THE UNIVERSITY *of* EDINBURGH

The architecture of the central region of factor H and its interaction with PspC of *S. pneumoniae*

Elisavet Makou

THE UNIVERSITY *of* EDINBURGH

PhD

2013

Abstract

The complement system is a major component of innate immunity and an effector of antibody-mediated immune responses. Unlike the other two activation pathways of the complement system, the alternative pathway is permanently switched on. Discrimination by complement between self and foreign is therefore achieved by selective protection of healthy host tissue and cells.

This study investigated the alternative pathway regulator factor H (FH), which is crucial for protection of self surfaces from complement. FH engages via its N- and C- terminal ends with activation-specific fragments of C3, C3b and C3d. The middle region of FH has no binding sites for complement components. It presumably ensures that the binding sites at either end of the extended and flexible FH molecule cooperate in recognizing C3b in fluid phase or on self surfaces, but not on foreign targets.

This study was aimed at achieving an atomic level understanding of the structure of the middle portion of FH, thereby testing hypotheses as to how it promotes the overall biological efficacy of the intact protein. High-resolution NMR-derived structures of two module pairs FH-10-11 and FH-11-12 were solved and combined with SAXS data to produce a model of FH-10-12. This was combined, *in silico*, with the previously solved FH-12-13 structure, then the model of FH-10-13 was used to revisit SAXS data for FH-10-15 and FH-8-15. A unique structure emerged, unlike any other encountered previously in the family of complement regulators, in which CCPs 13, 14 and 15 have a highly compacted organization that has repercussions for function.

While devoid of binding affinity for host ligands, this central region is a binding site for PspC, a virulence factor of *S. pneumoniae.* It has been speculated that the bacteria use this interaction to sequester FH in a conformation that resembles the one adopted by FH on self cells and makes it particularly good at regulating complement. *S*tructural and functional investigations of this interaction were performed to establish the molecular basis of the use of FH by this pathogen in order to avoid complement-mediated elimination. It was found that PspC and FH form a near-irreversible complex, while FH-8-15 binds PspC almost as tightly as intact protein. When bound to PspC, FH has a higher affinity for some of its targets, supporting the theory that this bacterial protein stabiles a particularly active conformation of the regulator.

Acknowledgements

The completion of my doctoral thesis would not have been possible without the help of a number of people.

First and foremost I would like to thank my supervisor, Prof. Paul N. Barlow for welcoming me into a stimulating and challenging academic environment. His knowledgeable and thoughtful guidance and sound advice and help, his constructive feedback and our extensive and productive discussions provided me with a strong foundation.

I would also like to express my gratitude to Dr Christoph Schmidt for his invaluable support and enjoyable company, as this made me feel a part of the academic community from the very beginning. His positive and respectful attitude, his willingness to spend as much time as was necessary to explain procedures and answer questions meticulously made working on my PhD a rewarding experience as well as a pleasant one.

Many thanks also go to Dr Andrew Herbert who was generous with his patience, help and advice, offering solutions at times when I felt overwhelmed. His wonderful sense of humor added the greatly needed element of fun in our lives.

Special thanks go to Nicky and Haris for their truly fun company and for being such supportive and encouraging friends during and beyond working hours.

I express my gratitude to Ilias and Dr Haydyn Mertens for their fundamental contribution to this thesis.

My further acknowledgments individually go to the following people: Carina, Dušan, Juraj, Janice, Mara, Matt, Stacey, Heather, Maria, Marie, Carla, John, Dinesh, Isabell, Patience, Liz, Vanesa, Janet.

I will also never forget my brother's enthusiastic support.

Last but by no means least, I am and always will be indebted to my parents for their support beyond measure and encouragement at all times, as well as for their financial provision that made it possible for me to complete my PhD.

Unless stated in the text, the work described in this thesis is my own work and has not been submitted for in whole or in part for a degree or other qualification at this or any other university.

Elisavet Makou

Contents

List of figures

List of tables

Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
AMD	age-related macular degeneration
ANA	anaphylatoxin domain
aHUS	Atypical haemolytic uraemic syndrome
AUC	analytical ultracentrifugation
BMG	buffered minimal glycerol
BMM	buffered minimal methanol
C4BP	C4b-binding protein
CA	cofactor activity
CCP	complement control protein
CNS	Crystallography and NMR system
CR1	complement receptor type 1
DAA	decay accelerating activity
DAF	decay accelerating factor
EM	transmission electron microscopy
FH	factor H
FHbp	factor H binding protein
FHL-1	factor H-like protein 1
FI	factor I
FID	free induction decay
GAG	glycosaminoglycan
GBM	glomerular basement membrane
HSQC	heteronuclear single quantum coherence
IMAC	immobilized metal ion affinity chromatography
K_D	equilibrium dissociation constant
LC-MS	liquid chromatography-mass spectrometry
MAC	membrane attack complex
MASP	mannan-binding lectin associated serine protease
MBL	mannan-binding lectin
MCP	membrane cofactor protein
MG	macroglobulin domain
MPGN	Membrano-proliferative glomerulonephritis
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
PAMP	pathogen-associated molecular pattern
PDB	protein data bank
PCR	polymerase chain reaction
PspA	pneumococcal surface protein A
PspC	pneumococcal surface protein C
RCA	regulator of complement activation
RMSD	root mean square deviation
RU	response unit

SAXS	small-angle X-ray scattering
SCR	short consensus repeat
SOC	super optimal with catabolite repression
SPR	surface plasmon resonance
TED	thioester-containing domain
TOCSY	total correlation spectroscopy
YPD	yeast extract peptone dextrose

“Πειρώ το μεν σώμα είναι φιλόπονον, την δε ψυχή φιλόσοφον”.

Isocrates, 436-338 B.C.
An ancient Athenian orator, rhetorician, and teacher.

CHAPTER 1

Introduction

1.1 The innate immune system

The immune system is a complex organization consisting of a network of molecules and cells that serves to protect the host against numerous infectious agents as well as other hazards. Historically, it has been divided into two major components: the innate (or non-specific) immune system and the acquired (or specific) immune system. While these components to some extent operate independently it is now recognized that there is extensive crosstalk between them, ultimately resulting in cooperative elimination of invading microorganisms and other dangerous entities.

In the presence of danger signals, the innate immune system is capable of a very rapid antibody-independent response. An important molecular component of innate immunity is the complement system of proteins. Following activation complement stimulates inflammation, coats or opsonises cells and particles for clearance, and undertakes direct cell killing via pore formation and cytolysis. The complement system, which first appeared in a primitive form in simple organisms more than 1000 million years ago (1), thus acts as a powerful first line of defense. It is activated by groups of pathogen- or danger-associated molecular patterns (2) (or a lack of host markers) and has no immunological memory, *i.e.* the extent of response is the same towards an infectious agent however many times it invades. In addition to acting independently of antibodies, the complement system is activated by antibody-antigen complexes whereupon it becomes an important effector arm of adaptive immunity. Moreover, following activation, complement can signal to B and T-cells leading to a heightened state of immune vigilance. The next section reviews the remarkable collection of protein collectively known as the complement system.

1.2 The complement system

1.2.1 Overview

The complement system (Fig. 1.1) is the major humoral molecular arm of innate immunity. The term “complement” was first introduced by Paul Ehrlich in the late 1890s and reflected its long-established role in complementing the actions of

antibodies. The complement system is now known to consist of approximately fifty serum and cell-surface proteins that act in a highly orchestrated fashion (3). The importance of the complement system for human health is a consequence of what are now recognized as its many physiological roles, which can be divided into three main overlapping categories: host defense against pathogens, maintenance of homeostasis and enhancement of the adaptive immune responses.

Complement-mediated host protection entails elimination of pathogens through the key process of opsonization whereby a cell is coated in activation-specific complement protein fragments. These then mark the cell out for destruction via phagocytosis. Direct cytolysis may also occur through activation of the terminal pathway of complement leading to formation of the membrane attack complex (MAC) (Figs. 1.1 and 1.3). The clearance of apoptotic cells and necrotic or metabolic cell debris can also occur through complement-mediated opsonization and phagocytosis hence allowing complement to fulfill its role in the maintenance of homeostasis. Likewise, immune complexes are opsonised by complement that is thus regarded as an important effector arm of antibody-driven immunity (4). Through the release, upon complement activation, of anaphylatoxic peptides, complement is pro-inflammatory and mediates chemotaxis and activation of leukocytes supporting its role in elimination of both pathogens and immune complexes. Finally activation-specific fragments of complement proteins are ligands for receptors on B-cells and T-cells, priming them for action. Figure 1.2 summarizes the main roles of complement.

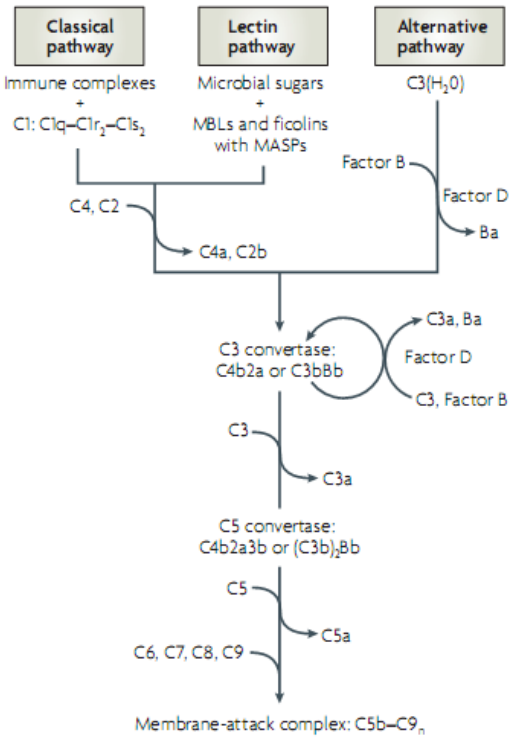


Figure 1.1 The three activation pathways of the complement system: The classical, mannose-binding lectin (MBL) and alternative pathway. All three converge at the point of formation of the C3 convertase and cleavage of C3. Each pathway can further induce the formation of the C5 convertases. Activation of C5 to C5b can further instigate the late steps which culminate in the formation of the MAC which can cause lysis and cell death (Figure from (5)).

Complement activity	Protein responsible
Host defense against infection	
Opsonization	Covalently bound fragments of C3 and C4
Chemotaxis and activation of leukocytes	Anaphylatoxins (C5a, C3a, C4a) and anaphylatoxin receptors on leukocytes
Lysis of bacteria and cells	Membrane-attack complex
Interface between innate and adaptive immunity	
Augmentation of antibody responses	C3b and C4b bound to immune complexes and to antigen; C3 receptors on B cells and antigen-presenting cells
Enhancement of immunologic memory	C3b and C4b bound to immune complexes and to antigen; C3 receptors on follicular dendritic cells
Disposal of waste	
Clearance of immune complexes from tissues	C1q; covalently bound fragments of C3 and C4
Clearance of apoptotic cells	C1q; covalently bound fragments of C3 and C4

Figure 1.2 The three main physiologic activities of the complement system (adapted from (4)).

Three pathways of the complement system (Fig. 1.1) are responsible for its activation: the classical pathway, the mannose-binding lectin pathway and the alternative pathway. Each of the three pathways is activated through a different route (Fig. 1.3). Antibody-antigen complexes are the main means of instigating the classical pathway. On the other hand, direct recognition of microbial carbohydrate motifs triggers the lectin pathway. By contrast with these pathways, which are triggered by specific recognition events, the alternative pathway works on a basis of selective protection of host surfaces. Thus, it is always activated at a low level and is thereby continuously probing every accessible surface for a lack of host-specific markers. For this reason, it is sometimes called the "tick-over" pathway (6) and it is maintained in a suppressed state in the healthy individual by regulatory proteins that are able to distinguish between host and non-host cells (or particles). As discussed below, it is very rapidly amplified in the absence of regulators as may occur on a bacterial surface.

All three pathways depend upon proteolytic cascades which results in rapid amplification of the initial response. In addition, all three pathways converge in the formation of the potentially cytolytic MAC. The enzymes that drive these cascades are initially circulating as inactive molecules in plasma. They become activated towards their substrates upon proteolytic cleavage followed by formation of complexes through exposure of nascent protein-protein recognition sites. By-products of cleavage serve to augment adaptive immune responses and to mediate inflammation.

As may be seen from figure 1.3, all three pathways converge at the cleavage of C3 to C3b. This reaction is performed by the binary complexes, called C3 convertases, consisting of activated fragments of C3 and factor B in the alternative pathway, or C4 and C2 in the classical pathway. The formation of C3b is a prerequisite of formation of the C5 convertases that cleave C5 to C5b, another key step in the complement pathway. C5b formation then triggers the non-enzymatic self-assembly of the MAC (7).

