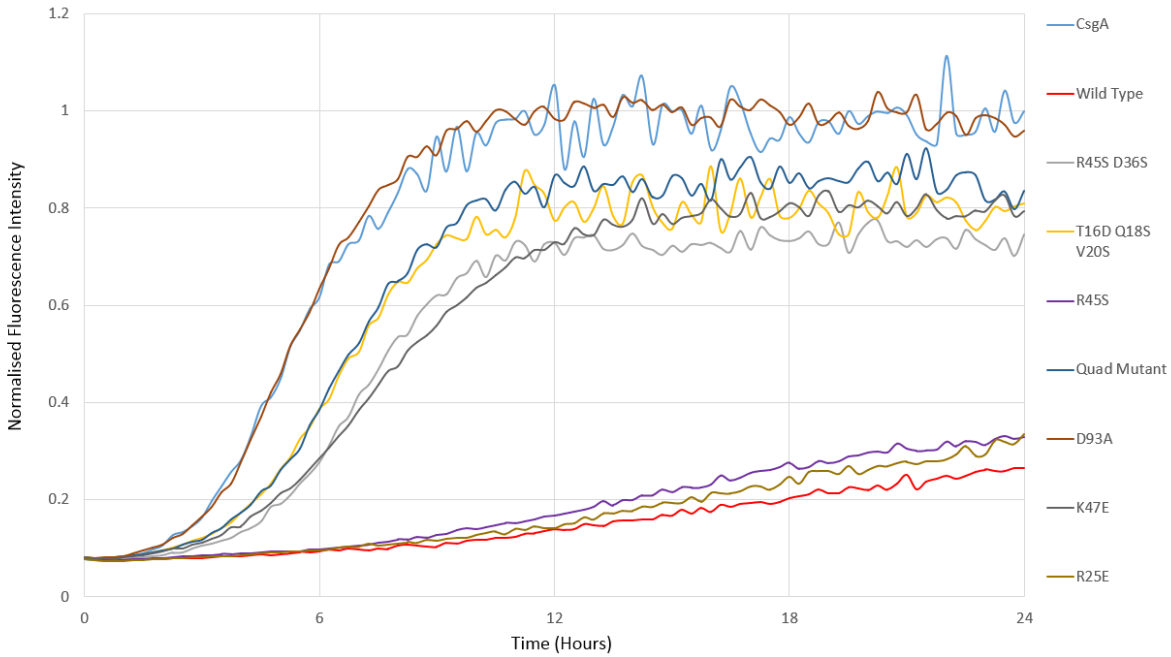


Figure 3.69: Normalised ThT Fluorescence Curves for initial CsgH Mutants. The fluorescence curves are normalised so that the final observed fluorescence of CsgA alone is 1. Many of the mutations show a significant decrease in the ability of the protein to inhibit amyloid formation however, R25 and R45 mutants have minimal effect of the protein function.

3.6.3 NMR of CsgH with CsgA

3.6.3.1 CsgH-CTH

Since CsgH can inhibit CsgA in vitro without any additional factors it is likely that the protein must interact with CsgA at some point during the process of amyloid formation. To probe this process ^{15}N labelled CsgH was expressed, purified and combined with freshly purified unlabelled CsgA. A series of 2D ^{1}H ^{15}N-HSQC spectra were collected for this mixture over time and then compared to the spectra of a sample of CsgH alone. Shifts can be observed between CsgH and CsgH with CsgA as well as between CsgH with CsgA samples over time (**Figure 3.70, Figure 3.71**).

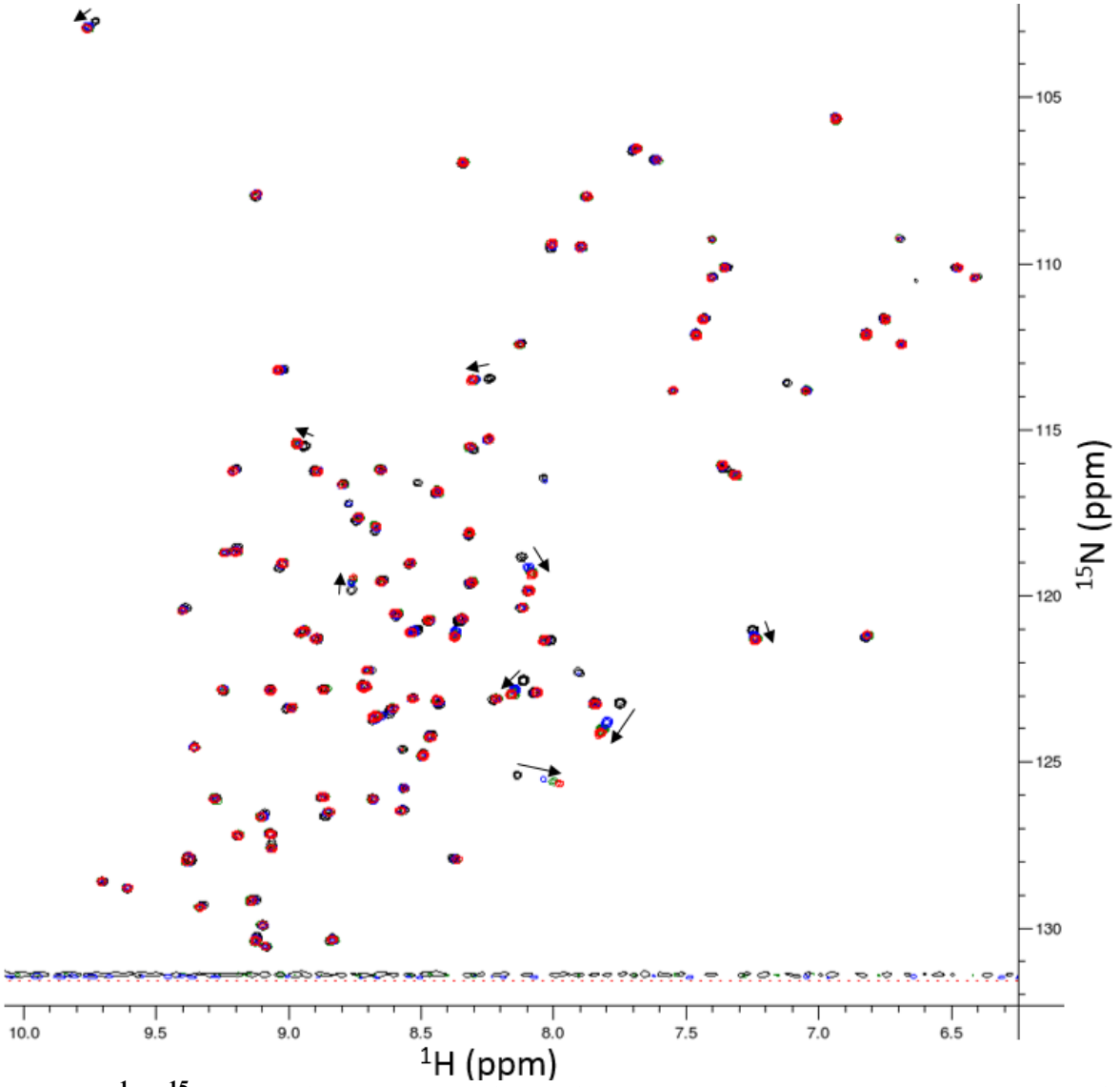


Figure 3.70: ^{1}H ^{15}N HSQC spectra of CsgH-CTH in the presence and absence of CsgA. Overlay of ^{1}H ^{15}N HSQC spectra of CsgH-CTH (Black), CsgH-CTH with CsgA (1:5) 0 hour incubation (Red), CsgH with CsgA (1:5) 4 hour incubation (Green) and CsgH with CsgA (1:5) 10 h incubation (Blue). The majority of the peaks show no significant movement but some peaks shift significantly in the presence of CsgA, arrows indicate the movement of peaks exhibiting particularly notable changes in chemical shift. The peaks can also be seen to move over time in the presence of CsgA possibly due to changes in the effective concentration of CsgA. Further details of peak changes are shown in **Figure 3.71**.

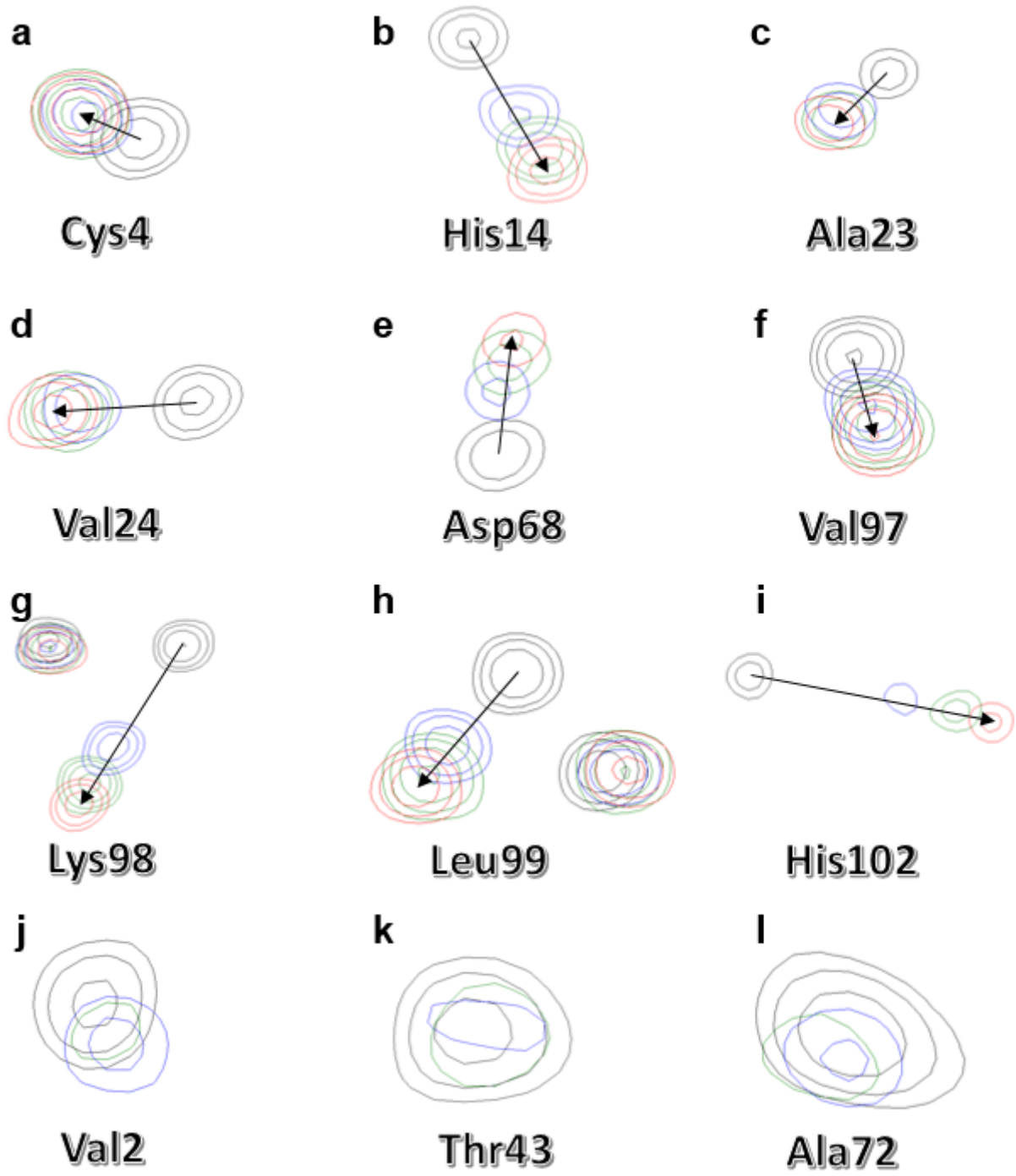


Figure 3.71: Examples of peak changes between CsgH with and without CsgA. Individual peaks have been selected from the HSQC overlay shown in **Figure 3.70**, CsgH-CTH alone is shown in black, CsgH-CTH with CsgA 0 hour incubation in red, CsgH with CsgA 4 hour incubation in green and CsgH with CsgA 10 h incubation in blue. Several peaks (**a-i**) show significant shifts when CsgA is added and then are seen to gradually move back towards the position of the CsgH alone over time. Notably several of these peaks are in or near the tag (**f-i**) and it is possible that the tag is interfering with the interaction. Some peaks are observed to disappear when CsgA is added to CsgH (**j-l**) and then slowly re-appear over time. Note that the chemical shift axis have been removed for clarity.

The peak changes between CsgH alone and CsgH with CsgA can be mapped onto the CsgH structure (**Figure 3.72**). Many of the largest peak changes localise to the region near the C-terminal polyhistidine tag and linker which suggests that the tag may be interfering with the CsgH-CsgA interaction. It is possible that the tag has some interaction with the nearby charged patch on the surface of CsgH (**Figure 3.72B**) and the shifts may reflect the tag being displaced from this position by CsgA. The majority of the peak perturbations (**Figure 3.71A, B, C, D, E, F, G, H and I**) show evidence of fast exchange which is indicative of a weak interaction however there are some peaks which are undergoing intermediate exchange (**Figure 3.71J, K and L**) which suggests they are experiencing a slightly stronger interaction. As the effective concentration of CsgA is uncertain at each time point of the experiment it would not be possible to accutately estimate the dissociation constant (K_d) from this experiment although it does appear to be a weak interaction from the exchange mechanism.

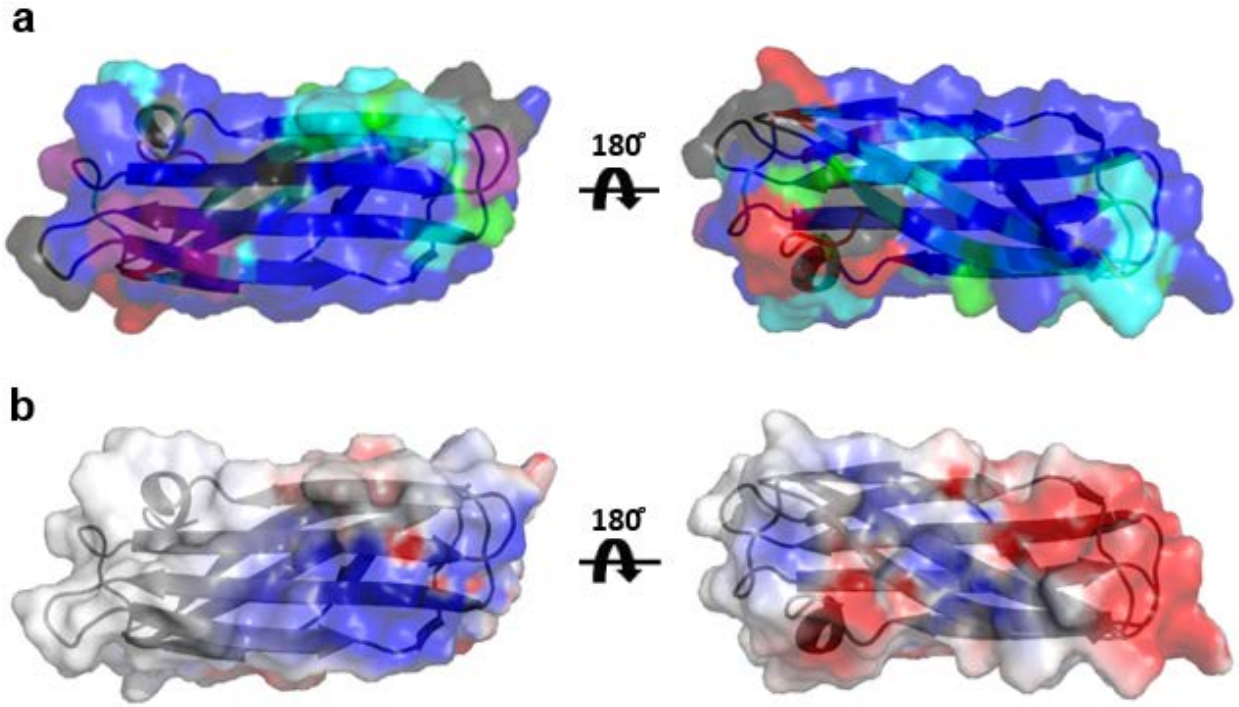


Figure 3.72: Chemical shift perturbations of CsgH-CTH and relationship to surface charge. a, Map of chemical shift perturbations of CsgH-CTH in the presence of CsgA. Residues which show no indications of change are shown in blue, peaks for which data was missing are shown in grey, peaks which disappear in the presence of CsgA are shown in purple, the remaining peaks show varying degrees of chemical shift changes in the presence of CsgA, the distances moved were calculated as described in the materials and methods (**2.5.1.3, Equation 2.1**), the structure was coloured from blue to red in order of ascending chemical shift displacement (D), with less than 0.02 shown in blue, 0.02-0.05 shown in light blue, 0.05-0.08 shown in green, 0.08-0.11 shown in yellow and more than 0.14 shown in red. **b**, Surface of CsgH coloured by charge with negative patches in red, positive patches in blue and neutral patches in white oriented to match the perspective of the images in section **a.** It is clear that the strong shifts observed at the C-terminus are near this negative patch.

3.6.3.2 CsgH-NTH

To eliminate the issues with the C-terminal histidine tag we reverted to the N-terminal His-tagged construct for the CsgA titration experiment. The CsgH-NTH construct was previously tested but not used for structure solution as described in section **3.5.1**. ^{15}N labelled CsgH-NTH was expressed and purified as before and then combined with freshly purified CsgA in the same manner as the CsgH-CTH construct. The strength of the chemical shifts was slightly smaller than those seen for CsgH-CTH (**Figure 3.73**), possibly reflecting a lower concentration of CsgA. Similar to the CsgH-CTH experiment the majority of peak perturbations still showing evidence of fast exchange, with a few showing evidence of intermediate exchange giving the overall indication that the interaction is weak.

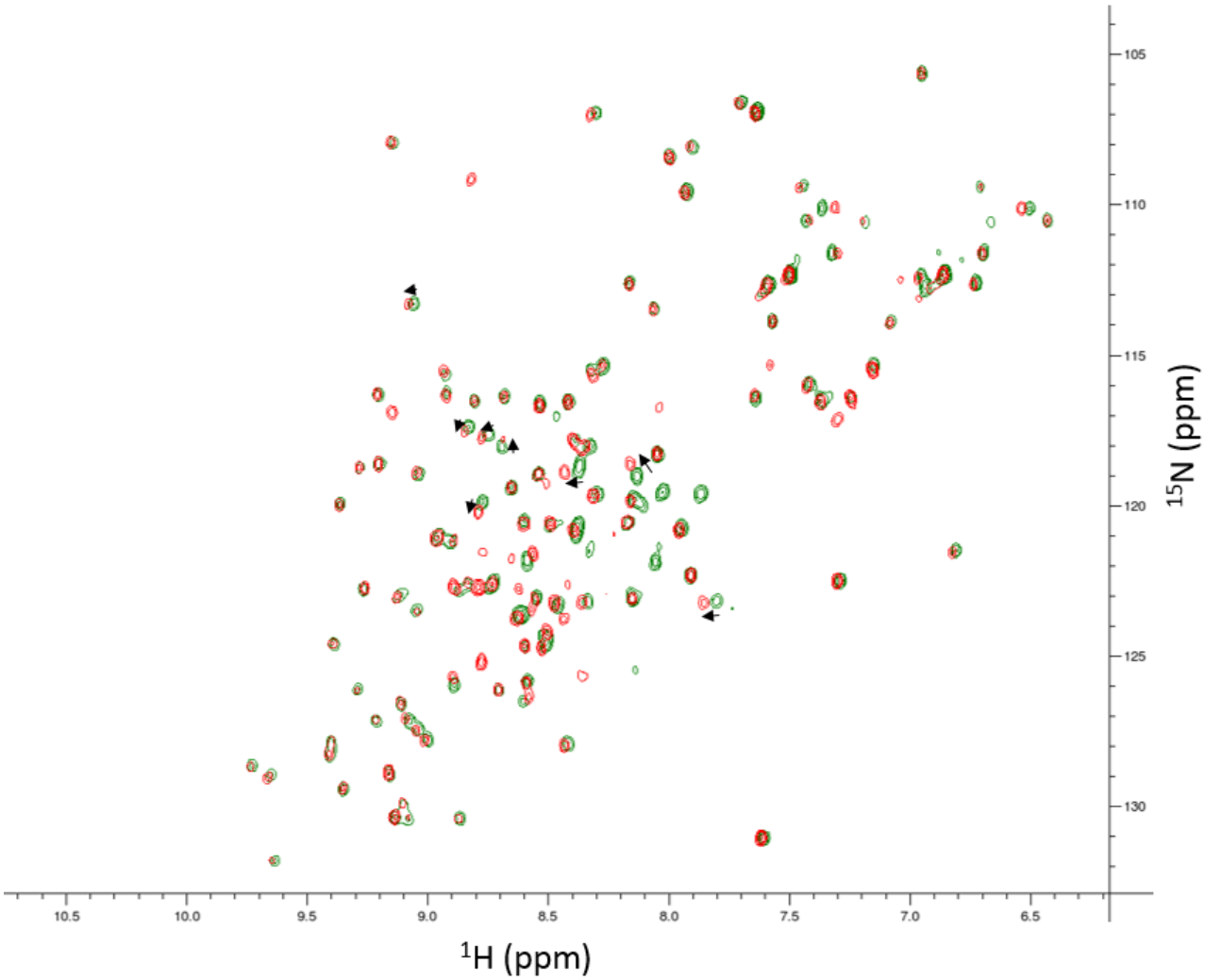


Figure 3.73: ^{1}H ^{15}N HSQC spectra of CsgH-NTH in the presence and absence of CsgA. Overlay of ^{1}H ^{15}N HSQC spectra of CsgH-NTH (Green) and CsgH-NTH with CsgA (1:3) 0 hour incubation (Red). The majority of the peaks show no significant movement but some peaks shift significantly in the presence of CsgA, arrows indicate the movement of peaks exhibiting changes in chemical shift.

A backbone assignment has not been completed for CsgH-NTH but the majority of peaks for the core protein (77/98) can be mapped directly from the CsgH-CTH assignments, these tentative assignments are shown in **Figure 3.74**. The missing assignments are a caused either due to the absence of peaks or ambiguity resulting from peak overlap or uncertainty over which peak has moved.

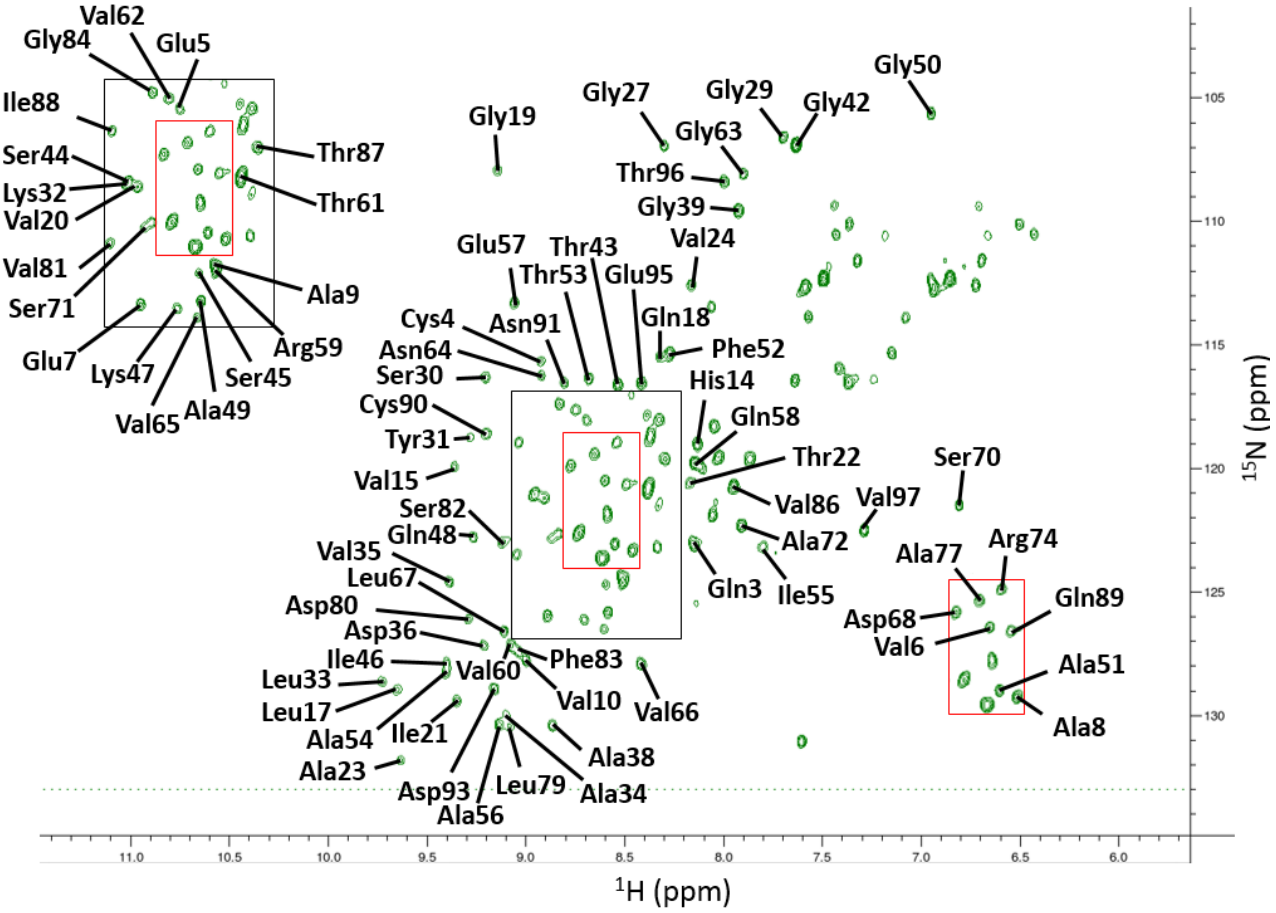


Figure 3.74: Backbone Assignments for CsgH-NTH. Assignments are based on known assignments for CsgH-CTH. 2D ^{15}N ^{1}H HSQC of CsgH-NTH is shown in green and labelled with the tentative assignments based on CsgH-CTH (**Figure 3.54**). Many of the unassigned peaks occur near the centre of the spectra where peak overlap and poor dispersion make precise assignment more difficult.