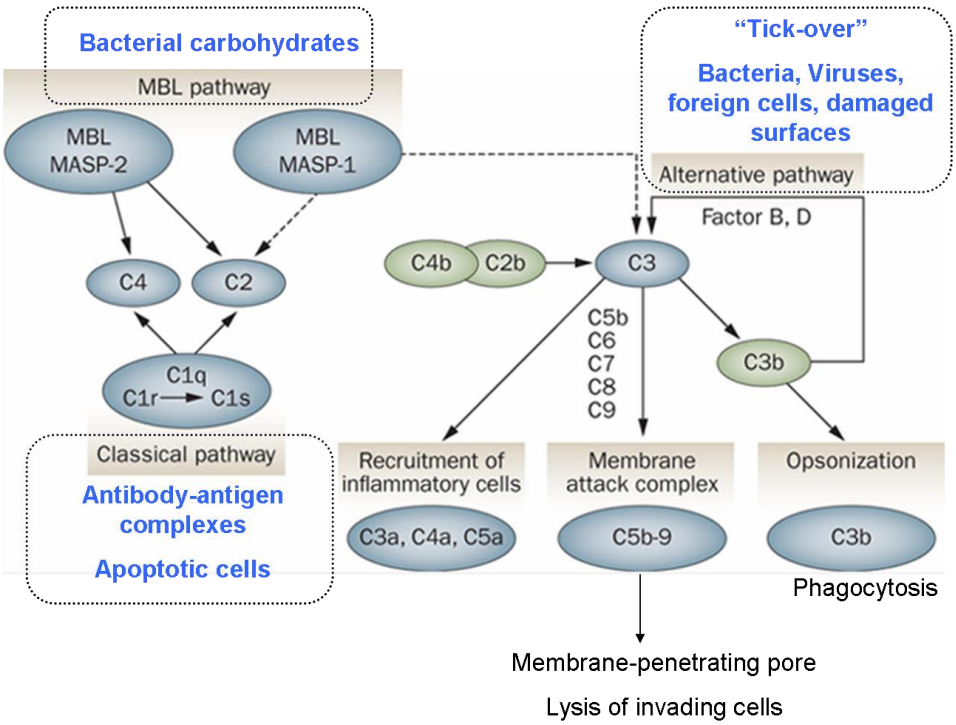


Figure 1.3 Depiction of the complement cascade. Activation can occur through three routes and factors that can instigate the complement cascade are shown on the figure. Terminal events include the formation of the membrane-attack complex, C3b opsonization and phagocytosis and recruitment of inflammatory cells (Figure modified from (8)).

1.2.2 The three activation pathways of the complement system

1.2.2.1 The classical pathway

The binding of the six globular headgroups of the C1q component (a hexamer of trimers) of the C1 complex of complement to the Fc portions of antibodies-bound-to-antigens (e.g. on the surface of an invading microorganism) initiates the classical pathway. A conformational change of the collagenous stem of C1q causes autoactivation (by conformational change and proteolysis) of the tightly associated C1r components that subsequently cleave and thereby activate the neighbouring C1s components. The activated form of C1s cleaves C4 to generate C4b that can attach to the bacterial surface via a short-lived thioester group. Cleavage of C4 also creates a new binding surface for C2 and formation of the pro-convertase complex, C4b2. As part of this complex, C2 is also susceptible to cleavage by activated C1s. This leads to the formation of C4b2a which is the C3 convertase of the classical pathway (4).

1.2.2.2 The lectin pathway

The lectin pathway is similar to the classical pathway except that it is triggered not by antibody-antigen complexes but by pathogen-associated molecular patterns (PAMP) such as the mannose moieties on a pathogen surface. Instead of C1 mannose-binding lectin in tight association with the serine proteases mannose-binding lectin-associated proteases 1 and 2 (MASP1 and MASP2) is utilized as the recognition protein. MASP1 and MASP2 act in a similar manner to C1r and C1s of the classical pathway and (in an analogous fashion to the actions of the C1 complex) lead to the formation of C4b2a, which is of course the C3 convertase complex (4). Thus the formation of C4b2a and cleavage of C3 to C3b is the point of convergence of both classical and lectin pathways.

1.2.2.3 The alternative pathway

As previously mentioned this pathway is always activated at a low level and in the absence of regulators can be rapidly amplified. The C3 convertase of the alternative pathway is composed of orthologous proteins C3b (from C3 (Fig. 1.4) that is similar to C4) and Bb (from factor B that is similar to C2). A version of C3 in which the internal thioester has been hydrolysed, called C3(H_2O), resembles C3b in terms of structure and ability to nucleate assembly of the C3 convertase.

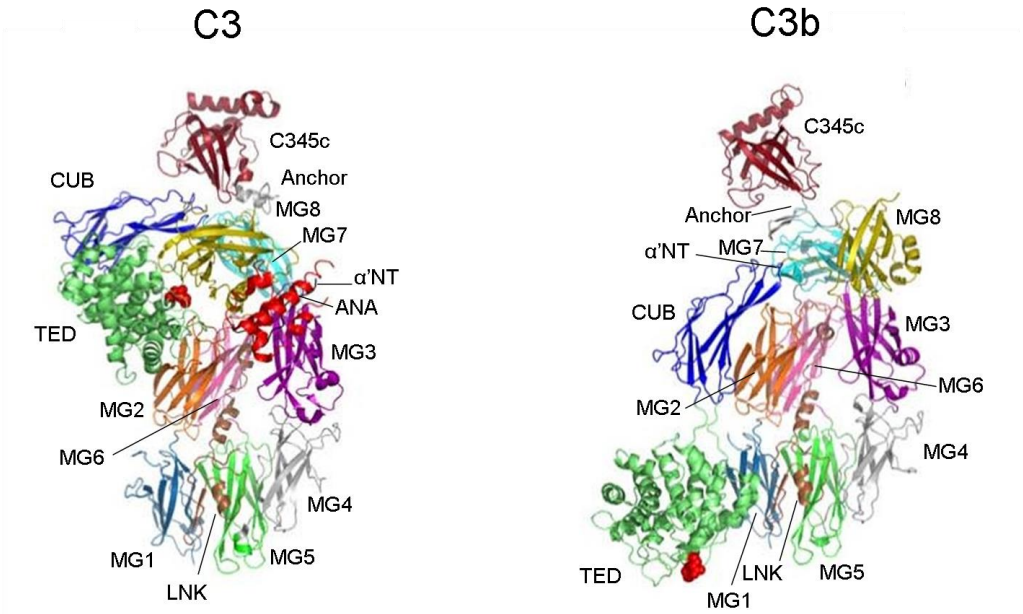


Figure 1.4 Crystal structures of C3 and C3b showing all of the 13 domains and the rearrangement they undergo upon formation of C3b (PDB access codes: C3: 2A73; C3b: 2I07) α'NT, N-terminus of the α-chain; ANA, anaphylatoxin domain; CUB, complement C1q/C1s, uEGF, BMP1; LNK, linker; MG, macroglobulin; TED; thioester-containing domain (9), (10).

Upon cleavage by the C3 convertase and removal of the ANA domain (generating C3a which is an anaphylatoxin), C3b undergoes extensive interdomain conformational changes (Figs. 1.4 and 1.5) exposing and activating the abovementioned thioester. This is highly reactive (unlike C3(H_2O) in which the thioester has been hydrolysed) towards nucleophiles and reacts rapidly either with solvent water or with a hydroxyl or amine groups on a near-by surfaces (5), (9), (11), (12). Thus both surface and fluid-phase complement activation is possible.

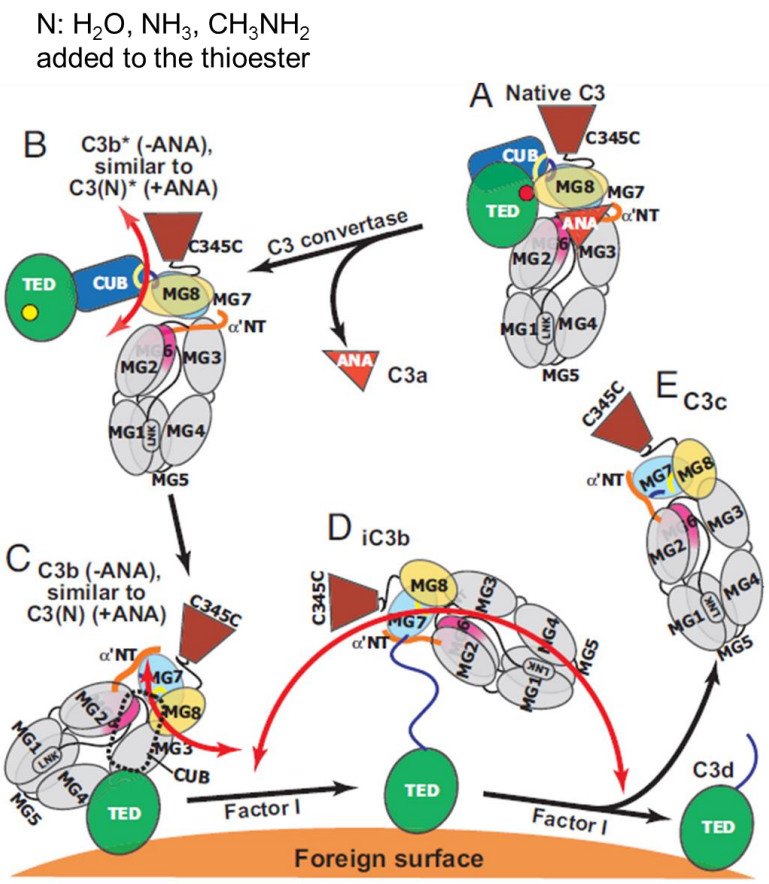


Figure 1.5 Schematic representation of the domain rearrangement C3 undergoes upon activation. The blue line in (D) corresponds to the unfolded N-terminal region of iC3b. The orange arrows represent the mobility range of the key moiety in the TED domain. (Figure adapted from (13)).

A similar conformational change/activation process occurs without cleavage of C3a- this much slower process leads to the tick-over convertase complex, C3(H_2O)Bb (Fig. 1.6).

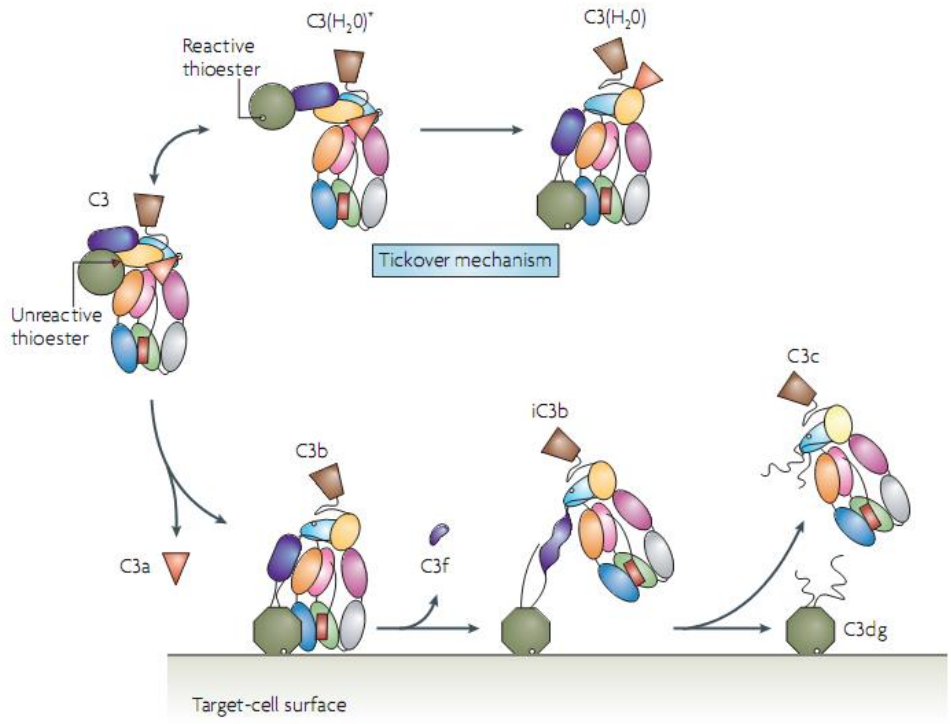


Figure 1.6 Schematic representations of the two activation mechanisms of C3 and its products. In both cases C3 undergoes conformational changes. The tick-over mechanism results in the formation of C3(H_2O) (through a stable intermediate denoted C3(H_2O)*). Activation of C3 by the C3 convertase results in the formation of C3b (similar to C3(H_2O) and the anaphylatoxin C3a. C3b gets processed further to form - in a sequential manner - iC3b (following removal of C3f), C3c and C3dg (Figure from (5)).

Formation of C3(H_2O) is a spontaneous (if rare) event and occurs ubiquitously albeit at a low level. C3(H_2O), like C3b, can bind to factor B. In the context of a complex with C3(H_2O) (or C3b) factor B is a substrate for a highly specific serine protease, factor D generating C3(H_2O)Bb, the so-called initiating C3 convertase (Fig. 1.7) that cleaves C3 to C3b. These C3b molecules can further nucleate formation of C3bBb complexes that can – unless they are down-regulated - produce additional C3b molecules (positive feedback amplification loop).

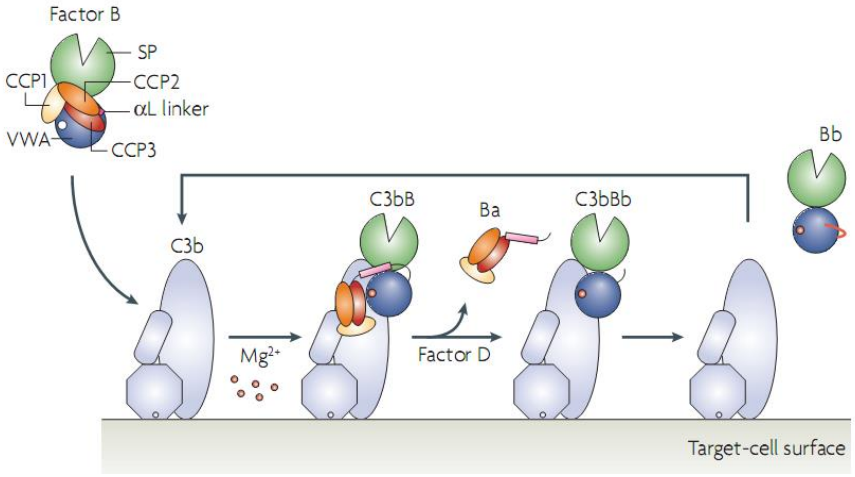


Figure 1.7 Schematic representation of the C3 convertase formation. Factor B associates with the surface-bound C3b in a Mg^{2+}-dependent manner and is then activated by factor D; the Ba fragment then dissociates form the complex. CCP, complement control protein; SP, serine protease; VWA, von Willebrand factor A (Figure and legend from (5)).

Both classical and alternative C3 convertases can cleave more C3 molecules to C3b, which can form more C3 convertases, thus leading to the amplification loop that explains the very fast response of the complement system (14). Additionally, both C3b and C4b molecules are opsonins helping to target foreign cells or other dangerous particles for phagocytosis.

1.2.2.4 Terminal events

In the absence of regulators, the bimolecular C3 convertases C3bBb and C4b2a interact with further C3b molecules to form the trimolecular C5 convertases. The C5 convertase of the classical/lectin pathway is designated C4b2b3b whereas that of the alternative is C3bBb3b. The C5 convertases cleave C5 (note that C3, C4 and C5 have high sequence similarity but differ in that C5 has no thioester moiety). This leads to the formation of the small C5a fragment and C5b. C5a along with C3a and C4a are anaphylatoxins, which are capable of initiating inflammatory responses. Formation of C5b is the first step towards the formation of the MAC. C5b interacts sequentially with C6, C7, C8 and multiple copies of C9 to form the MAC which creates a membrane-penetrating pore that leads to lysis of cell membranes (7).

Crucially, the MAC is unlikely to form on host cell surfaces thanks to their protection by natural regulators of all three pathways - this is the basis for specificity of the complement system. In the presence of regulators such as membrane cofactor protein (MCP) and FH, C3b is cleaved into inactive iC3b by factor I (FI) while the life-times of C3 and C5 convertases are curtailed (*i.e.* their decay is accelerated). These activities are termed cofactor activity (for FI) (CA) or decay-accelerating activity, respectively (DAA) (Fig. 1.8). iC3b remains attached to the surface as it contains the TED domain. In the presence of complement receptor type 1 (CR1) (or FH at low salt) iC3b gets further cleaved by FI into C3dg and C3c, and finally C3dg is cleaved by non-specific proteases to leave C3d; C3d corresponds almost exclusively to the TED (Fig. 1.8).

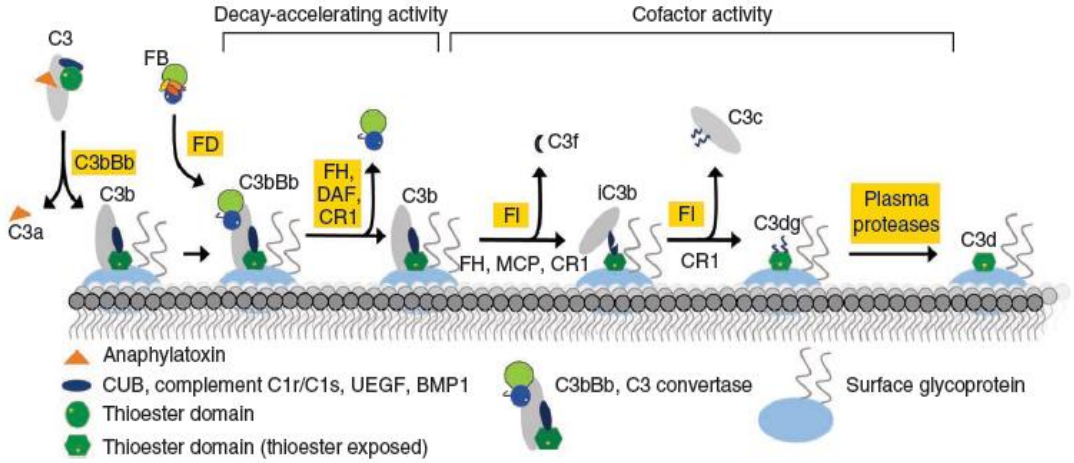


Figure 1.8 The central event in complement activation is the cleavage of C3 to C3b by C3 convertase, accompanied by attachment to surfaces mediated by the TED. In the presence of additional complement regulatory molecules, C3b may be further degraded sequentially to iC3b, C3c, C3dg and C3d. C3d corresponds to the TED and remains bound to the surface (Figure and legend from (15)).

In conclusion, the activation of complement can occur very rapidly via multiple pathways and a positive-feedback loop, and complement is essential given its potential to damage healthy host tissues as well as pathogenic or dangerous ones. In particular, deficiencies in regulators are a source of numerous disease symptoms (see section 1.7) that could be ameliorated if therapeutic means of complement inhibition were available.

1.3 Regulators of complement activation (RCA)

A fine balance between activation and regulation of the complement system must be maintained at all times so as to not jeopardize the healthy state of the host. This is achieved through a panel of regulatory proteins. These regulators, which can act both in fluid phase and on cell-surfaces, are mostly members of the regulators of the complement activation (RCA) family. They comprise multiple domains which consist entirely or almost entirely of the so-called complement control protein (CCP) modules which are arranged in a "beads-on-a-string" fashion. For example FH, a soluble regulator of the alternative pathway consists solely of twenty CCP modules. On the other hand, MCP (contains four CCP modules), DAF (contains four CCP modules), CR1 (usually contains 30 CCP modules) and C4BP (contains 59 CCP modules) consist almost entirely of the abovementioned module type. C4BP for example has seven α-chains, with eight CCP modules per chain, and one β-chain which contains 3 CCP modules. The α-chains have an additional C-terminal region that forms disulfide bonds with a C-terminal region in the β-chain resulting in the mature protein (16). The table below (Table 1.1) summarizes some of the regulators of the RCA family (16), (17).

Regulator	Location	Function
Factor H (FH)	In fluid phase and on surfaces (when C3b is also bound)	Recognizes self-surfaces; accelerates convertase decay (alternative pathway); cofactor for factor I
Factor H-like protein 1 (FHL-1)	In fluid phase	Accelerates convertase decay (alternative pathway); cofactor for factor I
C4b-binding protein (C4BP)	In fluid phase	Accelerates convertase decay (lectin/classical pathway); cofactor for factor I
Membrane co-factor protein (MCP or CD46)	Bound on surfaces	Cofactor for factor I
Complement receptor 1 (CR1 or CD35)	Bound on surfaces	Induces phagocytosis; accelerates decay of convertases; cofactor for factor I
Decay accelerating factor (DAF or CD55)	Bound on surfaces	Accelerates decay of convertases

Table 1.1 Regulators of complement activation (RCA) (3).

FH is unique in the RCA family in that it can act both in fluid phase and on cell surfaces (18) unlike membrane-restricted regulators (17). It ensures that complement-mediated damage arising from C3b amplification is directed away from host tissue and is particularly important for the protection of surfaces not enclosed by membranes (*i.e.* with no membrane-bound regulators).

1.4 FH

FH is encoded by the *CFH* gene in the RCA gene cluster on chromosome 1q32 (19). It is a 155-kDa soluble single-chain glycoprotein (20) expressed mainly by the liver and secreted into plasma where it circulates abundantly at concentrations of 116-562 μg/ml (0.8-3.6 μM) (21). The mature polypeptide consists of 1213 amino acids that may be divided into 20 CCPs of about 60 residues each. FH is the pivotal soluble regulator of the alternative pathway of the complement system (22). It has the following principal roles when acting in fluid phase and/or on self surfaces: it hinders the formation of the alternative pathway C3 convertase by competing with

factor B for binding to C3b, it accelerates the decay of the C3 convertase complexes (C3bBb) that do succeed in forming (*i.e.* DAA) and it acts as cofactor for FI-mediated cleavage of C3b into inactive iC3b (CA). The table below (Table 1.2) summarizes the role, mechanism and outcome of the main tasks of FH.

Effector function	Mechanism	Outcome
Hinders the formation of the alternative pathway C3 convertases, C3bBb	Competes with factor B and blocks its access for binding to C3b, by binding first to C3b	Complement is no longer activated: FH-bound C3b cannot bind factor B, C3 convertase cannot be formed, hence the proteolytic cascade and amplification loop are interrupted
Decay accelerating activity (DAA)	Bb of C3bBb may be displaced from C3bBb by electrostatic repulsion and steric obstruction	Complement activation and cascade are no longer functional
Cofactor activity (CA)	Acts as cofactor to serine protease FI for the further processing and inactivation of C3b to iC3b [33] When C3b in fluid phase: immediate binding and inactivation When C3b surface-bound: binding and inactivation dependent on the chemical composition of the surrounding surface	iC3b can no longer interact with convertases hence is no longer part of the amplification loop iC3b can be processed further by FI to C3dg and C3c, the first promoting phagocytosis and B cell stimulation while the second B cell stimulation

Table 1.2 Summary of the three main roles of FH, their mechanisms and the outcomes in the context of regulating the alternative pathway of complement for protection of the host.

1.4.1 Functionally relevant binding sites of FH and its importance in the complement pathway

It is of utmost importance that the complement system distinguishes between self and non-self surfaces and in this respect FH plays a crucial role. Since it is the only alternative pathway regulator that can act both in fluid phase and on cell surfaces, host cells that lack surface-bound regulators must rely upon FH for their protection.

The selectiveness towards host surfaces displayed by FH can be attributed to the existence of polyanions, such as glycosaminoglycans (GAGs) and sialic acid, on

mammalian cell surfaces (23). Earlier studies had suggested a sialic acid/GAG binding site located within FH-19-20, (24), (25) another one within FH-7 (26), (27) a third within FH-9 (28) and a fourth within FH-12-15 (29). However, more recent studies (30) showed that FH-9 (when produced without expression artifacts) does not encompass a GAG-binding site. Moreover, the whole of the region 10-15 failed to show any significant affinity for heparin (30). This leaves two well-established and universally agreed GAG-binding sites: on FH-19-20 and in FH-7 (Fig. 1.9).

A binding site for C3b is located within FH-1-4 and a crystal structure of that region in complex with C3b (Fig. 1.13) has been solved (31). Interestingly, these four modules are necessary and sufficient, in isolation from the remainder of FH, to perform the fluid-phase cofactor and decay-accelerating functions of FH although they do not distinguish between C3b bound on a self surface and C3b in fluid phase or on a complement-activating surface. At the C-terminal end of the molecule, FH-19-20 binds both C3b and polyanions (32). A third C3b-binding site in FH-12-14 was suggested (33) but more recently discounted as a primary C3b-interaction site (30). Finally, FH-6-8 showed a very much weaker affinity for C3b (Fig. 1.10) than the N-terminal and C-terminal sites (30) so it is not considered a primary C3b-binding site either.

Thus, in summary, the N-terminal (FH-1-7) and C-terminal regions (FH-19-20) of FH are predominantly responsible for CA and DAA and for self-surface discrimination by FH. It is suggested that some degree of avidity is conferred by simultaneous binding of the sites of the two termini (30). If these are binding to the same C3b molecule then the effect is presumably facilitated by the bent-back conformation of the central region of FH (30). On the other hand, it is possible that FH is fairly extended and spans two C3b molecules deposited on the sensor chip surface.