The changes in chemical shift can then be mapped onto the CsgH structure **(Figure 3.75)** the results indicated that there may be an interacting region in the protein near the negatively charged region of the protein**.**

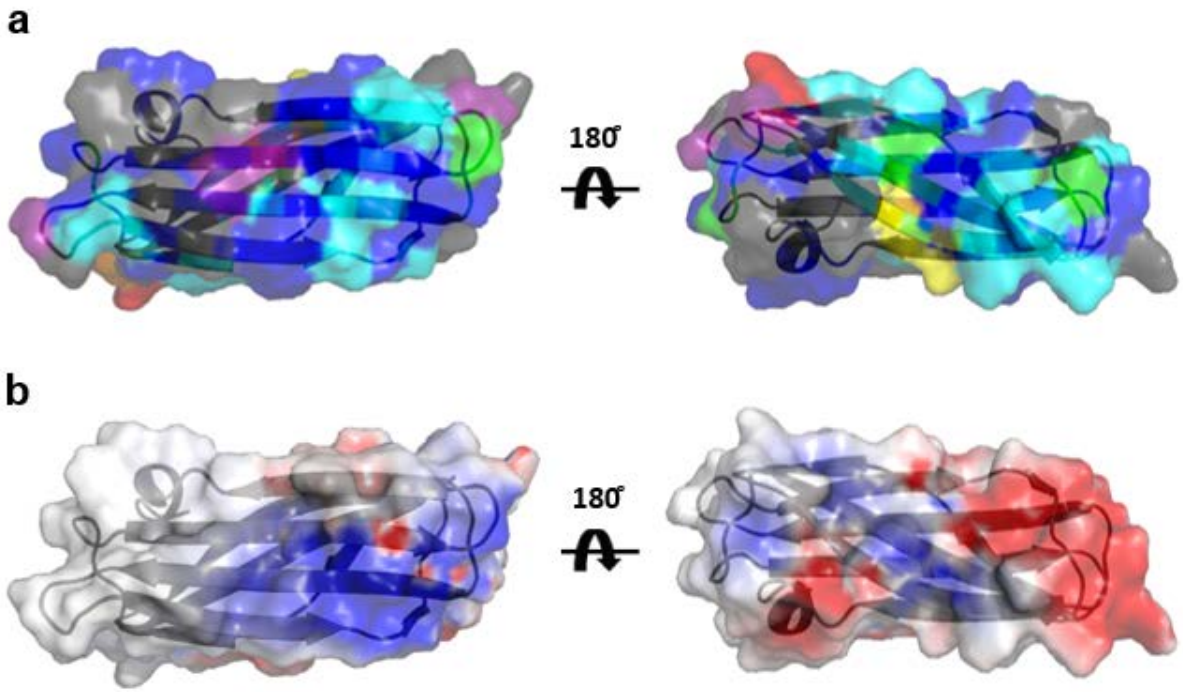


Figure 3.75: Chemical shift perturbations of CsgH-NTH and relationship to surface charge. a, Map of chemical shift perturbations of CsgH-NTH in the presence of CsgA. Residues which show no indications of change are shown in blue, peaks for which data was missing are shown in grey, peaks which disappear in the presence of CsgA are shown in purple, the remaining peaks show varying degrees of chemical shift changes in the presence of CsgA, the shift deviations were calculated as described in the materials and methods (**2.5.1.3, Equation 2.1**), the structure was coloured from blue to red in order of ascending chemical shift displacement (D), with less than 0.02 shown in blue, 0.02-0.05 shown in light blue, 0.05-0.08 shown in green, 0.08-0.11 shown in yellow and more than 0.14 shown in red. **b**, Surface of CsgH coloured by charge with negative patches in red, positive patches in blue and neutral patches in white oriented to match the perspective of the images in section A, the face of CsgH around the negatively charged region shows clustering of observed changes in chemical shifts, suggesting an interacting region, while the reverse face, where the positively charged patch is present shows minimal evidence of interaction.

To show that the N-terminally tagged CsgH was still active it was tested for inhibition of CsgA polymerisation using the ThT assay (**Figure 3.76**). The results showed that CsgH-NTH showed similar efficacy to CsgH-CTH.

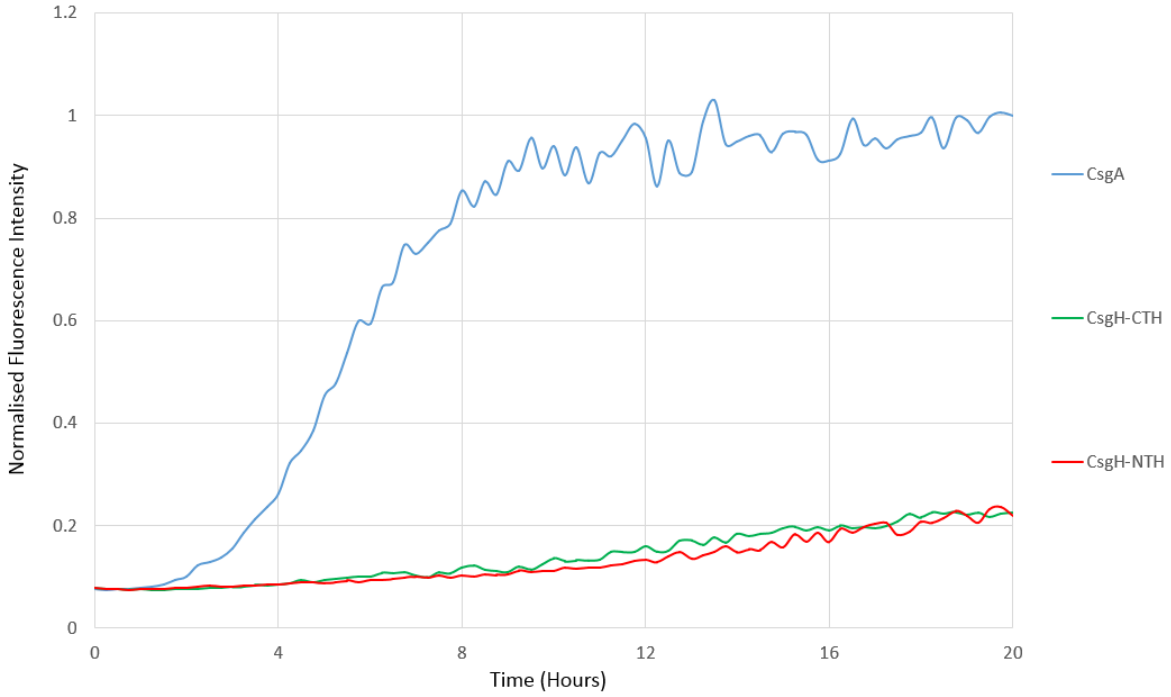


Figure 3.76: Comparing CsgH-CTH and CsgH-NTH Inhibition of Amyloid Formation. The fluorescence intensity is normalised so that the final measured fluorescence of CsgA alone is 1. The location of the hexahistidine tag on CsgH appears to have no significant effect on the ability of the protein to inhibit amyloid fibrillation, with the lag phase and overall intensity limited to a similar level.

Many of the chemical shifts showing shifts are conserved between the CsgH-CTH and CsgH-NTH. To try to eliminate the contributions from the tags in the two datasets were combined and then mapped onto the structure with all the chemical shifts averaged (**Figure 3.77B**) and also averaged and reduced to show only those peaks which shift significantly in both experiments (**Figure 3.77A**). This data can be compared to the charged surfaces of the protein illustrated in **Figure 3.77C**. Similar to the results for CsgH-NTH there is an indication of an interacting face near the negatively charged patch on the protein surface.

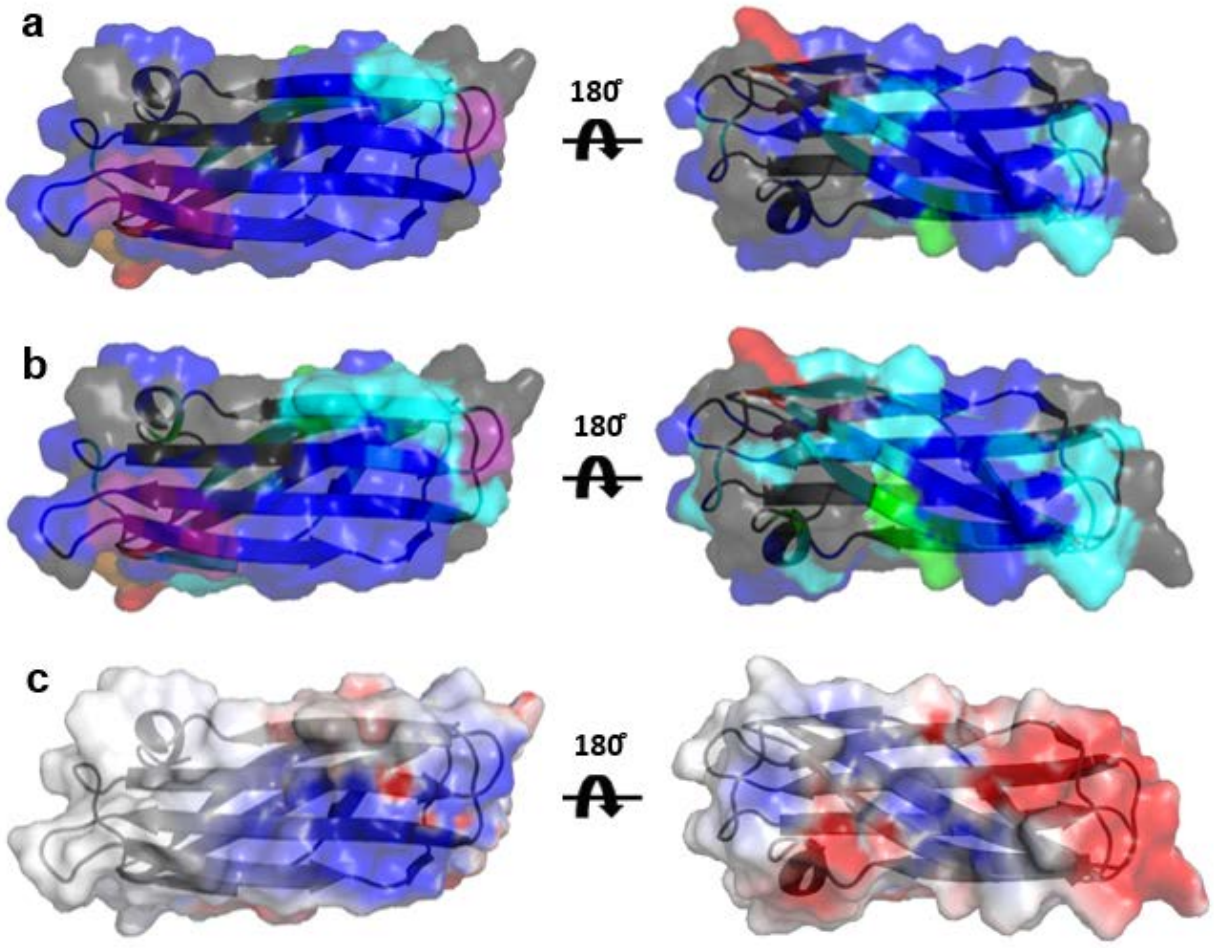


Figure 3.77: Comparison and amalgamation of CsgH-CTH and CsgH-NTH chemical shift data. The shift deviations were calculated as described in the materials and methods (**2.5.1.3, Equation 2.1**), the structures in **a** and **b** were coloured from blue to red in order of ascending chemical shift displacement (D), with less than 0.02 shown in blue, 0.02-0.05 shown in light blue, 0.05-0.08 shown in green, 0.08-0.11 shown in yellow and more than 0.14 shown in red. **a**, Map of chemical shift perturbations of CsgH in the presence of CsgA, with the D values averaged and a conservative cut-off is applied where if the chemical shift perturbation falls below 0.02 for either the experiment with CsgH-CTH or CsgH-NTH the residue in question coloured in Dark Blue (Indicating no significant shift). **b** ,Map of chemical shift perturbations of CsgH in the presence of CsgA, with the D values simply averaged, this should reduce the significantly reduce the impact of the tags on chemical shift perturbations but may not eliminate particularly large shifts, fortunately most of these large shifts are already discounted as the tag and linker cannot be compared between the CTH and NTH constructs. **c**, Surface of CsgH coloured by charge with negative patches in red, positive patches in blue and neutral patches in white oriented to match the perspective of the images in sections A and B.

3.6.4 CsgH and FapC

To test whether CsgH can work on other bacterial functional amyloid the protein was combined with FapC over a range of concentrations and amyloid formation was monitored using the ThT assay. Interestingly CsgH was capable of inhibiting FapC, albeit less efficiently than it could inhibit CsgA fibrillation (**Figure 3.78**), indeed comparison with previously observed CsgA inhibition (**Figure 3.67, Figure 3.78)** it appears that CsgH is fivefold more effective at inhibiting CsgA. The ability of CsgH to inhibit FapC is particularly interesting as the two amyloid proteins share little sequence similarity even in their amyloid repeats indicating CsgH has a robust, promiscuous mechanism of amyloid inhibition.

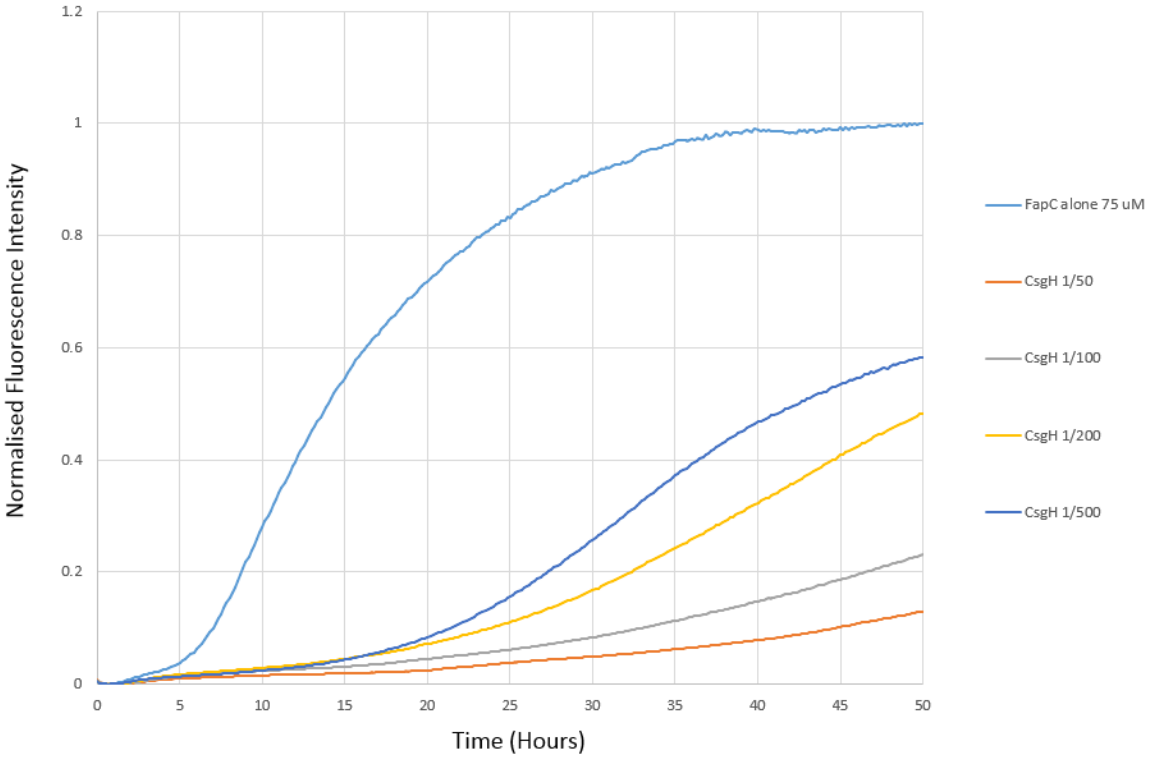


Figure 3.78: Normalised ThT Fluorescence Curves for FapC with varying concentrations of CsgH. The fluorescence curves are normalised so that the maximum observed fluorescence is 1. CsgH inhibits amyloid formation by FapC as it extends the lag phase and decreases the overall maximum fluorescence. The effectiveness of the inhibition increases with concentration.