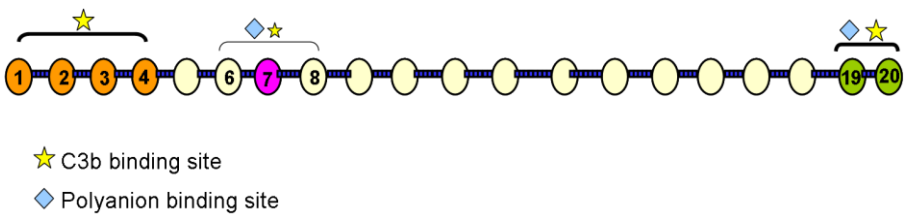


Figure 1.9 Functionally relevant binding sites for FH.

1.4.2 Structure of FH

The 20 domains that comprise FH consist of between 51 and 63 amino acid residues. Each has four invariant cysteines (I-IV) along with a near-invariant tryptophan between CysIII and CysIV. Other highly conserved residues include glycines, prolines, and various hydrophobic residues (Fig. 1.11) (34), (35). Each repeating sequence folds into a β-sheet rich domain, with two disulfide linkages (between CysI and CysIIII and between Cys II and CysIV) and has been called the complement control protein (CCP) module. The use of the term module arises from the occurrence of similar 3D structures, sharing the same underlying consensus sequence, in numerous proteins with diverse functions (16). This module is particularly dominant, however, in the RCA proteins (see also section 1.3) that control complement, hence the name CCP.

Figure 1.10 shows the first NMR structure of a CCP module, FH-16 (36). The N- and C- termini lie at opposite poles of the module; this is consistent with a "beads-on-a-string" arrangement of the 20 CCPs in the context of full length FH. Figure 1.12 shows a multiple sequence alignment of the 20 CCPs that comprise FH.

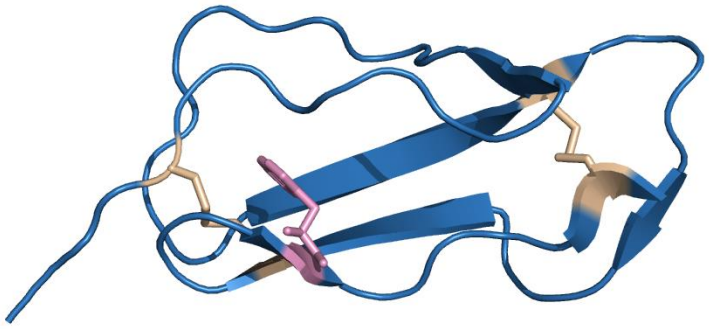


Figure 1.10 Cartoon representation of the NMR structure of FH-16. The N-terminus is on the left and the C-terminus on the right. The disulfide bonds are shown in beige while the consensus tryptophan residue is shown in pink (36), (37).

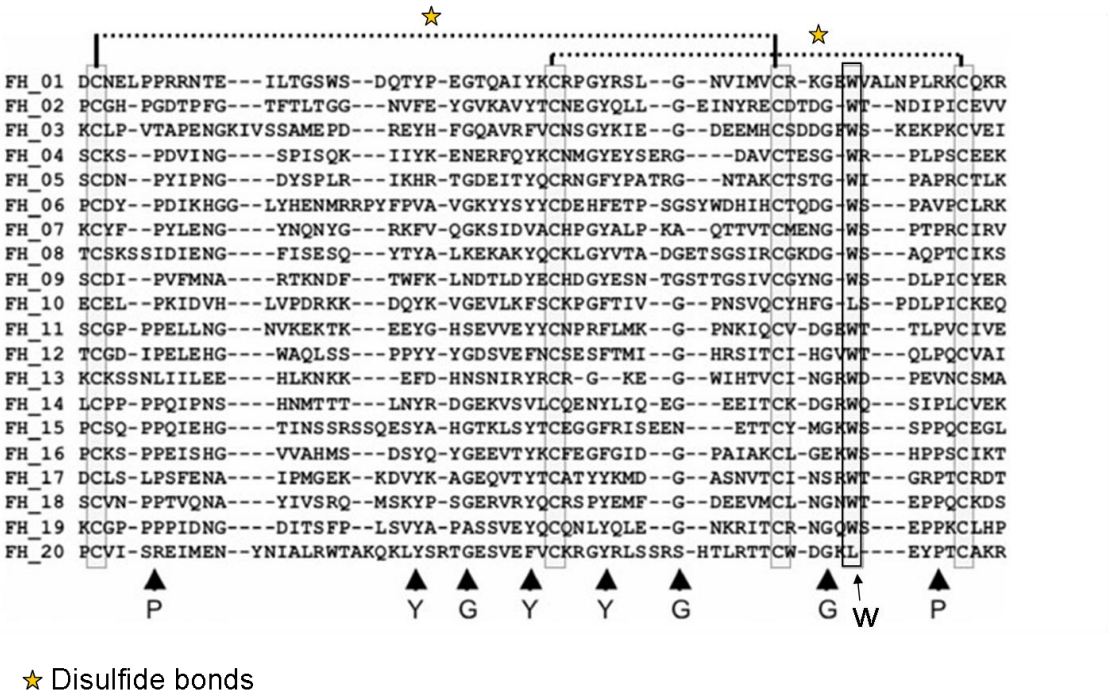


Figure 1.11 Multiple sequence alignment of 20 CCP modules in sequence of complement FH. One-letter codes are used throughout. Invariant Cys residues and almost invariant Trp residues are highlighted by boxes. Horizontal lines indicate disulfides within CCP modules (Figure and legend modified from (38)).

Several methods have been employed to gain structural knowledge of FH. It is too large a protein to be studied directly by nuclear magnetic resonance (NMR) spectroscopy while there are no reports of successful crystallization trials. Its modular composition lends it to a dissection strategy and there are numerous structures for various small fragments of FH (see Fig. 1.12).

1.4.3 Structures and characteristics of the CCP modules that comprise FH

Three-dimensional high-resolution structures have been determined by X-ray crystallography and/or NMR, in the context of fragments up to four CCPs, for all the CCPs in FH apart from 9, 10, 11, 14 and 17. In general, each CCP was shown to have a prolate ellipsoid shape of dimensions approximately 4 nm by 2 nm by 1.5 nm. The 20 CCPs of FH are joined by short linkers. Since CysI and CysIV lie very close to the beginning and end of each of the CCPs, linkers are normally defined as lying between CysIV of module (i) and CysI of module (i+1). By this definition the linkers

of FH are between three and eight residues in length. In reality module boundaries, as judged from three-dimensional structures, extend beyond these cysteine residues by a residue or two (36), (37).

NMR studies suggested FH-1-2 and FH-2-3 (39) exist in elongated conformations with little tilt between adjacent modules (Fig. 1.12) and this was backed up by a subsequent X-ray structure of FH-1-4 (in complex with C3b) that also showed a strongly tilted kinked FH-3-4 intermodular angle (31). This overall arrangement was further validated by a subsequent FRET study (40). Crystal structures of fragments of FH corresponding to FH-6-7 (41), (42) revealed an overall elongated shape. By way of contrast, an NMR-derived structure of FH-12-13 revealed a rigid 80^{o} "V-like" bend between the modules (43) (Fig. 1.12) Earlier NMR work on FH-15-16 failed to find more than a single NOE (distance restraint) between modules suggesting intermodular flexibility (44). Both X-ray and NMR studies showed that FH-19-20 (27), (45) adopt a rod-like shape, while X-ray studies on FH-18-20 showed that this region is "J"- shaped in which a ~122^{o} tilt is introduced between FH-18 and 19 (Fig.1.12). Low-resolution small-angle X-ray scattering (SAXS)-based studies showed however that in solution, FH-18-20 adopts multiple conformations mediated by flexibility between 18 and 19 (46). SAXS was also employed to study FH-11-14, 10-15 (43) and 8-15 (15). All pointed to a compact conformation of these regions which is consistent with these central modules (linker lengths within the 10-15 region are - in order - 6, 6, 8, 7 and 5) promoting a bent-back structure of FH that could be functionally important (see below). Finally, low resolution data on the regions 1-5, 6-8 and 16-20 pointed to not compactly arranged structures which in turn agrees with the high resolution information - where available - for these regions (47), (48).

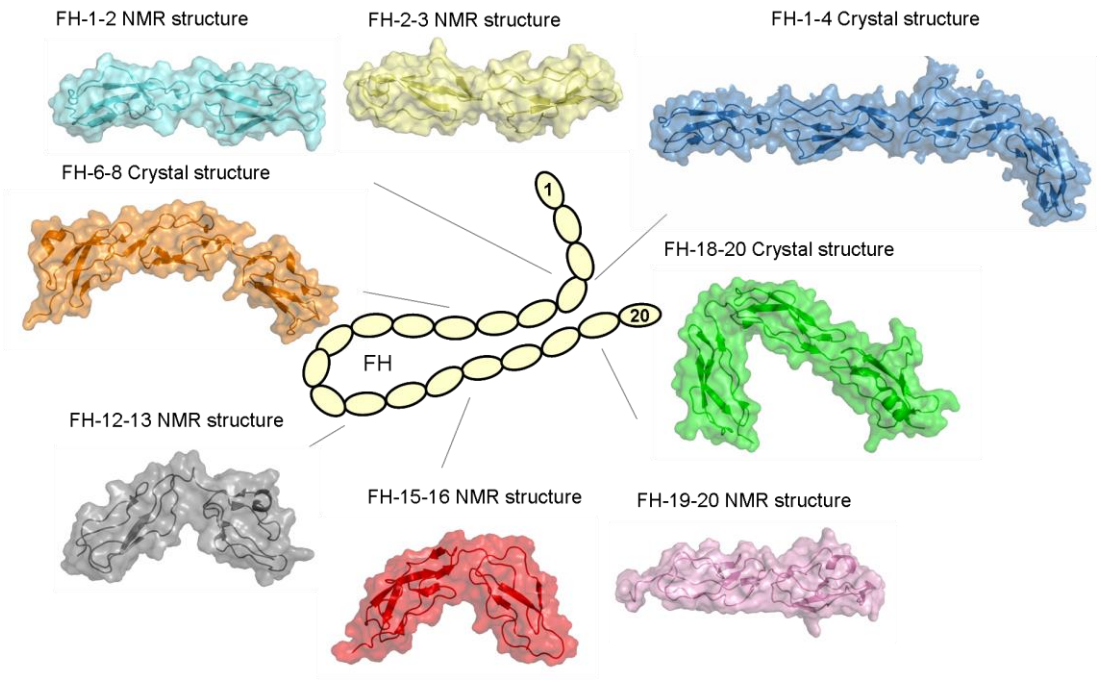


Figure 1.12 High resolution crystal and NMR structures for fragments of FH as indicated on the structures. The structures are shown in cartoon and semitransparent surface representation. The FH-15-16 structure shown is one conformer of many possibilities since only a single NOE could be found between the modules. Full length FH is shown in the middle. PDB accession codes are as follows: FH-1-2 NMR structure: 2RLP; FH-2-3 NMR structure: 2RLQ; FH-1-4 Crystal structure: 2WII; FH-6-8 Crystal structure: 2UWN; FH-18-20 Crystal structure: 3SW0; FH-12-13 NMR structure: 2KMS; FH-15-16 NMR structure: 1HFH; FH-19-20 NMR structure: 2BZM.

A bent-back overall average conformation for full-length FH is supported by various studies. Initially, a doubled-over appearance of the FH molecule was shown by SAXS and analytical ultracentrifugation (AUC) (49), while transmission electron microscopy (EM) (50) broadly agreed with the observation since the majority of the FH molecules observed on the grid appeared folded-back upon themselves. This would bring the two ends, N- and C- terminal regions that harbor the functional sites, close to one another. This can be further supported by the rather extended nature of the two termini of FH.

It is a matter of contention as to whether FH exists (and functions) in a monomeric, dimeric or higher order oligomeric form. It was inferred from scattering data and AUC analysis that FH exists predominantly as a monomer (49). Subsequent studies using the same techniques showed that there were a proportion of dimers, as well as higher order oligomers present (the resultant K_D for self-association was estimated to be 28 μM) (51).

In line with the favored bent-back conformation of FH, it was interesting that the protein became more extended with increase in salt concentration (52). Finally, the presence of metal ions such as zinc at higher concentrations induced oligomerization of FH (53).

FH encompasses nine potential N-glycosylation sites, of which eight - on FH-9, 12, 13, 14, 15 (2 sites), 17 and 18 are occupied mostly by non-fucosylated diantennary disialylated glycans (54), (55). Consistent with their locations away from functional sites, glycans do not appear to be required for function but their absence may have an impact on the solubility and stability of the protein.

1.4.4 Structure of FH-1-4 in complex with C3b

The crystal structure of FH-1-4 (31) with C3b showed a large contact interface consisting of four surface regions spanning the length of FH-1-4 (Fig. 1.13). The structure facilitates hypothetical mechanisms of DAA and FI CA of FH. It was suggested on the basis of the structure that FH-1-2 is predominantly responsible for competing with factor B, and perhaps also for the decay-accelerating function since by binding to the α'NT, MG2 and MG6-7 C3b domains, they could displace Bb from C3bBb by electrostatic repulsion and steric obstruction (31). In fact, the structure does not shed very much light on decay acceleration since it does not explain how FH-1-4 could access its binding site unless Bb vacates its binding site first. One possibility is that the complex of C3b with Bb is in fact dynamic with Bb modules peeling away from time to time allowing individual FH modules to take their place in an inchworm-fashion.

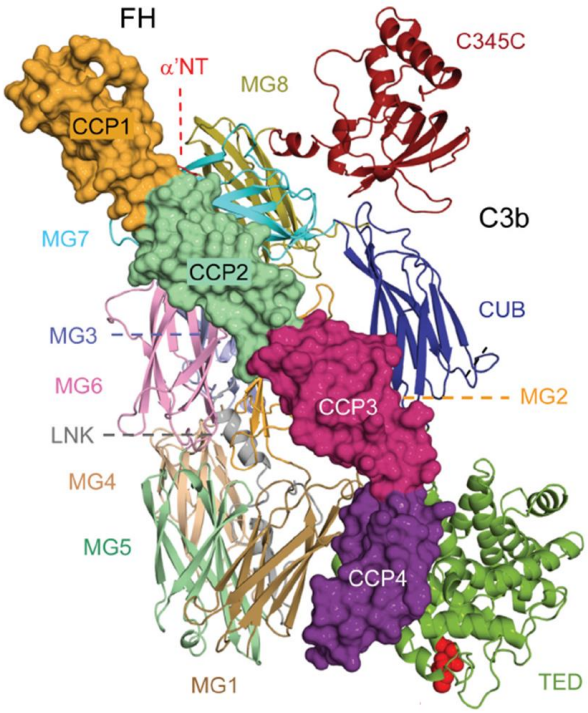


Figure 1.13 Overall structure of C3b in complex with FH-1-4 (PDB access code 2WII). C3b is shown in ribbon and the various domains are indicated. FH is shown in surface representations with its domains labeled. Residues that normally participate in the formation of the thioester (sidechains of Cys988 and Gln991) are shown as red spheres. In this study C3b was treated with iodoacetamide after generation by limited trypsinization (31) (Figure adapted from (31)).

1.4.5 Structure of FI and modeling of ternary complexes

The recent crystal structure of FI and the subsequent modeling of the ternary complex consisting of FI, FH-1-4 and C3b, (Fig. 1.14), allowed for an insightful proposed model for FH CA. Briefly, initially FH-1-4 of FH bind to C3b. This platform then enables FI to dock in the area formed around FH-1-3-bound to C3b. The binding of FI promotes allosteric activation of the light chain serine protease domain (SP) which then enables for the first cleavage of C3b to take place (56).

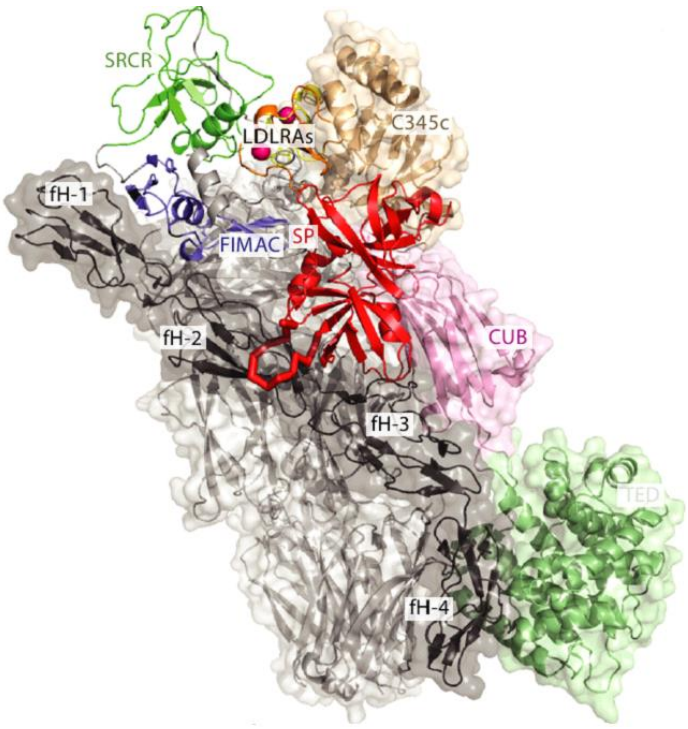


Figure 1.14 Modeling of the ternary complex of C3b, FH-1-4 and FI. FI is shown in cartoon representation (PDB access code: 2XRC); C3b and FH-1-4 are shown in cartoons and semi-transparent surfaces. The C3b are color coded as follows: C345C domain is shown in light bronze; the CUB domain in pink; the thioester-containing domain in light green; and the linker domain, C3 α'N-terminal domain and macroglobulin domains of C3b in light gray (Figure and legend from (56)).

1.4.6 A model for engagement of FH on self surfaces

The efficacy of FH in engaging with surface-attached C3b is greatly enhanced in the presence of the host's abundantly distributed polyanionic markers. Two independent studies presented crystal structures of the complex between C3d and FH-19-20 (15), (57). The results, together with the crystal structure of FH-1-4 of FH (31) with C3b, (Fig. 1.15) elucidated how FH could engage trivalently, with two sites on C3b and a third site involving GAGs on a self-surface. A merged superimposed model for the ternary complex of C3b with FH-1-4 and 19-20 showed that these the two FH segments bind to C3b adjacent to one another but they occupy distinct sites consistent with simultaneous binding. Furthermore the majority of FH-20 was available for interaction with GAGs, consistent with the concept of the C-terminus binding to both C3b and GAGs (15). An earlier study set to test this concept showed by NMR-chemical shift mapping that when a heparin sulfate analog was

titrated in a sample containing purified ^{15}N-lebeled FH-19-20 the chemical shifts of several residues were perturbed. These residues were mapped onto FH-20 and specifically in a region with complementary charge distribution. Furthermore, the same study showed, by NMR combined with computational docking, that the C3d part of C3b bound mostly to FH-20, at the opposing face of the GAG binding site (25). A bent-back central region of FH (FH-8-15) – consistent with the SAXS data – would enable FH to exhibit this binding mode (Fig.1.16). The possibility that FH-19-20 binds both C3b and a nearby C3d, when these are appropriately oriented, has also been suggested. This cannot be ruled out (57) although it does require a particular juxtaposition of the two molecules on a surface that might be rather rare. Most likely, FH is able to bend back upon itself in such a way that FH-1-4 can perform CA and DAA while FH-19-20 exercise discrimination between host and non-host polyanions (15) (Fig. 1.15).

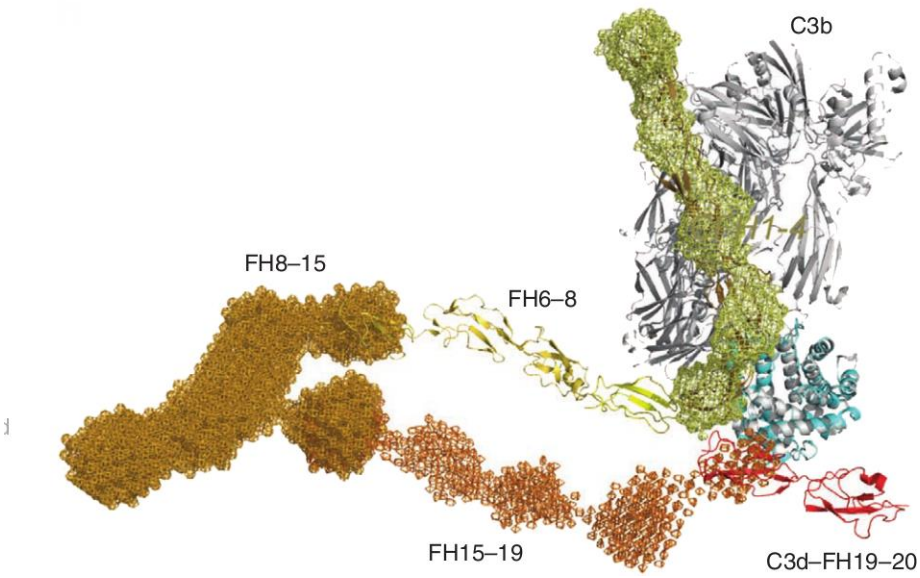


Figure 1.15 Potential model of FH engagement with surface-bound C3b. The FH-19-20-C3d structure (shown in red and cyan, pdb access code: 3OXU) superimposed (C3d on TED) onto the FH-1-4-C3b structure (tan and gray). This FH model has been constructed by juxtaposition or superposition of the structures of FH-5 (yellow) and FH-6-8 (yellow-orange) and the SAXS-shape envelopes pf FH-8-15 (bright orange) and FH-15-19 (orange). Structures are shown in cartoon representation and SAXS shape envelopes are shown in mesh. The FH-1-4 SAXS shape envelope (pale yellow) is superimposed onto the C3b-FH-1-4 structure (Figure and legend from (15)).

1.5 Central region of FH

As mentioned in the previous section, several studies by SAXS, AUC, EM and NMR suggested that a folded-back conformation of the FH molecule is favored that brings the two ends close to one another. This is interesting since the functional regions of FH sites involved in binding to the convertases and recognizing host-cell surfaces are located towards either end of the molecule. It remains uncertain whether such a putative bend is the result of an accumulation of small tilts between many modules or whether it is brought about by several large tilt angles involving the longer linkers between the central FH modules. This region, FH-8-15, devoid of recognition sites, could be of architectural importance, required for ensuring the two functional ends of the protein are able to cooperate in recognizing and de-activating convertases on host surfaces. It is unknown whether they represent a fully flexible tether that simply imposes a maximum distance between the ends, a rigid section that imposes a defined shape on FH, or a region that acts like a hinge able to open and close.

The FH-10-15 region deviates from the remaining CCP modules in terms of size and intermodular linker length. In contrast with three-to-five residues within linkers in the non-central part, linker lengths of up to eight amino acids are encountered within this central region.

1.5.1 Structural information on the central region of FH

The NMR-derived structure of FH-12-13 shows that the surface of FH-12 is predominantly electronegative (43). FH-13 is an unusual module as manifested through some notable features. First of all, it is the smallest CCP module known, just 51 amino acids long. It is connected to FH-12 by the longest linker, eight residues, in the whole RCA family. Interestingly, three residues in an aberrant loop of FH-13 (^{766}KNK^{768}), project to the side of the module and contribute to a striking electropositive surface patch. The NMR-derived data tally with observations from size-exclusion chromatography, AUC and SAXS. All are consistent with the conclusion that a rigid 80^{o} angle is formed between the two modules (43) (Fig.1.16).

FH-11-14, 10-15 and 8-15 were also subjected to biophysical studies. AUC and SAXS agree as to the monomeric state of all of these proteins in solution. For FH-11-14 and FH-10-15, size-exclusion chromatography excluded the possibility of fully elongated shapes but was consistent with a compact conformation (43). SAXS data support the presence of mutually tilted modules. The SAXS data for modules 10-15 suggest that the tilts between the modules are not in the same direction. This suggests a zigzag arrangement rather that a smooth bend. Both AUC and SAXS showed that FH-10-15 although containing two extra CCPs compared to FH-11-14 forms a less elongated structure thus implying that FH-10 and 15 tilt towards 11-14 (43). Thus, these central modules adopt a compact overall structure that seems to exhibit little intermodular flexibility (43) (Fig. 1.16). FH-8-15 was described based on SAXS as relatively compact and bent back which is consistent with the observations made for FH-11-14 and 10-15 (15) (Fig. 1.16).

The unusual organization of the CCPs that comprise the central region, relative to one another, along with the rigid defined bend of FH-12-13 give rise to the question of whether flanking CCP module pairs such as FH-10-11, FH-11-12 and FH-13-14 would exhibit such structural features, for example in terms of defined bent arrangements and lack of flexibility.