3.6.5 Summary

CsgH was shown to inhibit CsgA fibrillation with similar efficacy to CsgC indicating that the protein has a role in controlling amyloid fibrillation. Mutations of CsgH were made which reduced the ability of CsgH to inhibit CsgA fibrillation, these mutations include sets of mutations such as the double, triple and quad mutations described previously, single mutants include negatively charged residues such as D93 and K47 however some point mutations had no effect alone such as R25 and R45. These mutations suggest that surface charge is important for the function of CsgH but that the position of the charge is important as opposed to just the overall charge of the protein. NMR was used to probe the interaction between CsgH and CsgA, the results indicated that there was some form of specific interaction occurring and showing some localisation to a large negatively charged patch and the region around it on one side of the protein. Since the mutants which inhibited inhibition did not always exhibit strong shifts this may suggest that these, largely charged residues may have a role electrostatically attracting or orienting the molten form of CsgA before a more direct interaction which presumably leads to the adoption, in CsgA, of a less amyloidogenic fold. Interestingly CsgH was also shown to inhibit the *Pseudomonas* amyloid FapC despite the low sequence identity, indicating that the mechanism of CsgH/CsgC inhibition of amyloid is robust and promiscuous.

4) Discussion

4.1 Insights into the components of the Alf system

At the outset of this project our knowledge, at the molecular level, of *Pseudomonas* Alf was very limited. Bioinformatics analysis of the proteins gave us some indications as to the nature of the various components. FapD and FapF were indicated to be the main structured components in the system and showed homology to C39-like peptidases and β-barrels respectively. For the other components, the lack of related proteins, of known structure and function, meant that bioinformatics could only provide the most basic information. We have shown that FapD and FapF are indeed structured proteins and our results support the suggestion that FapA, FapB, FapC and FapE are unstructured proteins when isolated in solution. It is possible that FapA and FapE perform roles as adapters or accessory proteins to the complex similar to CsgE or CsgF which are also predicted to be disordered proteins in isolation. FapC was known to be an amyloid forming protein and the main component of the fibre, by homology it has been inferred that FapB is also an amyloid protein, our results support this suggestion as FapB appears to form amyloid in isolation based on the ThT assay. Our attempts to identify interactions between the components have been hampered by the unstructured nature of many of the proteins. Homology with subdomains of β-barrel proteins and the suggestion that FapD acts on FapF both support the possibility that FapD interacts with FapF, however we were unable to prove this interaction due to the incompatible buffers. The instability of FapD over the range of conditions and even in the presence of detergents may support the suggestion that FapD is bound to other proteins *in vivo* and is unstable in their absence, an effect which has been seen in other protein complexes [**196**].

The range of constructs produced for the system will prove useful for future work on *Pseudomonas* Alf. The repeat deletion constructs for FapB and FapC can be used to probe the effect of individual repeats on the amyloid fibrillation process. The periplasmic constructs are useful for co-expression with the full operon despite the difficulty of purifying the proteins with the full operon active; this could be resolved with some deletion mutants for the individual proteins, in particular deletion of FapC should improve the purity as the amyloid fibres should then no longer be produced. Alternatively the specificity of the pull-down experiment could be improved by using a more specific tag such as streptavidin or maltose-binding protein (MBP). Additionally if a more stable construct or more stable conditions can be identified for FapD which allow it to remain soluble in the presence of detergents it would be interesting to probe for the potential interaction between the proteins which is implied by homology. Since FapC amyloid fibres could be produced they could be used to perform solid state NMR experiments to study the structure of the amyloid fibres themselves which could be interesting for comparison to known systems [**197, 198, 199**]. As observed in the introduction both the FapC and CsgA repeats contain a putative Q/N – X_{10} – Q/N

motif, notable as the proteins share little else in common (**Table 4.1**). Our experimental results support the suggestion of a common basis for amyloid fibrillation between the two proteins as CsgH is capable of inhibiting both amyloid proteins. The presence of these similar sequences between the proteins might implicate that this is the sequence which CsgH interacts with. FapC is a much larger protein than CsgA with long loop regions dividing its three repeats and so it may be useful to isolate the individual amyloid repeat of FapC, produce it as a peptide and use it for study of the system, possibly in NMR titrations with CsgH, where it could be used at a much higher molar ratio than the full length protein. Isolation of the short Q/N – X_{10} – Q/N peptide might also be interesting for study as whether or not it retains any amyloidogenic properties it may still be able to interact with the amyloid inhibitors CsgH and CsgC.

Table 4.1: Properties of CsgA and FapC. Comparison of CsgA and FapC (UK4) features, using Protparam [**181**] to estimate MW and pI.

Feature	CsgA	FapC
pI	5	8.49
MW	15 kDa	25 kDa
Number of Amyloid Repeats	5	3
Amyloid Repeat Sequence	S-X_5-Q-X-G-X-G-N-X–A-X_3-Q	X_{15}-G-X_4-N-X_3-G-X_6-N-X_7
Total Asp+Glu	10	14
Total Lys+Arg	6	16

4.2 Towards an NMR structure of FapD

The FapD PAO1 construct was too unstable for study however construct optimisation. However with a change of *Pseudomonas* strain to PA7, truncation of the protein to remove the potentially disordered N-terminus, followed by buffer optimisation using DSF and NMR has improved the stability considerably to the point where the protein is stable enough to conduct 3D NMR experiments. Significant progress has therefore been made towards the structure of FapD. With the current construct and buffer solving the NMR structure would be an intensive task, further optimisation of the buffer has good potential to resolve this issue, particularly titrating the level of glycerol and the experimental temperature or finding an alternative additive to stabilise the protein.

Although many constructs were made in an attempt to stabilise the protein it could still be a productive avenue in pursuit of the protein structure. A fusion construct of the *Pseudomonas* PA7 version of FapD with a stabilising protein such as MBP could serve the dual purpose of stabilising the protein and provide crystal contacts for crystallisation. Having a large additional domain is not ideal for NMR studies of the protein, an alternative approach would be to revisit the homolog constructs using the purification methods and additives which proved effective in stabilising the *Pseudomonas* PA7 version of FapD as these may show better long term stability. Interactions with other components of the Alf system might also improve the stability of the protein so if interaction partners are identified in the future these could be used to stabilise the protein either by combining the proteins *in vitro* or producing a fusion construct between them. The potential presence of a transmembrane helix in some homologs of FapD as well as the extended C-terminal helices predicted by I-TASSER hint at the potential for an interaction with FapF at the outer membrane, however it has not been possible to reconstitute a FapF-FapD complex, it is possible that another protein acts as an adapter between the two structured components.

Since FapD has been suggested to cleave FapF [**92**], it would be interesting to test this process in vitro, unfortunately FapD is unstable in the presence of detergents tested so far. It may be possible to get around this aggregation problem by binding FapD to a surface and then exposing the FapF solution to this surface. To facilitate this process it would be desirable to produce an FapF construct (for *Pseudomonas* PA7) with an alternative binding tag, preferably one large enough to make N-terminal cleavage easier to observe, for example MBP or GST. The full operon plasmid might also be useful for testing this process in case accessory components are required, the problem with our current periplasmic construct for this purpose is that the tag is on the N-terminus and would be removed if the protein was cleaved, preventing us from observing cleaved species. To probe for FapF cleavage using the full operon we would ideally want a knock-out for FapF in the operon plasmid, a knock-out for FapD in the operon plasmid and a mutant of FapF with a histidine tag inserted into one of the loops.

4.3 Towards the crystal structure of the membrane protein FapF

FapF was successfully expressed insolubly and purified to produce a sample which could be refolded and was shown to be folded by NMR. The protein could be crystallised but the crystal size was very small and the diffraction weak. Optimisation of the construct and work on homologous forms of FapF did produce some improvements but the main breakthroughs were to express the protein to the membrane and to use a C8E4-LDAO mixture. This combination could produce crystals which diffracted to high resolution, a critical step in getting the protein structure; phasing will require the production of further crystals for direct phasing. Phasing with will depend upon the success of heavy metal soaking or SeMethionine incorporation, the former can be screened to find conditions for isomorphous binding, the latter may be complicated by the slightly low proportion of Methionine in the native protein sequence. The sequence of our FapF construct is approximately 420 residues long but contains only five methionine residues (as one was removed with the signal sequence), producing a ratio 1:84, it has been suggested that you need one methionine per 75 residues to reliably phase [**200**]. One way to get around this problem would be to introduce additional methionine residues by mutagenesis, preferably into predicted loop regions of the protein. Another method to allow phasing is to introduce a lanthanide binding tag and using it to bind heavy atoms to phase the protein [**201**]. A further issue with our crystals is that they are still small and so enhancing the crystallisation of the protein could produce dividends in terms of diffraction, higher diffraction is also beneficial for phasing. Enhanced crystallisation could be achieved by the incorporation of lysozyme into loop regions as has been successful with some membrane proteins such as G protein-couple receptors [**202**]. Alternatively membrane expressed versions of FapF were also produced for the *Burkholderia* and TRCM homologs which might produce better quality crystals.

4.4 CsgH Solution Structure and Functional Insights

CsgH was identified as a potentially interesting candidate for structural study due to its unknown function and the suggestion of a role as a CsgC substitute in some bacterial strains, the discovery that it is capable of inhibiting not only CsgA, but also FapC, a genetically distinct form of functional amyloid, only serves to increase the interest in the protein. A truncated construct for CsgH from *Rhodopseudomonas palustris* was selected for structural study using NMR. The structure of CsgH was found to be very similar to the crystal structure of CsgC both in terms of backbone position ~RMSD and in surface charge distribution. The implication being that the structure and charge distribution are important for the protein function. Function was further probed using mutagenesis with the ThT assay to test function. This indicated specific residues which effected the function including several surface charged residues, the position of the charge removed appears to be important as removal of charged residues on other parts of the surface did not impair function. NMR of labelled CsgH in the presence of CsgA allowed us to probe for an interaction interface, although the tag appears capable of interfering with the interaction, possibly due to its charge, there was certainly evidence of protein-protein interaction and some indication of an interface (**Figure 4.1**). The mutant residues on both faces are close to the charged patches on the protein and these regions also show some degree of conservation at least in terms of residue properties on the protein surface (**Figure 4.1**). Combining the conservation and chemical shift data does indicate that there may be two main regions of potential interaction, around the positively charged region near R45/K47 and on the end of the protein around E57 and T87. Interestingly K47 was one of the mutants which effected the function of CsgH while R45 did not have an effect. The regions near E57 and T87 have not been probed much by our mutagenesis, apart from the R25 mutant which did not have any significant effect. conservation of residues between CsgC and CsgH (ClustalW2 [**137**]) and between CsgH homologs (ConSurf [**203, 204, 205, 206**]) does not exactly match residues for which mutagenesis has been shown to effect CsgH activity, although there are certainly regions of the protein which are more conserved than others and these do show some relationship to our mapped chemical shifts.