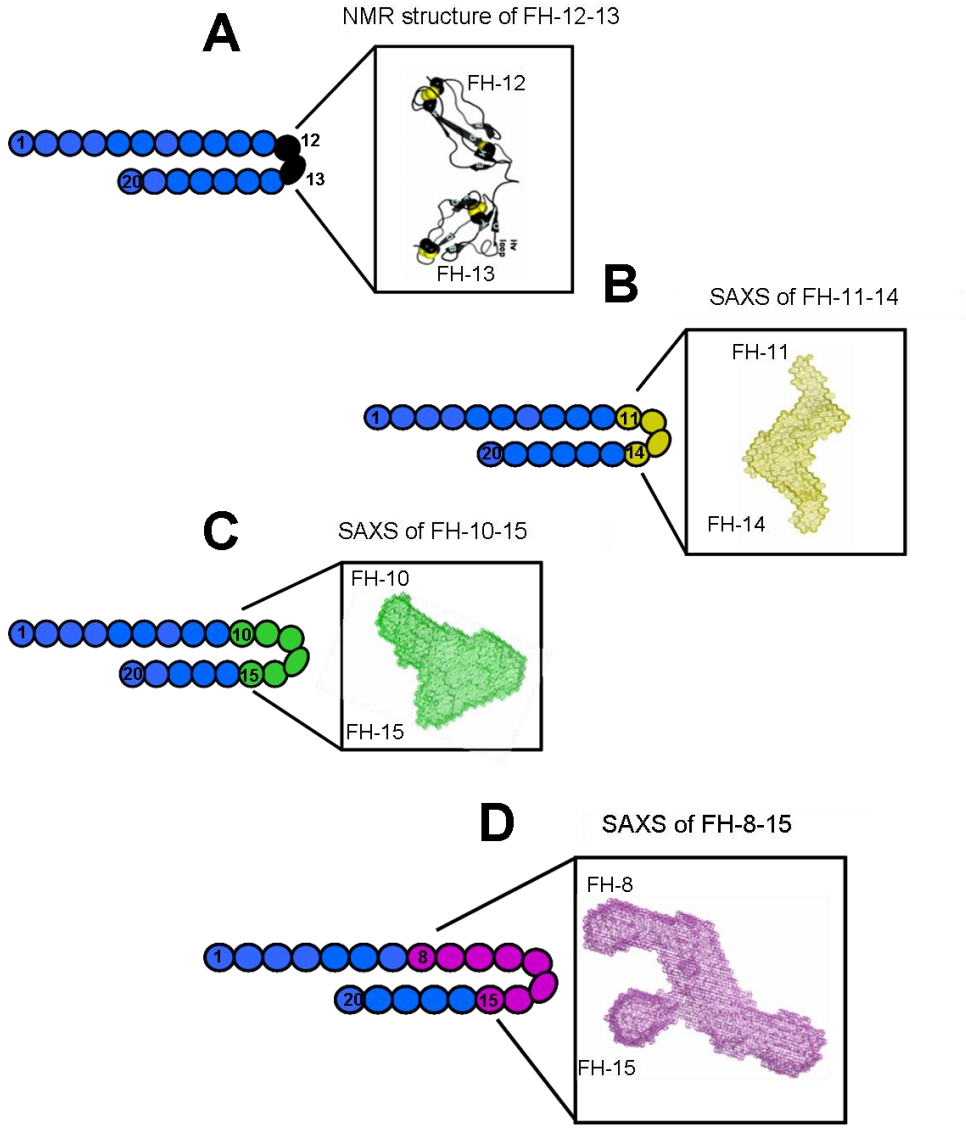


Figure 1.16 Structural analysis of the central region of FH by NMR and SAXS. (A) NMR-derived structure of FH-12-13 (PDB access code: 2KMS) (43). (B), (C), (D) SAXS-derived shape envelopes for FH-11-14, FH-10-15 and FH-8-15 respectively (43), (15).

1.6 FH protein family

The FH protein family consists of FH, FH-like protein 1 (FHL-1) and five FH-related proteins (FHR-1, -2, -3, -4, -5). All members are composed solely of CCP modules. FHL-1 is a FH splice variant consisting of the first seven N-terminal modules of FH and a C-terminal four-amino acid extension, and displays similar functional regulatory activities to FH (58), (59), (60). The functional role of the FHR (FHR1-FHR5) is not yet fully characterized. The FHR proteins are encoded by their

own genes, located in the regulators of complement activation cluster. They are built from five to nine CCP modules and some have been shown to bind to C3b and polyanions (61), (62). FHR-1 was shown to regulate C5 convertases (63) while FHR-4 may bind to C-reactive protein and take part in opsonization of necrotic cells (64). These proteins circulate in plasma at much lower levels than the main regulator FH.

1.7 FH and disease

When the complement pathway is not properly regulated, pathological conditions are likely to arise. Specifically, with respect to FH, mutations or polymorphisms in the gene of FH have been found to be associated with membranoproliferative glomerulonephritis type II (MPGN2), atypical haemolytic uraemic syndrome (aHUS) and age-related macular degeneration (AMD) (65). In a different context of disease occurrence in which FH is involved, the ability of this regulator for immune surveillance is utilized by an extended array of microbes in order to achieve immune invasion (66).

1.7.1 Age-related macular degeneration

Age-related macular degeneration (AMD) is the leading cause of loss of vision in the Western world for individuals of 50 and above. Complement dysregulation plays an important role in its pathogenesis. Specifically, the Tyr402His polymorphism that lies in FH-7 as well as in FHL-1 (which equates to the first 7 CCP modules of FH) increase the risk of contracting AMD up to seven-fold (in homozygotes). FH-7 encompasses one of the polyanion-binding sites in FH. Early stage AMD is characterized by the formation of drusen, lipoproteinaceous deposits, between the retinal pigment epithelium (RPE) and Bruch's membrane. The Y402H polymorphism has also been linked to increased risk of MPGN II (see below) (65).

1.7.2 Membranoproliferative glomerulonephritis type II

Membranoproliferative glomerulonephritis type II (or Dense Deposit Disease, DDD), is a rare condition characterized by dense deposits within the glomerular basement membrane (GBM). These deposits consist to a large extent of C3c. In addition, C3 serum levels are distinctively low. The disease has been associated with mutations in the *CFH* gene. Reduced levels of FH in plasma, or deficiency in its ability to regulate the alternative complement pathway, result in excessive complement activity with repercussions distinct in the GBM. In both MPGN II and AMD, accumulations of complement-containing debris occur within the kidney and eye, respectively (65).

1.7.3 Atypical haemolytic uraemic syndrome

Haemolytic uraemic syndrome is characterized by hemolytic anemia and often ends in acute renal failure. The typical form is associated with infections by contrast with the atypical form, aHUS, which occurs in 5-10 % of HUS patients. aHUS is connected to mutations or polymorphisms in the *CFH* gene, the majority of which cluster in the C-terminal region, FH-19-20, *i.e.* in the aforementioned polyanion and C3b-binding sites. In this case, the mutant FH molecules appear to be capable of normal activity in plasma but cannot protect host cells from lysis (thus uncontrolled complement activation takes place on cell surfaces). This is consistent with an important role of the C-terminus in distinguishing host versus foreign surfaces (65).

1.7.4 Pathogen invasion

Many pathogens disguise themselves from the complement system by mimicking properties of the host cell surfaces or by binding to FH and other regulators thus facilitating their escape from complement-mediated destruction. The number of invading agents that have been shown to bind to FH is constantly increasing. These include amongst others *Streptococcus pneumoniae (S. pneumoniae), Streptococcus pyogenes (S. pyogenes)* and *Neisseria meningitidis (N.*

meningitidis). One of the strategies employed, for example, by *S. pneumoniae* for escaping the immune system is by expressing a major virulence factor: the surface protein PspC which binds to FH and minimizes C3b deposition on pneumoccoci (66).

The following sections (section 1.9 onwards) discuss the topic of complement evasion by pathogens with particular focus on *S. pneumoniae* and more specifically on the interaction with FH via its surface –attached virulence factor, the PspC protein.

1.8 Concluding remarks

Complement is an important and potent immune surveillance system which, however, has to be tightly regulated by natural fluid-phase and steady-state regulators. FH is the main regulator of the always activated alternative pathway and ensures the protection of the host by acting both in fluid-phase and on cell surfaces.

The crystal structures of FH-1-4-C3b, and FI shed light on the corresponding interactions and the implicated mechanisms. The crystal structure of FH-19-20 with C3d and subsequent model extrapolation to C3b provided a structural basis for engagement by FH of C3b on a self surface. Binding studies demonstrated the cooperative binding of the two termini to C3b and polyanions such as GAGs. This bivalency is facilitated by a bent-back conformation that FH adopts. Indeed FH-8-15 is arranged in a compact, unique manner, unlike any other seen in the RCA family and enables the two functional ends to come close. FH can then perform its biological tasks. The only high-resolution structure available for the central region is that of FH-12-13. It is evident that the unique architecture of the 8-15 part needs to be further elucidated at an atomic level as the question of how it promotes overall biological activity of FH still remains. Taking this one step further, it may also help understand more about the molecular basis of the use of FH by pathogens in order to avoid their complement-mediated elimination.

1.9 The use of the complement system as an evasion strategy by pathogens

To survive in the harsh environment of the bloodstream nearly all pathogens (of vertebrates) have evolved means to evade complement-mediated destruction. Some pathogens take this strategy further and actively subvert the complement system to gain entry into a host cell where they can avoid triggering complement activation altogether while they divide and multiply. Many pathogens produce proteins that directly interfere with the initial steps of activation, *i.e.* they act to block C3b amplification. Others target the terminal pathway by, for example, blocking MAC assembly (67). Some produce enzymes that proteolytically degrade complement components, or they produce proteins that interact with host complement proteins and thereby block their actions (68).

Furthermore, some bacterial proteins remain on the bacterial cell surface where they recruit and tether host regulators such as FH, FHL1, and C4 binding protein (C4BP) (67). This strategy is advantageous to the microorganism as the natural regulators are already present at high levels. Figure 1.17 schematically illustrates these and other examples employed by pathogens for protection from complement. (68), (69). Examples of pathogens that bind to complement inhibitors include *S. pneumoniae*, *S. pyogenes*, *Escherichia coli (E. coli)*, *Haemophilus influenzae (H. influenzae)*, *Neisseria gonorrhoeae (N. gonorrhoeae)* and *N. meningitidis* (69).

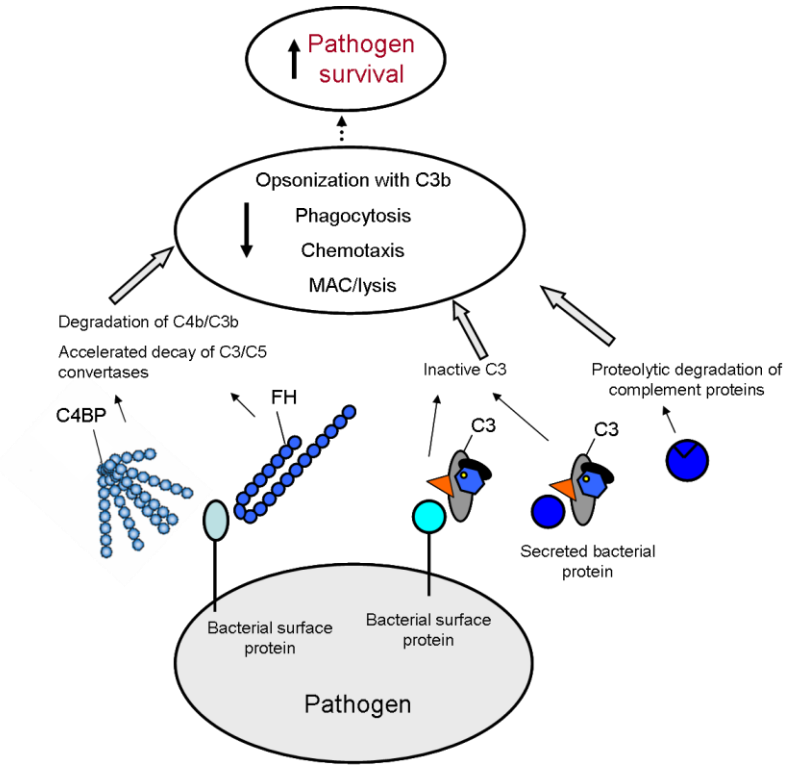


Figure 1.17 Examples of how pathogens can interfere with complement in order to increase their survival chances. The regulators FH and C4BP can be recruited onto the pathogen surface where they inhibit formation of the convertases and participate in FI-mediated degradation of C3b and C4b. Furthermore secreted pathogen proteins can degrade complement components. All these lead to incompetent opsonophagocytosis arising from decreased C3b opsonization and chemotaxis. Finally lysis through the formation of the membrane attack complex (MAC) may also be compromised (Figure and legend adapted from (69)).

1.10 *S. pneumoniae*

S. pneumoniae (pneumococcus) is a Gram-positive bacterium that resides naturally in the human upper respiratory tract flora in low numbers (70). However it is also an important pathogen causing several serious infections such as pneumonia, meningitis and septicemia (71). Severe infections have been associated with high mortality rates and death. Upon occurrence, clearance of such infections is dealt with by both the innate and the adaptive immune responses (72). Upon activation by the presence of pneumococci, complement-mediated phagocytosis takes place (73), (74). The major phagocytic cells of the gas-exchange areas of the lung are the human alveoral macrophages. Pneumococci bind to their surface, become internalized and are killed by phagolysosomes and it has been demonstrated that this process and its outcome can be compromised in the absence of opsonization with complement (75).

1.11 Evasion of complement by pneumococci

The classical and the alternative pathways of complement are predominantly responsible for clearance of pneumococci (73), (74). The bacteria are Gram-positive hence inherently resistant to MAC-mediated lysis. Since multiple constituents of the complement system target *S. pneumoniae* for destruction, these bacteria have evolved more that one evasion mechanism. They can compromise the alternative pathway C3 convertase (see below) and thereby reduce opsonisation with C3b (74). In addition, pneumococci can secrete toxins that attack host tissue and are able to activate the classical pathway (76), (77). Finally they can recruit host regulators onto their surface thus escaping complement-mediated opsonophagocytosis (78), (79), (80).

1.11.1 Virulence factors

Several virulence-associated factors are involved in the aforementioned mechanisms, which provide resistance to the immune response. Some of these will be mentioned in this section.

Pneumolysin is an intracellular toxin that does not become secreted until lysis of the pneumococci occurs by the action of autolysin (an enzyme that can break down the peptidoglycan layer of the cell wall). Mutant strains that do not express pneumolysin are non-virulent. This potent toxin can, for example, attack host tissue and cells, it can activate the classical pathway "with a concomitant reduction of serum opsonic activity" (77) and cause inflammatory responses; this being key to pneumoccocal-induced pathogenesis (77).

Cell-surface virulence proteins produced by pneumococci have become the subject of many studies due to their potential use as vaccine antigens. These proteins can be divided into three groups: choline-binding proteins, lipoproteins and proteins that are attached to the cell wall via an LPXTG motif (where X is any amino acid). For example, choline-binding proteins include PspA (Pneumococcal surface protein A), and PspC (Pneumococcal surface protein C). PspC has been reported to bind the polymeric immunoglobulin receptor that transports secretory IgA (81), to C3 (82) as

well as FH (83), (84), (85). The study of the interaction of PspC with FH is a main objective of the current project, and therefore will be discussed in more detail below.

1.11.2 Pneumococcal surface protein C (PspC)

The *pspC* locus of *S. pneumoniae* is polymorphic. Sequencing data from 43 strains enabled the organization of the variants into 11 groups. Groups 1-6 are the classical PspC proteins, anchored to the bacterial cell wall by a C-terminal choline-binding domain; groups 7-11 utilize a C-terminal LPXTG motif (86) for this purpose. The domain organization of the classical PspC proteins is as follows (see Fig. 1.18): a leader peptide and an N- terminal domain is followed by one or two repeats designated R1, R2 then a proline-rich region (87) and, finally, the choline-binding domain. In for example, strain D39 of capsular serotype 2, there is a 96-residue N-terminal domain followed by R1 and R2, each of 107 amino acids, then a 97-residue proline-rich region (Fig. 1.20) and lastly, the C-terminal choline-binding domain made up from 10 repeats of 20 amino acids (88). The R domains of PspC are highly conserved in sequences when compared to other pneumococcal strains (89). In strain TIGR4, for example, the R1 and R2 domains each comprise three α-helices that altogether form a flat, raft-like structure. Furthermore, the N-terminal region of PspC proteins is α-helical and has a stable fold. Thus, the whole of the NR1R2 region appears to consist of α-helices (89).

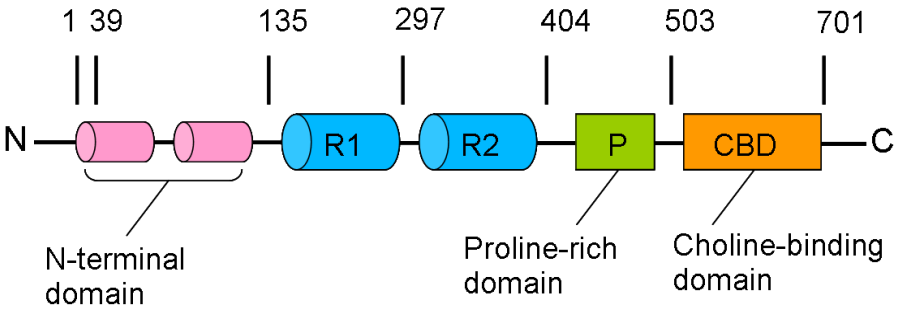


Figure 1.18 Schematic illustration of the domain organization of the PspC protein of strain D39, capsular serotype 2. The amino acid positions are indicated on the right of each domain (87).

Figure 1.19 shows a multiple sequence alignment of the N-terminal domain of PspC of strain D39 (which is the focus of this study) with PspC allelic variants.

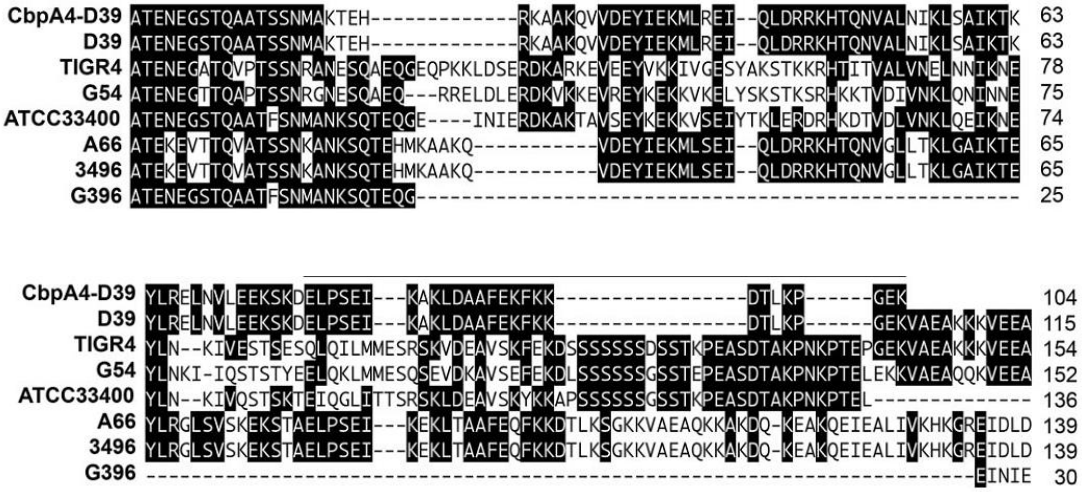


Figure 1.19 Sequence alignment of the N-terminal PspC of strain D39 with the PspC allelic variants as indicated on the left of the alignments. Identical residues are shown shaded. Dashed lines indicate the gaps that were introduced for optimal alignment. Amino acid positions are marked on the right according to the relative distance from the first amino acid following the signal sequence (Figure and legend from (88)).

1.12 FH binding motif on PspC

The ability of PspC to bind FH is an important mechanism by which pneumococci can resist complement-mediated phagocytosis. Furthermore, it has been demonstrated that FH retains many functional abilities even after engaging with PspC (90). As described in subsequent paragraphs, several recent studies have set out to pinpoint the domains of PspC and FH that mutually contribute to this key interaction.

Initial studies used Western blotting to map FH binding region on PspC. These experiments used a fragment corresponding to residues 1-225 of an isogenic strain of D39, a fragment corresponding to residues 1-445 and another spanning residues 225-445 of PspC. All fragments were incubated with FH. Only the 225-445 fragment did not bind, hence the FH-binding domains in PspC must reside within residues 1-225 of the α-helical N-terminal domain (90); this includes the N-terminal domain extended to the R1 repeat, figure 1.18).

A subsequent study further localized the FH-binding region. Using Western blots, the wild-type D39 strain showed strong binding to FH while a mutant that lacked the 89-amino acid region of the N-terminus showed no appreciable binding. Furthermore mutants in which the R1, R1/R2 or proline-rich regions were deleted suggested these regions did not contribute to binding. Thus the N-terminal 89-amino acid region is essential for FH-binding activity. Similar conclusions arose from use of epi-immunofluoresence microscopy of intact pneumococci (88).

By isothermal titration calorimetry (ITC) a K_D of <1 nM was observed for the interaction of FH with the N-terminal region of PspC and a formation of a 1:1 complex between the two was inferred (88). Spot-synthesized peptide analysis identified a 12-amino acid stretch within the N-terminal (ALNIKLASAIKTK, amino acids 88-99 in PspC of strain D39) which is required for binding to FH, however a synthetic peptide representing this 12-amino acid motif alone was incapable of mediating binding (88).

A separate study suggested the proline-rich region was important for PspC binding to FH. Specifically, ELISA assays demonstrated that a construct (residues 1-445) lacking the C-terminal choline-binding domain showed two-threefold higher affinity for FH than a shorter construct (residues 1-371) additionally lacking the proline-rich region (91). Finally a study using PspC deletion mutants and surface plasmon resonance narrowed down the FH-binding site to the N-terminal residues 38-158, of PspC (87).

1.13 PspC-binding regions on FH

Several studies have attempted to identify the FH domain/s responsible for interacting with PspC. This section will briefly present the findings to date, particularly with respect to PspC from the virulent *S. pneumoniae* strain D39 (since this was also the focus of the present work).

In 2002, Duthy *et al*. localized the PspC binding site on FH to FH-8-15 by ELISA assays while FH-1-7 and 16-20 did not show discernible binding (91). In order to further localize the site within the FH-8-15 region FH-10-12, 8-12, 8-13 and 8-14 (which were produced recombinantly and expressed in *Pichia pastoris* (*P.*

pastoris)) were tested for binding. Only 8-13 and 8-14 showed interaction. The binding, however, for 8-14 was roughly two-thirds the binding of 8-13. It was therefore suggested that the additional CCP 14 in the 8-14 construct may be obstructing the binding site residing in FH-13 while on the other hand it seems that addition of CCPs 14-15 to the 8-13 construct can remove this inhibition resulting in the increased binding demonstrated by FH-8-15. Based on the comparable binding shown by 8-15, it appeared that some complex higher order structure might exist in the 13-15 region of FH. Such a conclusion was further reinforced by whole-cell ELISA, which confirmed that the presence of both modules 13 and 15 was required in the constructs in order for optimum binding to occur. This could mean (and remains to be tested) that the two sites act together to form the complete site while displacing the weaker-binding 8-14 from PspC, or that a second dite exists in FH-15. Finally in this study, heparin, C3b and sodium chloride did not inhibit the FH-PspC interaction. This result indicates that the heparin (GAG)-binding site and C3b-binding site do not overlap with the PspC-binding site (91).

In 2004, Dave *et al.* concluded that the binding site for PspC (strain D39) was located within FH-6-10, and that FH did not lose its functional activities in the presence of PpsC. In more detail, ligand dot-blotting assays localized the PspC-binding site on FH between modules 6 and 10 (including FH-6 and FH-10) because all recombinant truncated constructs that contained FH-1-10 bound to PspC and the FH mutant lacking FH-6-10 (rFH Δ6-10) failed to show binding. Moreover, inhibition of haemolysis was observed both upon addition of FH as well as a mixture of FH and PspC (90).

In 2007 Hammerschmidt *et al.* investigated the effect of bacterial cell-surface bound FH on pneumococcal adhesion to host cells. Bacteria-bound FH enhanced adherence to host cells (both epithelial and endothelial) especially when the amount of capsular polysaccharide was low. With respect to identifying the PspC-binding site in FH, SPR was used. For this experiment PspC2.1 (produced by strain ATCC33400, capsular type 1) was immobilized and the FH deletion constructs FH-1-7, 8-11, 11-15, 8-20 and 15-20 were tested for binding. It was surmised that two binding sites exist within FH located in FH-8-11 and FH-19-20 (88).

A follow-up study by Agarwal *et al*. in 2010 investigated how pneumococci can be ingested by epithelial cells through a novel mechanism involving FH acting as a bridge between pathogen and host cells. Specifically, it was proposed that *S. pneumoniae* recruits FH via sites in 8-11 and 19-20 recognized by PspC. The FH polyanion-binding region in FH-19-20 then mediates binding of FH to host cells. Through the dynamics of the actin cytoskeleton and kinase activities of signal transduction molecules, pneumococci use a FH-integrin complex for invasion (92). (FH can interact with integrin receptors, CR3 (92)).

1.14 Concluding remarks

Pathogens have developed sophisticated mechanisms to escape their destruction by the host's complement system. Virulence determinants displayed on the surface of pathogens can recruit regulators like FH in order to survive. Examples of such invaders - other than *S. pneumoniae* - are *N. meningitidis* (93), *S. pyogenes* (94), *H. influenzae* (95) and *Borrelia burgdorferi (B. burgdorferi)* (96). Figure 1.20 illustrates where these bacteria can bind to the host regulator FH.