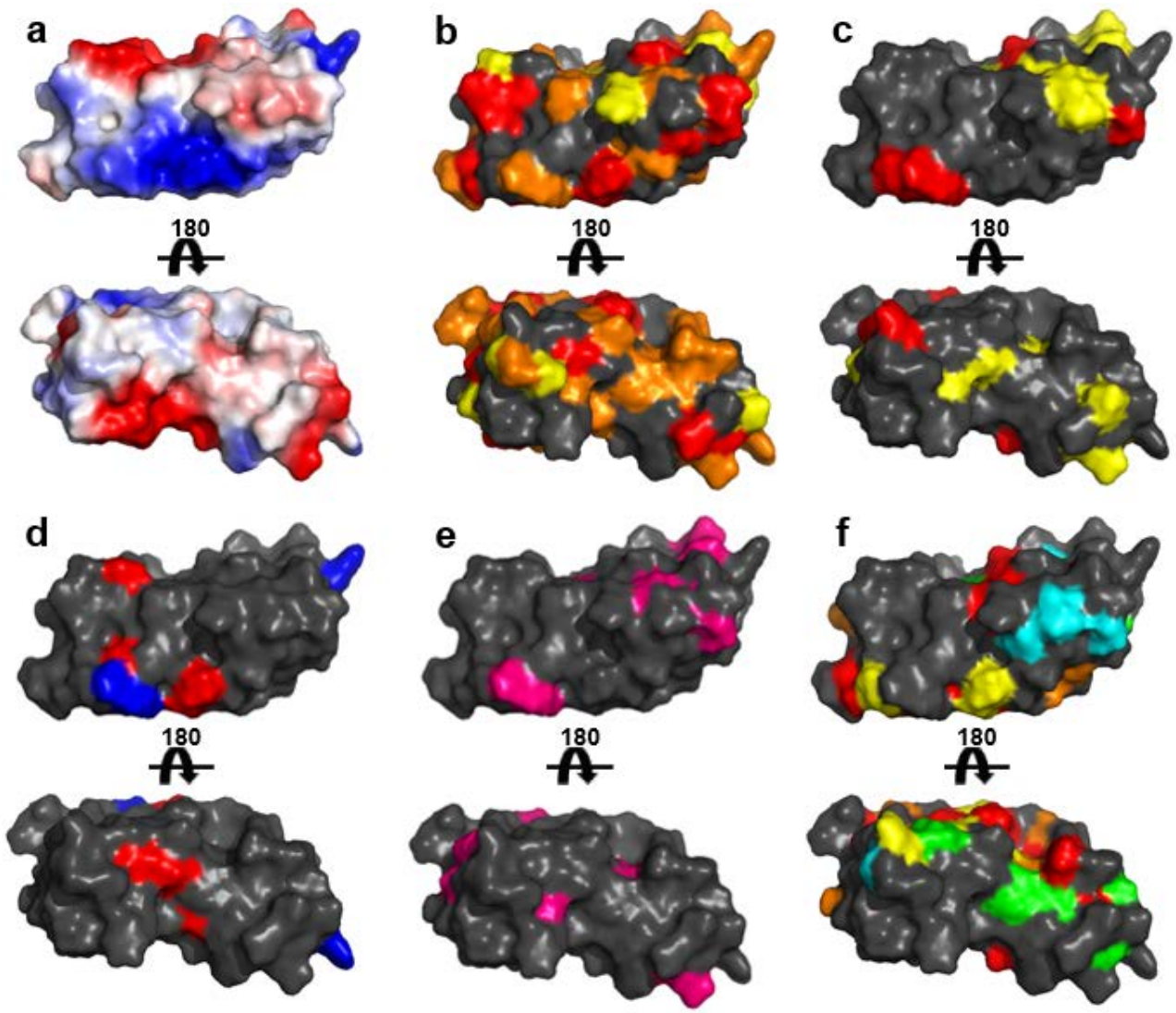


Figure 4.1 Conservation between CsgH and CsgC & suggestion of an interaction interface. These surface diagrams, showing either side of the structure of CsgH, have been mapped with a range of different data sets to illustrate potential regions of interest on the surface for interaction. **a**,CsgH surface coloured by charge with negative patches in red, positive patches in blue and neutral patches in white; **b**, CsgH surface coloured according to conservation between CsgC and CsgH based on clustalw2 alignment [137] with non-conserved regions in grey, fully conserved residues in red, residues which conserve size and charge in orange and residues which conserve either size or charge in yellow. **c**, CsgH surface coloured to show residues which show averaged chemical shift changes of residues in both experiments with CsgH-CTH and CsgH-NTH, in grey are residues showing no significant conserved shift, in yellow are residues which show a weak shift ($D<0.05$) and in red residues with stronger shifts ($D>0.05$). **d**, CsgH surface coloured to show mutants which inhibit the function of CsgH in red and those which did not in blue; all other residues are shown in grey. **e**,CsgH surface showing in pink residues which both show chemical shift changes in the presence of CsgA and are conserved between CsgH and CsgC, other residues are shown in grey **f**, CsgH surface showing conservation between CsgH homologs as calculated by Consurf [**203, 204, 205, 206**], in red are fully conserved residues (Conservation Score 9), the other residues are coloured in decreasing conservation, in orange (Conservation Score 8), in yellow (Conservation Score 7), in green (Conservation Score 6), in blue (Conservation Score 5), in grey are all residues with a Conservation Score lower than 5.

The mechanism by which CsgH and CsgC inhibit amyloid formation is currently unknown and it is interesting to speculate how these potent amyloid inhibitors may work in the light of structural and functional insights. Comparison of the cysteine positions in CsgC and CsgH indicate that the CxC motif, which had previously been highlighted in CsgC [**106**], does not have a role in the inhibition of amyloid since the cysteine residues are not conserved and the cysteines in CsgH appear to fulfil a structural role. It is true that CsgC and CsgH, as proteins with an IG-like fold, are structurally similar to many chaperones, and many chaperone proteins are known to inhibit amyloid formation such as DnaK, αβ-crystalin, Hsp33 and Spy [**207, 208**]. However the potency of these inhibitors is far lower than CsgH generally with equimolar or near equimolar ratios [**208**], suggesting the chaperones merely sequester free monomers preventing aggregation. Similarly charge has been shown to play an important role in inhibiting amyloid fibrillation, in Aβ amyloid mutants altering the charge of small hydrophilic proteins scMN and CB, the number of charges were shown to correlate with the efficiency of inhibition with the most positively charged protein showing the best inhibition (1:100) [**209**]. CsgH is also a small hydrophilic protein and conserved patches of negative and positive charge are prominent on both CsgC and CsgH and several charged residues appear to be important for function, however mutations of charged residues at other positions have no effect on efficiency indicating the position of the charge is important rather than overall charge alone. This fits with data collected for CsgC by Dr Jonathan Taylor which indicated that specific charges in CsgC are important for function (**Personal Communication, Unpublished Data**). The NMR experiments conducted on CsgH in the presence and absence of CsgA indicate that there is a transient interaction between the proteins, but that the interaction extends to end of the protein around the protruding R25 as well as the charged patches which were the primary targets of our mutagenesis. This could suggest a mode of action where the electrostatics on the surface of CsgH/CsgC guide the CsgA as it moves in to interact, the CsgH/CsgC molecule then interacts more tightly. There is no decline in signal from CsgH in the presence of CsgA suggesting that CsgH is not depleted from the solution as seen for chaperone inhibitors such as DNAJB6 [**198**].

The high potency of CsgH/CsgC inhibition, capable of inhibiting amyloid formation at ratios of 1:1000, indicates that the proteins must have an efficient method of inhibiting the fibre formation beyond simply sequestering CsgA molecules via a chaperone-like activity, since only ~0.1 % of the CsgA monomers could be directly bound by CsgH at a time. In principle this could be explained in several different ways: CsgH could act by capping the ends of nascent fibres inhibiting fibre extension as seen in Cytochalasin inhibition of actin [**210**]; CsgH could act on polymerised CsgA to stabilise the fibre to prevent new ends from forming or to destabilise the fibre breaking it down; CsgH could interfere with fibre nuclei, interfering with the initial templating steps of fibre formation; CsgH could also act to prevent the initial steps of fibre formation by sequestering or

altering CsgA which is in an intermediate state on the path to amyloid formation possibly converting CsgA to a transiently inactive form. Interestingly the DNAJB6 chaperon from the Hsp40 family was shown to inhibit amyloid formation at substoichiometric ratios by interacting with oligomers [**211**]. It has been suggested that intrinsically disordered proteins that convert to amyloid have an inverted energy landscape where they are disordered at the energy minimimum sample a large range of conformations in their natively disordered state with more ordered states occupying a higher energy level [**213**], this could be the case for CsgA with the monomers interchanging between a broad range of non-amyloidogenic structures, a smaller subset of the protein may then sample the amyloid forming conformation seeding the nucleation-precipitation of the fibres; If CsgH acts on this partially structured subset of CsgA and the population of this group is small enough it may explain the ability of CsgH to inhibit despite a very low stoichiometric ratio. Although the ThT assay is a well established method for checking for amyloid fibre formation it was still important to show that fibre formation is in fact inhibited as it could conceivably be possible that CsgC/H has an effect on the fluorescence of the fibres themselves. Dr Jonathan Taylor conducted electron microscopy experiments showing that fibre formation was retarded by CsgC, and that the final fibres appear similar in the presence or absence of the inhibitor after 22 h (**Figure 4.2**). It is noticeable that the formation of small almost spherical objects, possibly large oligomers of CsgA is significantly delayed in the presence of CsgC (up to 5 hours). The delay in the formation of these aggregates, possibly nuclei for fibre formation may suggest that CsgC/H acts early on in the pathway to amyloid formation.

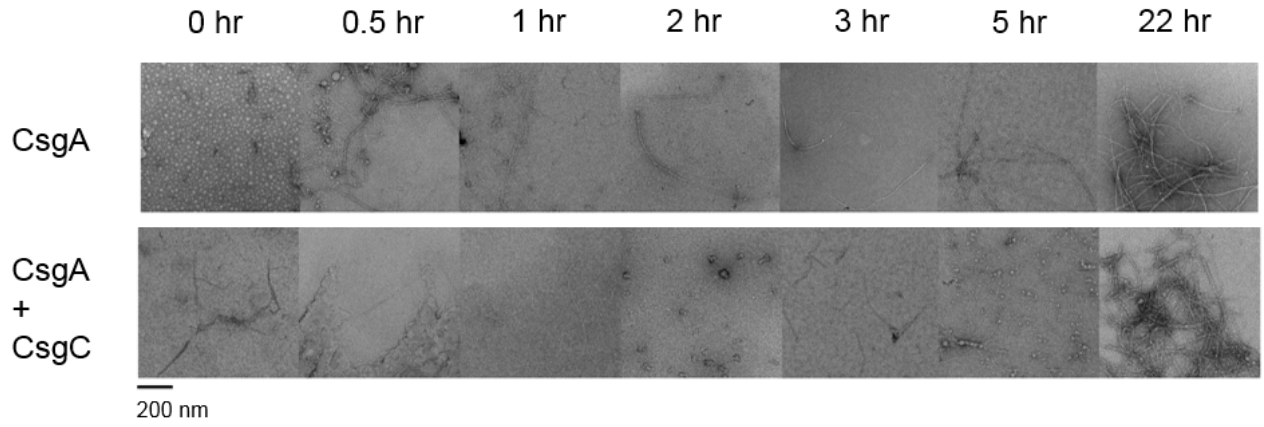


Figure 4.2 Electron Microscopy Images of CsgA aggregation over time in the presence and absence of CsgC. Samples of CsgA (35 uM) were purified and incubated at room temperature in the presence and absence of CsgC (1:200 CsgC:CsgA) for 22 h and small aliquots were used to produce EM grids. The grids were analysed and representiative images were selected for comparison. The scale bar indicates 200 nm. These images were kindly provided by Dr Jonathan Taylor who conducted the experiments (**Unpublished data**).

There are several methods which could be used to further probe the molecular mechanism by which CsgH inhibits amyloid. Although many mutants have been tested, further mutants could be made to probe the surface of CsgH more extensively. The conserved residues S85, R74, Q58, T16, Q3 and P94; the residues on the surface near interesting charged regions E5, E7, E57, E95, T87 and D80 as well as the residues which show distinct chemical shifts D68, G84 and T43 could all be mutated to test the effect on the protein function (**Figure 4.3**). E5 and E7 would be particularly interesting as they correspond to the largest negatively charged patch on the protein surface and are adjacent to the triple mutant which had an effect on protein function. The mutations to Q3, Q58, E57, S85 and G84 would probe a region of the protein where several residues showing chemical shifts and where there is conservation between CsgH and CsgC, which may therefore constitute and interaction interface. The NMR experiment combining CsgA with CsgH-NTH could be repeated with a higher concentration of CsgA to enhance the magnitude of chemical shifts which should improve the accuracy and detail of our chemical shift mapping. Alternatively it may be useful to change the approach of our NMR experiments and probe the interface using paramagnetic relaxation enhancement, a construct of CsgA tagged with paramagnetic species could be combined with labelled CsgH and distance constraints can then be derived from the changes in relaxation.

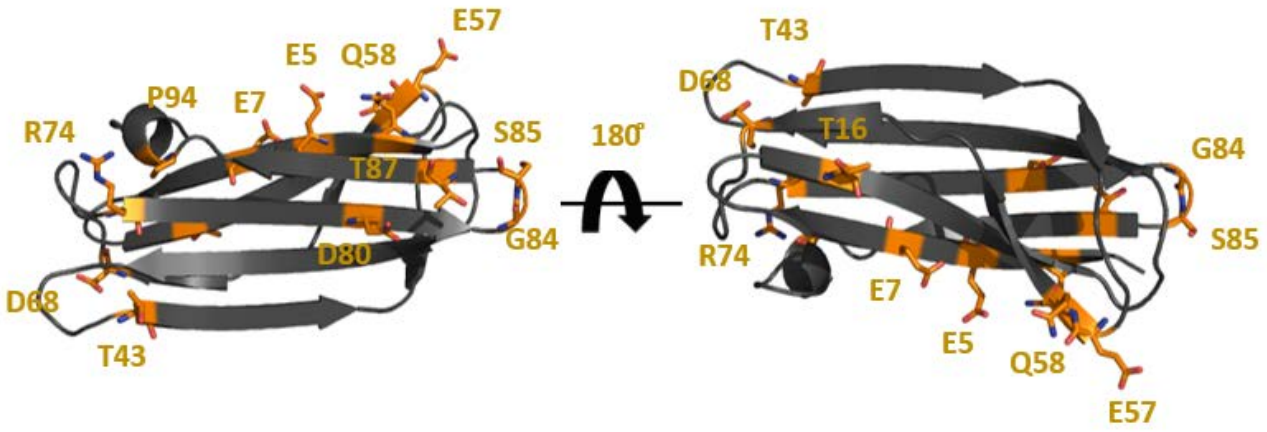


Figure 4.3 Suggested Further Mutants of CsgH. Cartoon structure of CsgH with residues suggested for future mutation shown as sticks, with carbon atoms coloured orange, nitrogen atoms blue and oxygen atoms red and labelled with the residue number and type. The focus of these mutants are charged residues not tested so far, conserved residues and residues which show or are near regions which have been shown to exhibit chemical shifts in the presence of CsgA.

Although the potency of CsgH at inhibiting amyloid *in vitro* is interesting in its own right, to enhance our understanding of controlling amyloid and despite and undoubtedly a useful function for the protein in nature, this may not be its only role *in vivo*. In the curli system the CsgA molecules must be transported through the periplasm, where the inhibition of amyloid is certainly useful, it must be directed to the CsgG pore where it is transported through to form amyloid external to the cell. CsgH acts to keep CsgA in a transport competent state, keeping the CsgA monomeric, it may also perform a role both in chaperoning the CsgA molecules to CsgG and/or it may alter the properties of the resulting fibre, as we cannot tell from our experiment what the effect of exposure to CsgH is on the atomic structure of the fibre, both of these areas are interesting for future study. The fibre structure in the presence and absence of CsgH/CsgC could be probed using solid state NMR or electron microscopy and mutants of CsgC could be tested in the context of the Curli system to see the effect *in vivo*. Although interesting insights have been made into the function of CsgH/CsgC amyloid inhibition there are many details still to be elucidated; with the structure of CsgH solved and progress made towards our understanding of the *Pseudomonas* amyloid we are hopeful that that the molecular mechanisms underlying control of functional amyloid in bacteria may be revealed in the near future.

References

[1]: Virchow, R. (1855). Ueber den Gang der Amyloiden Degeneration. *Arch. Pathol. Anat. Physiol.* **8**, 364-369.

[2]: Cohen, A.S. (1986) General introduction and a brief history of the amyloid fibril. In Marrink, J., and Van Rijswijk, M.H. (Eds.), *Amyloidosis*, Nijhoff, Dordrecht, 3–19.

[3] Elghetany MT, Saleem A. (1988) Methods for Staining Amyloid in Tissues: A Review, *Biotechnic and Histochemistry* **63** (4), 201-212

[4] Friedreich, N. and Kekulé A. (1859) Zur Amyloidfrage. *Virchows Arch Pathol Anat* **16**, 50–65.

[5] Bolder S.G., Sagis L.M., Venema P. & van der Linden E. (2007) Thioflavin T and birefringence assays to determine the conversion of proteins into fibrils. *Langmuir* **23** (8), 4144–4147

[6] Hawe A., Sutter M. and Jiskoot W. (2008) Biochemical basis of ThT amyloid fluorescence Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* **25** (7), 1487–1499.

[7] Divry P. *(1927) Etude histochimique des plaques seniles J. Neurol. Psychiatry **27, 643-657***

[8] Howie A.J., Brewe D.B., Howell D. and Jones A.P. (2008) Physical basis of colors seen in Congo red-stained amyloid in polarized light Laboratory Investigation 88, 232-242

[9] Ghafoor A., Hay I.D and Rehm B.H.A. (2011) Role of Exopolysaccharides in Pseudomonas aeruginosa Biofilm Formation and Architecture *Applied and Environmental Microbiology* **77** (15), 5238-5246

[10] M. Fändrich (2007) On the structural definition of amyloid fibrils and other polypeptide aggregates *Cellular and Molecular Life Sciences* **64** (16), 2066-2078

[11] del Mercato L.L., Pompa P.P., Maruccio G., Antonio Della Torre A.D., Sabella S., Tamburro A.M., Cingolani R. and Ross Rinaldi R. (2007) Charge transport and intrinsic fluorescence in amyloid-like fibrils *PNAS* **104** (46), 18019-18024

[12] Cohen, A.S. and Calkins, E. (1959) Electron microscopic observations on a fibrous component in amyloid of diverse origins. *Nature*, **183**, 1202-1203.

[13] Eanes, E.D. and Glenner, G.G. (1968) X-ray diffraction studies on amyloid filaments. *Journal of Histochemistry and Cytochemistry*, **16** (11), 673-677.

[14] Sunde M., Serpell L.C., Bartlam M., Fraser P.E., Pepys M.B. and Blake C.C. (1997) Common Core Structure of Amyloid Fibrils by Synchrotron X-ray Diffraction *J. Mol. Biol.* **273** (3), 729-739

[15] Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426**, 884–90

[16] Gazit, E. (2002) The "correctly folded" state of proteins: is it a metastable state. *Angew. Chem*. Int. Ed. WILEY-VCH Verlag GmbH, Weinheim **41** (2)**,** 257–259

[17] Baldwin A.J., Knowles T.P.J., Tartaglia G.G., Fitzpatrick A.W., Devlin G.L., Shammas S.L., Waudby C.A., Mossuto M.F., Meehan S., Gras S.L., Christodoulou J., Anthony-Cahill S.J., Barker P.D., Vendruscolo M., and Dobson C.M. (2011) Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, (36) 14160–14163.

[18] Knowles T.P.J., Vendruscolo M. and Dobson C.M. (2014) The amyloid state and its association with protein misfolding diseases *Nature Review* **15**, 384-396

[19] Scherzinger E., Lurz R. Turmaine M., Mangiarini L., Hollenbach B., Hasenbank R., Bates G.P., Davies SW, Lehrach H and Wanker E.E. (1997) Huntingtin-Encoded Polyglutamine Expansions Form Amyloid-like Protein Aggregates In Vitro and In Vivo *Cell* **90** (3), 549-558

[20] Tracz S.M., Abedini A., Driscoll M. and Raleigh D.P. (2004) Role of Aromatic Interactions in Amyloid Formation by Peptides Derived from Human Amylin *Biochemistry*, **43** (50), 15901–15908

[21] Hilbich C., Kisters-Woike B., Reed J., Masters C.L. and Beyreuther K. (1992) Substitutions of hydrophobic amino acids reduce the amyloidogenicity of Alzheimer's disease βA4 peptides *Journal of Molecular Biology* **228** (2), 460–473

[22] Chapman M.R., Robinson L.S., Pinkner J.S., Roth R., Heuser J., .Hammar M, Normark S. and Hultgren S.J. (2002) Role of Escherichia coli Curli Operons in Directing Amyloid Fiber Formation **295** (5556), 851-855

[23] Wanga X. and Chapman M.R. (2008) Curli provide the template for understanding controlled amyloid propagation *Prion* **2** (2), 57-60

[24] Thomas PJ, Qu B.H. and Pedersen P.L. (1995) Defective protein folding as a basis of human disease *Trend in biochemical science* **20** (11), 456-459

[25] Kirkitadze, M.D., Bitan, G. and Teplow, D.B. (2002) Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: The emerging role of oligomeric assemblies. *J. Neurosci. Res.* **69** (5), 567–577.

[26] Klein WL, Krafft GA, Finch CE. (2001). Targeting small Abeta oligomers: the solution to an Alzheimer's disease conundrum? *Trends Neurosci* **24** (4), 219 –224.

[27] McGowan D.P. van Roon-Mom W., Holloway H., Bates G.P., Mangiarini L., Cooper G.J., Faull R.L. and Snell R.G. (2000) Amyloid-like inclusions in Huntington's disease *Neuroscience* **100**(4),677-680

[28] Irwin D.J., Lee V.M. and Trojanowski J.Q. (2013) Parkinson's disease dementia: convergence of α-synuclein, tau and amyloid-β pathologies Nat Rev Neuroscience **14**(9), 626-636

[29] Mucchiano G.I., Jonasson L., Häggqvist B., Einarsson E. and Westermark P. (2001) Apolipoprotein A-I-derived amyloid in atherosclerosis. Its association with plasma levels of apolipoprotein A-I and cholesterol. *Am J Clin Pathol* **115**(2),298-303

[30] Andrade C. (2007) A peculiar form of peripheral neuropathy *Acta Psychiatrica* **26**(3-4),251-257

[31] Meretoja J (1969) Familial systemic paramyloidosis with lattice dystrophy of the cornea, progressive cranial neuropathy, skin changes and various internal symptoms: a previously unrecognized heritable syndrome. *Ann Clin Res* **1**(4):314-324

[32] Kahn S.E., Andrikopoulos S. and Verchere C.B. (1999) Islet amyloid, a long-recognized but underappreciated pathological feature of type 2 diabetes. *Diabetes* **48**(2): 241–253

[33] Benson M.D. and Cohen A.S. (1979) Serum amyloid-A protein in amyloidosis, rheumatic and neoplastic diseases. *Arthritis Rheum* **22**(1) 36-42.

[34] Pepys MB1, Hawkins PN, Booth DR, Vigushin DM, Tennent GA, Soutar AK, Totty N, Nguyen O, Blake CC, Terry CJ, Feest T.J., Zalin A.M. and Hsuan J.J. (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis *Nature* **362**(6420), 553 – 557

[35] Lobo A., Launer L.J., Fratiglioni L., Andersen K., Di Carlo A., Breteler M.M., Copeland J.R., Dartigues J.F., Jagger C., Martinez-Lage J., Soininen H. and Hofman A. (2000) Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. *Neurology. **54** (11 Suppl 5), S4-9*

[36] Hoyert D.L. and Xu J. (2012) Deaths: Preliminary Data for 2011, *National Vital Statistics Reports* 61 (6) Hyattsville, MD: National Center for Health Statistics. 2012.

[37] Glenner, G.G. and Wong, C.W. (1984) Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **120**(3),885–890

[38] Hardy, J.A. and Higgins, G.A. (1992) Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**(5054), 184–185

[39] Alonso A., Zaidi T., Novak M., Grundke-Iqbal I. and Iqbal K. (2001) Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments *PNAS* **98**(12),6923-8

[40] Nordberg A. (2001) Nicotinic receptor abnormalities of Alzheimer's disease: therapeutic implications. *Biol Psychiatry* **49**(3),200-210.

[41] Ball, M. J. (1982). Limbic predilection in Alzheimer dementia: is reactivated herpesvirus involved? *Can. J. Neurol. Sci.* **9**(3), 303–306.

[42] Tanzi R.E. (2012) The genetics of Alzheimer disease *Cold Spring Harb. Perspect. Med.* **2** (10), a006296.

[43] Perez-Nievas B.G., Stein T.D., Tai H.C., Dols-Icardo O., Scotton TC, Barroeta-Espar I, Fernandez-Carballo L, de Munain EL, Perez J, Marquie M, Serrano-Pozo A,Frosch MP, Lowe V, Parisi JE, Petersen RC, Ikonomovic MD, López OL, Klunk W, Hyman BT, and Gómez-Isla T (2013) Dissecting phenotypic traits linked to human resilience to Alzheimer's pathology. *Brain.* **136**(8), 2510-26.

[44] Giannakopoulos P., Herrmann F.R., Bussière T**.**, Bouras C**.**, Kövari E**.**, Perl D.P., Morrison J.H., Gold G. and Hof P.R. (2003) Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology*. **60**(9), 1495-500

[45] Ingelsson M., Fukumoto H., Newell K.L., Growdon J.H., Hedley-Whyte E.T., Frosch M.P., Albert M.S., Hyman B.T. and Irizarry M.C. (2004). Early Aβ accumulation and progressive synaptic loss, gliosis, and tangle formation in AD brain. *Neurology* **62**(6): 925–931

[46] Holmes C., Boche D., Wilkinson D., Yadegarfar G., Hopkins V., Bayer A., Jones R.W., Bullock R., Love S., Neal J.W., Zotova E. and Nicoll J.A. (2008) Long-term Effects of Abeta42 Immunisation in Alzheimer's Disease: Follow-up of a Randomised, Placebo-controlled Phase I Trial. Long-term Effects of Abeta42 Immunisation in Alzheimer's Disease: Follow-up of a Randomised, Placebo-controlled Phase I Trial. *Lancet* 372(9634),216-23

[47] Arriagada P.V., Growdon J.H., Hedley-Whyte E.T. and Hyman B.T. (1992) Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology*. **42**(3), 631-9.

[48] Duyckaerts C., Colle M.A., Dessi F., Grignon Y, Piette Y. and Hauw J.J. (1998) The progression of the lesions in Alzheimer disease: insights from a prospective clinicopathological study. *J Neural Transm Suppl* **53**, 119–126

[49] Goedert M. and Jakes R. (2005) Mutations causing neurodegenerative tauopathies. *Biochim Biophys Acta.* **1739**(2-3):240-50.

[50] Poorkaj, P., Bird, T.D., Wijsman, E., Nemens, E., Garruto, R.M., Anderson, L., Andreadis, A., Wiederholt, W.C., Raskind, M. and Schellenberg, G.D. (1998), Tau is a candidate gene for chromosome 17 frontotemporal dementia. Ann Neurol., 43(6),815–825.

[51] Jin M, Shepardson N, Yang T, Chen G, Walsh D and Selkoe DJ. (2011) Soluble amyloid beta-protein dimers isolated from Alzheimer cortex directly induce Tau hyperphosphorylation and neuritic degeneration. *PNAS* **108** (14), 5819–5824.

[52] Roberson E.D., Scearce-Levie K., Palop J.J., Yan F., Cheng I.H., Wu T., Gerstein H., Yu G.Q. and Mucke L. (2007) Reducing endogenous tau ameliorates amyloid beta-induced deficits in an Alzheimer's disease mouse model. *Science*. **316**(5825),750–754

[53] Ittner L.M. and Gotz J. (2011) Amyloid-beta and tau--a toxic pas de deux in Alzheimer's disease. *Nat Rev Neurosci* **12**(2),65–72

[54] Lambert M.P., Barlow A.K., Chromy B.A., Edwards C., Freed R., Liosatos M., Morgan T.E., Rozovsky I., Trommer B., Viola K.L., Wals P., Zhang C., Finch C.E., Krafft G.A. and Klein W.L. (1998) Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins *PNAS* **95**(11),6448-6453

[55] Lacor P.N., Buniel M.C., Chang L., Fernandez S.J., Gong Y., Viola K.L., Lambert M.P., Velasco P.T., Bigio E.H., Finch C.E., Krafft G.A. and Klein W.L. (2004) Synaptic targeting by Alzheimer's-related amyloid beta oligomers *J Neurosci* **24**(45),10191–10200

[56] Laurén J., Gimbel D.A., Nygaard H.B., Gilbert J.W. and Strittmatter S.M. (2009) Cellular Prion Protein Mediates Impairment of Synaptic Plasticity by Amyloid-β Oligomers Nature **457**(7233),1128-32

[57] Wisniewski H.M., Wen G.Y., and Kim K.S. (1989) Comparison of four staining methods on the detection of neuritic plaques Acta Neuropathol **78**,22-27

[58] Prusiner, S.B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science* **216**(4542),136–144.

[59] Prusiner, S.B. (1998). Prions. *PNAS* **95**(23),13363–13383

[60] Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ and Cohen FE (1993) Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *PNAS* **90**(23)10962-10966

[61] M. W. Head and J. W. Ironside (2012) Review: Creutzfeldt–Jakob disease: prion protein type, disease phenotype and agent strain *Neuropathology and Applied Neurobiology* **38**(4), 296–310

[62] Manuelidis L., Yu Z.X., Barquero N., Banquero N. and Mullins B. (2007). Cells infected with scrapie and Creutzfeldt-Jakob disease agents produce intracellular 25-nm virus-like particles. *PNAS* **104**(6),1965–70.

[63] Biljan I., Giachin G., Ilc G., Zhukov I., Plavec J. and Legname, G. (2012) Structural basis for the protective effect of the human prion protein carrying the dominant negative E219K polymorphism *Biochem J*, **446** (2), 243-251

[64] Govaerts C., Wille H., Prusiner S.B. and Cohen F.E.(2004) Evidence for assembly of prions with left-handed-helices into trimers *PNAS* **101** (22), 8342–8347

[65] Desvaux M., Hebraud M., Talon R. and Henderson I.R. (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue, *Trends Microbiol*. **17**(4), 139–145.

[66] Olsén A., Jonsson A. and Normark S. (1989) Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. *Nature.* **338**(6217):652-5.

[67] Claessen D., Rink R., de Jong W., Siebring J., de Vreugd P., Boersma F.G., Dijkhuizen L., Wosten H.A.. (2003) A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in Streptomyces coelicolor by forming amyloid-like fibrils. Genes Dev. **17**(14):1714-26

[68] Romero D.,Aguilar C, Losick R, and Kolter R (2010) Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. *PNAS* **107**(5): 2230–2234.

[69] Larsen, P., Nielsen, J. L., Dueholm, M. S., Wetzel, R., Otzen, D. and Nielsen, P. H. (2007), Amyloid adhesins are abundant in natural biofilms. *Environmental Microbiology*, **9**(12): 3077–3090.

[70] Gophna U., Barley M., Seijffers R., Oelschlager T.A., Hacker J. and Ron E.Z. (2001) Curli fibers mediate internalization of Escherichia coli by eukaryotic cells. Infect. Immun. **69**(4), 2659-65

[71] Low A., Chandrashekaran I.R., Adda C.G., Yao S., Sabo J.K. amd Zhang X. (2007) Merozoite surface protein 2 of Plasmodium falciparum: expression, structure, dynamics, and fibril formation of the conserved N-terminal domain. *Biopolymers* **87**(1),12-22.

[72] Olsén A, Jonsson A and Normark S (1989) Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. *Nature* **338**(6217),652-5.

[73] Olsén A, Arnqvist A, Hammar M, Sukupolvi S and Normark S (1993) The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA, the subunit gene of fibronectin-binding curli in Escherichia coli. *Mol Microbiol* **7**(4),523-36.

[74] Nguyen P.Q., Botyanszki Z., Tay P.K. and Joshi N.S. (2014) Programmable biofilm-based materials from engineered curli nanofibers *Nat Commun.* **5**,4945

[75] Van Gerven N., Goyal P., Vandenbussche G., De Kerpel M., Jonckheere W., De Greve H. and Remaut H. (2014) Secretion and functional display of fusion proteins through the curli biogenesis pathway *Mol Microbiol* **91**(5),1022-35

[76] Wingender, J., Neu, T. & Flemming, H.-C. (1999) Microbial Extracellular Polymeric Substances. Wingender, J., Neu, T. & Flemming, H.-C. (Eds.) Springer, Heidelberg 1–19

[77] Falsetta M.L., Klein M.I., Colonne P.M., Scott-Anne K., Gregoire S., Pai C.H., Gonzalez-Begne M., Watson G., Krysan D.J., Bowen W.H. and Koo H (2014) Symbiotic relationship between Streptococcus mutans and Candida albicans synergizes virulence of plaque biofilms in vivo. *Infect. Immun.* **82**(5), 1968-81

[78] Tyson G.W., Chapman J., Hugenholtz P., Allen E.E, Ram R.J., Richardson P.M., Solovyev V.V, Rubin E.M., Rokhsar D.S. and Banfield J.F.(2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428**, 37–43

[79] Lawrence J.R., Korber D.R., Hoyle B.D., Costerton J.W. and Caldwell D.E. (1991) Optical sectioning of microbial biofilms *J. Bacteriol* **173**(20),6558-6

[80] Costerton J.W., Lewandowski Z., Caldwell D.E., Korber D.R., and Lappin-Scott H.M. (1995) Microbial Biofilms *Annu. Rev. Microbiol.* **49**,711-745

[81] Costerton, J. W., Stewart, P. S. and Greenberg, E. P., (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284**(5418), 1318–1322.

[82] Nickel, J.C., Ruseska, I., Costerton, J.W. 1985 Tobramycin resistance of cells of Pseudomonas aeruginosa growing as a biofilm on urinary catheter material. *Antimicrob. Agents Chemother*. **27**(4):619- 24

[83] Mulcahy, H., Charron-Mazenod, L. and Lewenza, S. (2008) Extracellular DNA chelates and induces antibiotic resistance in *Pseudomonas aeruginosa* biofilms. *PLoS Pathog* **4**(11): e1000213

[84] Sutherland, I.W. (2001) Biofilm exopolysaccharides: a strong and sticky framework. *Microbiology* **147**(pt 1), 3-9

[85] Flemming H.C. and Wingender J. (2010) The biofilm matrix *Nature Reviews Microbiology* **8**,623-633

[86] Flemming, H.C. and Wingender, J. (2001) Relevance of microbial extracellular polymeric substances (EPSs) – Part I: Structural and ecological aspects. *Water Sci. Technol.* **43**(6), 1-8.

[87] Mann, E.E. and Wozniak,D.J. (2011) Pseudomonas biofilm matrix composition and niche biology. *FEMS MicrobiolRev*. **36**(4), 893-916

[88] Saldaña Z., Xicohtencatl-Cortes J., Avelino F., Phillips A.D., Kaper J.B., Puente J.L., Girón J.A. (2009) Synergistic role of curli and cellulose in cell adherence and biofilm formation of attaching and effacing Escherichia coli and identification of Fis as a negative regulator of curli. *Environ Microbiol.* **11**(4):992-1006

[89] Jakubovics N.S, Shields RC, Rajarajan N, Burgess JG (2013). Life after death: the critical role of extracellular DNA in microbial biofilms. *Lett. Appl. Microbiol.* **57**(6): 467–75.

[90] McCrate O.A., Zhou X., Reichhardt C., Cegelski L. (2013) Sum of the parts: Composition and architecture of the bacterial extracellular matrix. *J Mol Biol.* **425**(22):4286-94.

[91] Stoodley P., Sauer K., Davies D.G., Costerton J.W. (2002) Biofilms as complex differentiated communities. *Annu Rev Microbiol* **56:** 187–209.

[92] Dueholm M.S., Petersen S.V., Sønderkær, M., Larsen P., Christiansen G., Hein, K.L., Enghild, J.J., Nielsen J.L., Nielsen P.H. and Otzen, D.E. (2010) Functional amyloid in Pseudomonas *Molecular Microbiology* 77(4), 1009-1020

[93] Karmali M.A., Steele B.T., Petric M. and Lim C. (1983) Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing Escherichia coli in stools. *Lancet.* **1**(8325):619-20

[94] O'Toole, G.A., Kaplan, H. and Kolter, R. (2000). Biofilm formation as microbial development. *Annu Rev Microbiol* **54**, 49–79.

[95] Murray CJL and Lopez AD (Eds.) (1996) *The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020.* Harvard University Press, Cambridge.

[96] Esrey SA, Potash JB, Roberts L, Shiff C (1991) Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization*, **69**(5):609–621.

[97] Serra D.O., Richter A.M., Klauck G., Mika F. and Hengge R. (2013) Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm *MBio*, **4**(2), e00103–e00113

[98] Olsen A., Arnqvist A., Hammar M., Normark S. (1993) Environmental regulation of curli production in Escherichia coli I*nfect. Agents Dis.*, **2**(4), 272–274

[99] Prigent-Combaret C., Vidal O., Dorel C. and Lejeune P. Abiotic surface sensing and biofilm-dependent regulation of gene expression in Escherichia coli *J. Bacteriol*., **181**(19), 5993–6002

[100] Hammar, M., Arnqvist, A., Bian, Z., Olsen, A. & Normark, S. (1995) Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12 *Mol Microbiol* **18**(4), 661–670.

[101] Dudin O, Geiselmann J, Ogasawara H, Ishihama A, Lacour S (2014) Repression of flagellar genes in exponential phase by CsgD and CpxR, two crucial modulators of Escherichia coli biofilm formation *J Bacteriol* **196**(3), 707–15

[102] Hammar M, Bian Z, Normark S (1996) Nucleator-dependent intercellular assembly of adhesive curli organelles in Escherichia coli *PNAS* **93**(13):6562–6566

[103] Nenninger A.A., Robinson L.S., Hammer N.D., Epstein E.A., Badtke M.P., Hultgren S.J. and Chapman M.R. (2011) CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation *Mol. Microbiol.* **81**(2), 486-99

[104] Nenninger A.A., Robinson L.S. and Hultgren S.J. (2009) Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF *PNAS* **106**(3),900-905

[105] Goyal P., Krasteva P.V., Gerven, N.V., Gubellini F., Van den Broeck I., Troupiotis-Tsailaki A., Jonckheere W., Pehau-Anaudet G., Pinker J.S., Chapman M.R., Hultgren S.J., Howorka S., Fronzes R and Remaut H. (2014) Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG *Nature* **516**, 250–253

[106] Taylor J.D., Zhou Y., Salgado P.S., Patwardhan A., McGuffie M,m Pape T., Grabe G., Ashman E., Constable S.C. Simpson P.J., Lee W.C., Cota E., Chapman M.R. and Matthews S.J. (2011) Atomic resolution insights into curli fiber biogenesis. *Structure* **19**(9-16), 1307-1316

[107] Evans M.L., Chorell E., Taylor J.D., Aden J., Gotheson A., Li F., Koch M., Sefer L., Matthews S.J., Wittung-Stafshede P., Almqvist F. and Chapman M.R. (2015) The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation *Molecular Cell* **57**(3), 1–11

[108] Dueholm M.S., Albertsen M., Otzen D. and Nielsen P.H. (2012). Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. *PLoS One* **7**(12):e51274

[109] Wang X., Hammer N.D. and Chapman M.R. (2008) The Molecular Basis of Functional Bacterial Amyloid Polymerization and Nucleation *J. Biol. Chem.* **283**(31),21530–21539

[110] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17),3389-3402.

[111] Cao B., Zhao Y., Kou, Y., Ni D., Zhang X.C. and Huang Y. (2014) Structure of the nonameric bacterial amyloid secretion channel *PNAS* **111**:E5439-5444

[112] Dueholm M.S., Nielsen S.B., Hein K.L., Nissen P., Chapman M., Christiansen G., Nielsen P.H. and Otzen, D.E. (2011) Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation, Biochemistry **50**(39), 8281–8290.

[113] Wang X., Smith D.R., Jones J.W. and Chapman M.R, (2007) In vitro polymerization of a functional Escherichia coli amyloid protein, *J. Biol. Chem*. **282**(6),3713–3719.

[114] Collinson, S.K., Parker, J.M., Hodges, R.S. and Kay, W.W. (1999) Structural predictions of AgfA, the insoluble fimbrial subunit of Salmonella thin aggregative fimbriae *J. Mol.Biol.* 290, 741–756

[115] Shewmaker F., McGlinchey R.P., Thurber K.R., McPhie P., Dyda F., Tycko R. and Wickner R.B. (2009) The functional curli amyloid is not based on in-register parallel beta-sheet structure *J. Biol. Chem* **284**(37),25065-25076.

[116] Wang X. and Chapman M.R. (2008) Sequence determinants of bacterial amyloid formation *J. Mol. Biol*. **380**(3),570–580.

[117] Wang X., Zhou Y., Ren J.J., Hammer N.D. and Chapman M.R. (2010) Gatekeeper residues in the major curlin subunit modulate bacterial amyloid fiber biogenesis, *PNAS* **107**(1) 163–168

[118] Hammer N.D., Schmidt J.C. and Chapman M.R. The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization *PNAS* 104(30), 12494–12499

[119] White A.P., Collinson S.K., Banser P.A., Gibson D.L., Paetzel M., Strynadka N.C. and Kay W.W. (2001) Structure and characterization of AgfB from Salmonella enteritidis thin aggregative fimbriae, *J. Mol. Biol.* **311**(4), 735–749.

[120] Hammer N.D, McGuffie B.A, Zhou Y., Badtke M.P. Reinke A.A., Brännström, K., Gestwicki J.E., Olofsson A., Almqvist F. and Chapman M.R. (2012) The C-terminal repeating units of CsgB direct bacterial functional amyloid nucleation. *Journal of Molecular Biology*, **422(**3), 376–389.

[121] Sievers F., Wilm A., Dineen D.G., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J.D. and Higgins D. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega *Molecular Systems Biology* **7**,539

[122] Taylor J.D and Matthews S.J. (2015) New insight into the molecular control of bacterial functional amyloids *Front Cell Infect Microbiol* **5**, 33.

[123] Robinson, L.S., Ashman, E.M., Hultgren, S.J. and Chapman, M.R. (2006) Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol Microbiol* **59**(3), 870-81

[124] Zhou Y., Smith D.R., Fugnagel D.A. and Chapman M.R. (2013) Experimental manipulation of the microbial functional amyloid called curli *Methods Mol Biol* **966**,53-75

[125] Evans M.L., Schmidt J.C., Ilbert M., Doyle S.M., Quan S., Bardwell J.C., Jakob U., Wickner S. and Chapman M.R. (2011) E. coli chaperones DnaK, Hsp33 and Spy inhibit bacterial functional amyloid assembly *Prion* **5**(4), 323-34.

[126] Jones D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**(2),195-202.

[127] Livermore D.M. (2002) Multiple mechanisms of antimicrobial resistance in Pseudomonas aeruginosa: our worst nightmare? *Clin Infect Dis.* **34**(5):634-40

[128] Oliver A., Cantón R., Campo P., Baquero F. and Blázquez J. (2000) High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science* **88**(5469),1251-4.

[129] Henrichfreise B., Wiegand I., Pfister W. and Wiedemann B. (2007) Resistance mechanisms of multiresistant Pseudomonas aeruginosa strains from Germany and correlation with hypermutation. *Antimicrob Agents Chemother.* **51**(11),4062-70

[130] Van Eldere J. (2003) Multicentre surveillance of *Pseudomonas aeruginosa* susceptibility patterns in nosocomial infections *J. Antimicrob. Chemother.* **51**(2),347-352.

[131] Neuhauser M.M., Weinstein R.A., Rydman R., Danziger L.H., Karam G. and Quinn J.P. (2003) Antibiotic resistance among gram-negative bacilli in US intensive care units: implications for fluoroquinolone use. *JAMA* **289**(7):885

[132] Hart C.A. and Winstanley C. (2002) Persistent and aggressive bacteria in the lungs of cystic fibrosis children. *Br Med Bull* **61**,81-96

[133] Hassett D., Cuppoletti J., Trapnell B., Lymar S., Rowe J., Yoon S., Hilliard G., Parvatiyar K., Kamani M., Wozniak D., Hwang S., McDermott T. and Ochsner U. (2002). Anaerobic metabolism and quorum sensing by Pseudomonas aeruginosa biofilms in chronically infected cystic fibrosis airways: rethinking antibiotic treatment strategies and drug targets *Adv Drug Deliv Rev* **54**(11),1425–1443.

[134] Espinosa-Urgel M., Salido A. and Ramos J.L. (2000) Genetic analysis of functions involved in adhesion of *Pseudomonas putida* to seeds. *J. Bacteriol.* **182**(9),2363–2369

[135] Haas D. and Défago G. (2005) Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat. Rev. Microbiol.* **3**(4):307–319.

[137] Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) ClustalW and ClustalX version 2.0 *Bioinformatics* **23**(21),2947-2948

[138] Dueholm M.S., Søndergaard M.T., Nilsson M., Christiansen G., Stensballe A et al. (2013) Expression of Fap amyloids in *Pseudomonas aeruginosa*, *P. fluorescens*, and *P. putida* results in aggregation and increased biofilm formation. *Microbiologyopen* **2**(3),365-82

[139] Håvarstein L.S., Diep D.B. and Nes I.F. (1995) A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export *Molecular Microbiology* **16**(2):229-240

[140] Bleves S., Viarre V., Salacha R., Michel G.P., Filloux A. and Voulhoux R. (2010) Protein secretion systems in *Pseudomonas aeruginosa*: a wealth of pathogenic weapons *Int. J. Med. Microbiol.* **300**(8), 534–543

[141] Dirix G., Monsieurs P., Dombrecht B., Daniels R., Marchal K., Vanderleyden J. and Michiels J. (2004) Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters *Peptides* **25**(9), 1425–1440

[142] Lecher J., Schwarz C.K., Stoldt M., Smits S.H., Willbold D. and Schmitt L. (2012) An RTX transporter tethers its unfolded substrate during secretion via a unique N-terminal domain *Structure* 20(10), 1778–1787

[143] Delaglio F., Grzesiek S., Vuister G.W., Zhu G., Pfeifer J. and Bax A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**(3), 277–293

[144] Ulrich E.L., Akutsu H., Doreleigers J.F., Harano Y., Ioannidis Y.E., Lin J., Livny M., Mading S., Maziuk D., Miller Z., Nakatani E., Schulte C.F., Tolmie D.E. Wenger K., Yao H. and Markley J.L. (2008) BioMagResBank *Nucleic Acids Research* **36**, D402-D408

[145] Jung Y.S. and Zweckstetter M. (2004) MARS—robust automatic backbone assignment of proteins. *J Biomol NMR* **30**:11–23

[146] Marchant J., Sawmynaden K., Saouros S., Simpson P. and Matthews S. (2008) Complete resonance assignment of the first and second apple domains of MIC4 from *Toxoplasma gondii*, using a new NMRView-based assignment aid. *Biomol. NMR Assign*. **2**(2), 119–121

[147] Shen Y., Cornilescu G. and Bax A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts *J. Biomol. NMR* **44**(4), 213-223

[148] Rieping W., Habeck M., Bardiaux B., Bernard A., Malliavin T.E. and Nilges M. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**(3), 381-382.

[149] Brunger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.S., Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T. and Warren G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D*. **54**(Pt 5):905–921.

[151] Chen V.B, Arendall W.B. 3rd, Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S. and Richardson D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crytsallogr* **66**(Pt 1), 12-21

[152] Davis I.W., Leaver-Fay A., Chen V.B., Block J.N., Kapral G.J. Wang X., Murray L.W. Arendall W.B. 3rd, Snoeyink J., Richardson J.S. and Richardson D.C. (2007) *Nucleic Acids Res* **35**(Web Server issue),W375-83

[153] Williamson M.P. Using chemical shift perturbation to characterise ligand binding (2013) *Prog. Nucl. Mag. Res.* **73**,1-16

[154] Caffrey, M., and V. Cherezov. (2009) Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**(5): 706-731.

[155] Kelley L.A., Mezulis S., Yates, C.M., Wass M.N. and Sternberg M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis *Nature Protocols* **10**, 845-858

[156] Yang J., Yan R., Roy A., Xu D, Poisson J. and Zhang Y. (2015) The I-TASSER Suite: Protein structure and function prediction. Nature Methods **12**:7-8

[157] Buchan D.W.A., Minneci F., Nugent T.C.O., Bryson K. and Jones D.T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench *Nucleic Acids Research* **41**(Web Server issue),W349-57

[158] Petersen T.N., Brunak S., von Heijne G. and Nielsen H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions *Nature Methods*, **8**:785-786

[159] Bagos P.G., Liakopoulos T.D., Spyropoulos I.C. and Hamodrakas S.J. (2004) PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res* **1**;32(Web Server issue):W400-4.

[160] Ward, J.J., McGuffin, L.J., Bryson K., Buxton, B.F. and Jones, D.T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**(13),2138-2139.

[161] Bryson, K. Cozzetto D. and Jones D.T. (2007) Computer-assisted protein domain boundary prediction using the DomPred server. *Current Protein and Peptide Science* **8**(2), 181-188

[162] Lin D.Y., Huang S. and Chen J. (2015) Crystal structures of a polypeptide processing and secretion transporter. *Nature* **523**, 425-430

[163] Buchanan S.K. (1999) β-Barrel proteins from bacterial outer membranes: structure, function and refolding *Current Opinion in Structural Biology* **9**(4),455-461

[174] Micsonai A., Wien F., Kernya L., Lee Y.H., Goto Y., Réfrégiers M. and Kardos J. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy *PNAS* **112**(24),E3095-E3103

[175] Krogh A., Larsson B., von Heijne G. and Sonnhammer E.L.L. (2001) Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes *J Mol Biol* **305**,567-580

[176] Terpe K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial system Appl Microbiol Biotechnol **72**(2),211–222

[177] Hautbergue G.M. and Golovanov A.P. (2008) Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt. *J Magn Reson.* **191**(2):335-9

[178] Tamm L.K., Hong H. and Liang B. (2004) Folding and assembly of β-barrel membrane proteins *Biochimica et Biophysica Acta - Biomembranes* **1666**(1-2),250-263

[179] Wimley W.C. (2003) The versatile beta-barrel membrane protein *Current Opinion in Structural Biology* **13**(4):404–411

[180] Mohammad M., Howard K.R. and Movileanu L. (2011) Redesign of a Plugged β-Barrel Membrane Protein *Journal of Biological Chemistry* **286**(10),8000-8013.

[181] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. and Bairoch A. (2005) Protein Identification and Analysis Tools on the ExPASy Server (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press, New York

[182] Tiburu EK, Moton DM, Lorigan GA. (2001) Development of magnetically aligned phospholipid bilayers in mixtures of palmitoylstearoylphosphatidylcholine and dihexanoylphosphatidylcholine by solid-state NMR spectroscopy. *Biochimica Et Biophysica Acta (BBA) - Biomembranes*. **1512**,206–214.

[183] Ujwal R. and Bowie J.U. (2011) Crystallizing membrane proteins using lipidic bicelles. *Methods*. **55**(4),337-41

[184] Faham S and Bowie JU. 2002 Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol*. **31,**1–6.7

[185] Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H.J., Kugn P., Weis W.I., Koblilka B.K. and Stevens R.C. (2007) High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein–Coupled Receptor. *Science*. **318**,1258–1265.

[186] Struyve M., Moons M. and Tommassen J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein *J Mol Biol* **218**(1),141-148

[187] Robert V., Volokhina E.B., Senf F., Bos M.P., Van Gelder P. and Tommassen J. (2006) Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol* **4**(11),e377

[188] Vranken W.F., Boucher W., Stevens T.J., Fogh R.H., Pajon A., Llinas M., Ulrich E.L., Markley J.L., Ionides J. and Laue E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**(4),687-96

[189] Cheung M.S., Maguire M.L., Stevens T.J. and Broadhurst R.W. (2010) DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance* **202**(2),223-233

[190] Johnson, B.A and Blevins R.A. (1994) NMR View: A computer program for the visualization and analysis of NMR data *Journal of Biomolecular NMR* **4**(5),603-614.

[191] Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V, (1963) Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**, 95–9

[192] Neal S., Nip A.M. Zhang H. and Wishart D.S. (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts *Journal of Biomolecular NMR* **26**(3),215-240

[193] Holm L. and Rosenström P. (2010)] Dali server: conservation mapping in 3D *Nucl. Acids Res.* **38**(Web Server issue),W545-549

[194] Tung C.H., Huang J.W. and Yang J.M. (2007) Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for fast protein structure database search *Genome Biology* **8(**3),R317

[195] Yang J.M and Tung C.H. (2006) Protein structure database search and evolutionary classification *Nucleic Acids Research* **34**(13),3646-59

[196] Miliara X., Garnett J.A., Tatsuta T., Ali F.A., Baldie H., Perez-Dorado I., Simpson P., Yague E., Langer T and Matthews S. (2015) Structural insight into the TRIAP1/PRELI-like domain family of mitrochondrial phospholipid transfer complexes. *EMBO* reports **16**, 824-835

[197] Iwata, K., Fujiwara, T., Matsuki, Y., Akutsu, H., Takahashi, S., Naiki, H. and Goto, Y. (2006) 3D structure of amyloid protofilaments of beta2-microglobulin fragment probed by solid-state NMR. *PNAS* **103**(48): 18119-18124

[198] Tycko R. (2006) Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q Rev Biophys* **39**(1),1-55.

[199] Van Melckebeke H., Wasmer C., Lange A., Ab E., Loquet A., Bockmann A. and Meier B.H. (2010) Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state NMR spectroscopy. *J.Am.Chem.Soc*. **132**(39),13765-13775

[200] Hendrickson W.A. and Ogata C.M. (1997) Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276**: 494–523

[201] Barthelmes K., Reynolds A.M., Peisach E., Jonker H.R.A., DeNunzio N.J., Allen K.N. Imperiali B. and Schwalbe H. (2011) Engineering Encodable Lanthanide-Binding Tags into Loop Regions of Proteins *J. Am. Chem. Soc.* **133**(4),808-819

[202] Matthew E., Ding F.X., Naider F and Dumont M.E. (2013) Functional fusions of T4 lysozyme in the third intracellular loop of a G protein-coupled receptor identified by a random screening approach in yeast *Protein Eng Des Sel* **26**(1),59-71

[203] Celniker G., Nimrod G., Ashkenazy H., Glaser F., Martz E., Mayrose I., Pupko T. and Ben-Tal N. (2013) ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function *Isr. J. Chem*. **53**(3-4),199-206

[204] Ashkenazy H., Erez E., Martz E., Pupko T. and Ben-Tal N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucl. Acids Res*. **38**(Web Server issue),W529-33

[205] Landau M., Mayrose I., Rosenberg Y., Glaser F., Martz E., Pupko T. and Ben-Tal N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucl. Acids Res*. **33**(Web Server issue)W299-302

[206] Glaser F., Pupko T., Paz I., Bell R.E., Bechor D., Martz E. and Ben-Tal N. (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **19**(1),163-164

[207] Shammas S.L., Waudby C.A. Wang S., Buell A.K., Knowles T.P.J., Ecroyd H., Welland M.E., Carver J.A., Dobson C.M. and Meehan S. (2011) Binding of the Molecular Chaperone αB-Crystallin to Aβ Amyloid Fibrils Inhibits Fibril Elongation *Biophys J* **101**(7),1681-1689

[208] Evans M.L., Schmidt J.C., Ilbert M., Doyle S.M., Quan S., Bardwell J.C., Jakob U., Wickner S. and Chapman M.R. (2011) E. coli chaperones DnaK, Hsp33 and Spy inhibit bacterial functional amyloid assembly. *Prion.* **5**(4),323-34

[209] Assarsson A., Hellstrand E., Cabaleiro-Lago C. and Linse S. (2014) Charge Dependent Retardation of Amyloid β Aggregation by Hydrophilic Proteins *ACS Chem Neurosci.* **5**(4), 266–274.

[210] Flanagan M.D. and Lin S. (1980) Cytochalasins block actin filament elongation by binding to high affinity sites associated with F-actin. *J Biol Chem* **255**,935-838

[211] Mansson C., Arosio P., Hussein R., Kampinga H.H., Hashem R.M., Boelens W.C., Dobson C.M., Knowles T.P., Linse S. and Emanuelsson C. (2014) Interaction of the molecular chaperon DNAJB6 with growing amyloid-beta 42 (Aβ42) aggregates leads to sub-stoichiometric inhibition of amyloid formation. *J. Biol Chem.* **289**(45),31066-76

[212] Ivanova M.I., Sawaya M.R., Gingery M., Attinger A. and Eisenberg D. (2004) An amyloid-forming segment of β2-microglobulin suggests a molecular model for the fibril *PNAS* **101**(29),10584–10589

[213] Granata D., Baftizadeh F., Habchi J., Galvagnion C., De Simone A., Camilloni C., Laio A. & Vendruscolo M. (2015) The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci Rep.* **26(**5),15449

[214] Perez-Iratxeta C., and Andrade-Navarro M.A. (2008) K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Structural Biology* **8**,25