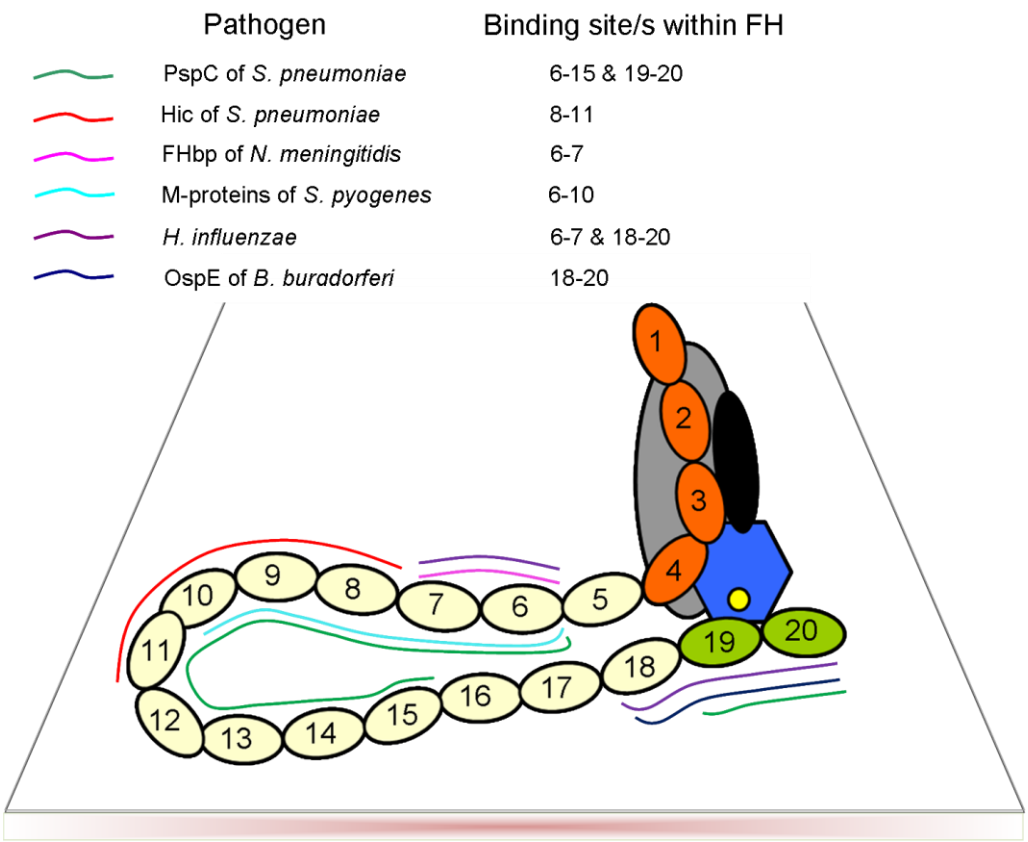


Figure 1.20 Schematic summary of some of the pathogens that recruit the complement regulator FH. FH is shown bound to C3b on a surface. CCPs within FH that bind to the pathogens are indicated color-coded for each pathogen (83), (93), (94), (95), (96).

S. pneumoniae is a harmful pathogen and one mechanism by which it can evade complement is by recruiting FH. It expresses PspC, a surface protein and chief virulence factor which binds FH through its α-helical N-terminal domain. Regions in FH that mediate this interaction have been mainly localized in the regions FH-6-10 and 19-20. Figure 1.21 schematically summarizes the PspC-binding sites on FH. In summary, although partially conflicting data exist regarding the location of the PspC binding site with the complement regulator FH, it appears that most of the sites cluster around the central region of FH and clearly away from the FH-1-4 regulatory region.

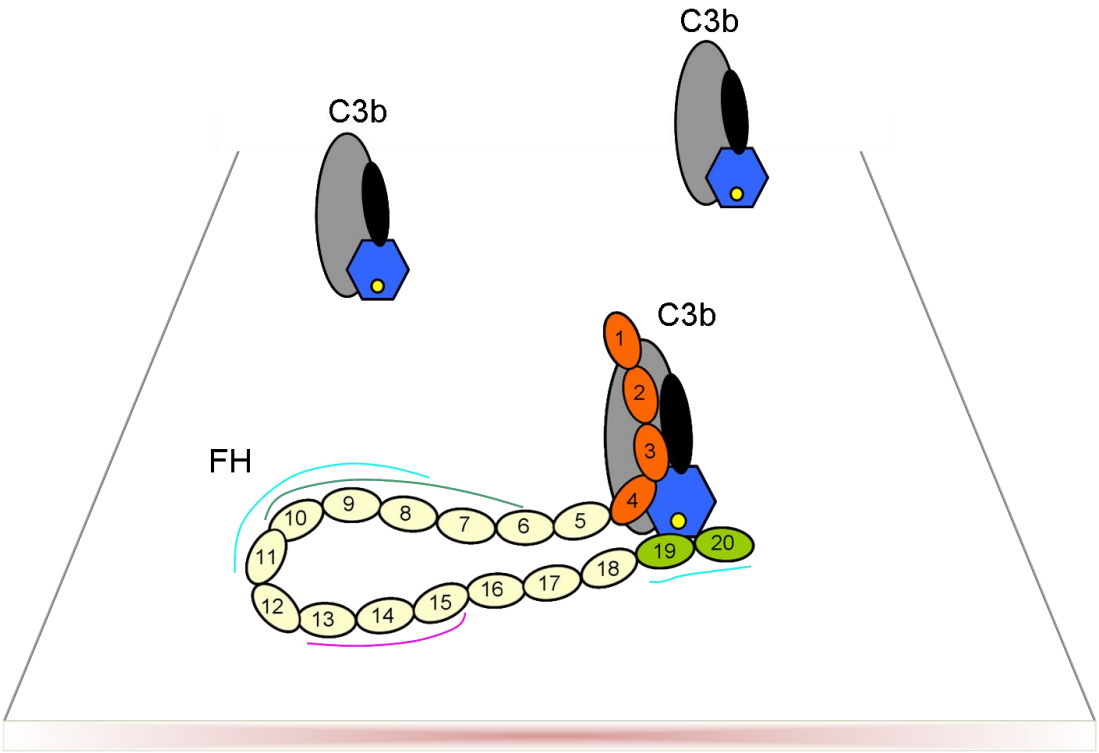


Figure 1.21 Schematic summary of the locations of the PspC binding sites on FH as determined by three studies in particular. FH is shown in complex with C3b in this representation bound on a surface. FH-6-10 constitutes one region encompassing the PspC binding site as demonstrated by Dave *et al.* (90) and is highlighted by a green curvy line in the figure. FH-8-11 and 19-20 correspond to the PspC-binding-site-containing domains as determined by Hammerschmidt *et al.* (87) and are highlighted by cyan curvy lines. Finally FH-13-15 emerged as the binding site containing region by studies from Duthy *et al.* (91) and is highlighted by a pink curvy line.

Overall it appears (see Figs. 1.20 and 1.21) that important bacterial protein-binding regions in FH are domains 6-7, the middle compact region and the C-terminal 19-20 region. This observation has mechanistic implications. For example, the FHbp of *N. meningitidis* binds with a very high affinity to FH-6-7 and mimics the mechanism by which host cells regulate complement activation on their surface using amino acid residues instead of charged sugars (FH-7 contains a GAG-binding site) (93). The FH C-terminal region contains binding sites for both C3b and polyanionic components while the central region is devoid of such recognition sites. With respect to PspC of *S. pneumoniae* the binding sites are located in regions 6-15 and 19-20 (as determined by three studies in particular) following the same pattern as the other pathogens (87), (90), (91).

Understanding the exact molecular mechanisms of the interactions of PspC with FH could prove to be a significant step towards developing new vaccines.

Indeed, the FH-binding fragment of PspC has been investigated as a vaccine candidate and its use in combination with other antigens could lead to successful vaccine development (97). What is more, light will be shed on how FH operates in the context of several biological processes, thus expanding our understanding on the works of this structurally unique and increasingly important complement regulator.

1.15 Overall goals of the project

The complement regulator, FH, controls the positive feedback cycle of C3b production both in the fluid phase and on cell surfaces. While in the former case, FH prevents futile depletion of C3 and factor B, in the latter case it is critical for protecting self surfaces from harm by complement. It is not surprising that sequence variations in FH are linked to disease or that FH and rationally engineered analogues have therapeutic potential. The overall goal of this project is to understand how the central region of FH may promote the overall biological efficacy of the intact protein.

FH engages simultaneously with GAG/polyanions and C3b (and its fragments) on a self surface through binding sites in FH-1-7 and 19-20. The central region, FH-10-15, contains the longest linker lengths and smallest modules and forms a compact, potentially hinge-like structure. This region has not been as fully characterized as the terminal regions and forms the main focus for the current project.

One of the sophisticated strategies by which pathogens circumvent the complement system is by recruiting FH. While most seem to latch onto FH-6-8, which contains a GAG-binding site, or FH-18-20, which contains both a GAG and a C3b-binding site, some bacterial proteins bind to FH-8-15 which does not include such regulatory regions. This intriguing observation led to a second focus of the current work – the study of how PspC interacts with the central region of FH. The figure below, (Fig. 1.22), summarizes the abovementioned points.

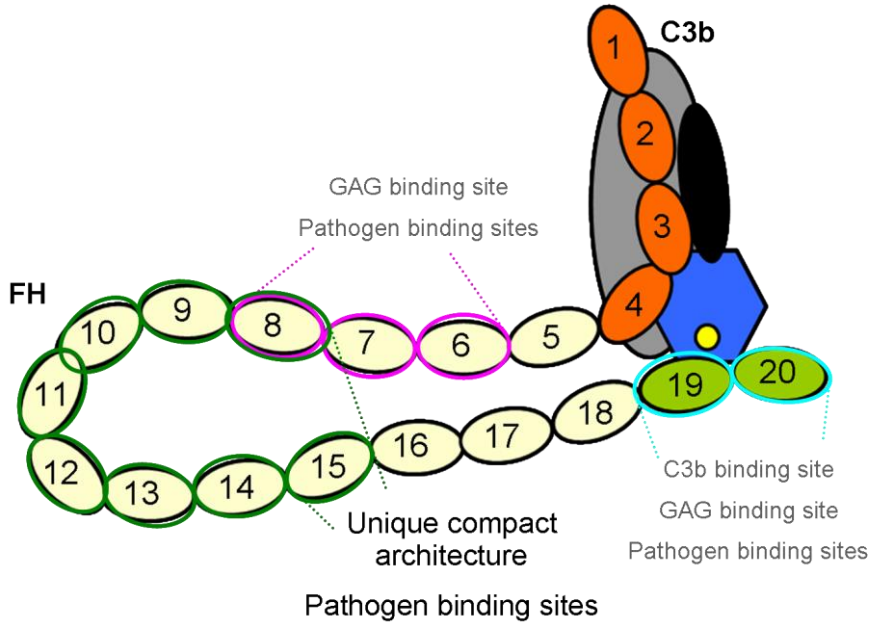


Figure 1.22 Representation of the FH molecule bound to C3b. CCPs that have a functional role and to which pathogens bind are highlighted as follows: FH-6-8 in pink circles, FH-8-15 in green circles and FH-19-20 in cyan circles.

1.16 Specific aims

A combination of high-resolution data combined with lower-resolution techniques will be used to allow more accurate modeling of the spatial organization of the CCPs of the central region. Biophysical techniques will also be used to investigate an intuitively attractive hypothesis, that binding of the pathogen protein PspC to FH may stabilize a preferred conformation that optimizes its complement-regulatory activities.

1. The first aim is to determine the solution structures of FH-10-11 and FH-11-12 so as to determine whether these are "V-shaped", like FH-12-13, and to allow a model of FH 10-13 to be created by concatenation of bimodule structures.

2. SAXS data will be collected on those fragments as well as FH-10-12, providing orthogonal validation of intermodular tilt angles and the spatial arrangement of modules. The NMR-derived structural data will be used in order to distinguish between two existing SAXS-derived models of FH-10-15 structure.

3. The interaction between FH and PspC will be studied to resolve conflicting reports in the literature regarding which FH modules are involved. Furthermore the interactions of complexes consisting of FH-PspC constructs with the C3 fragments C3b, C3c and C3d will be investigated. The objective here is to determine whether PspC really binds to two sites in FH and to test the hypothesis that it stabilizes a conformation of FH that has enhanced ability to regulate complement.

CHAPTER 2

Materials and Methods

Overview

This study entailed production of recombinant proteins corresponding to segments of complement regulator FH, and of *S. pneumoniae* virulence factor PspC (see section 2.5, Fig. 2.9). The FH-derived protein segments were produced in the yeast expression host *P. pastoris* whereas the PspC constructs were produced from genetically modified *E. coli*. Procedures for cloning, digestion and transformation reactions, culture growth and manipulation of protein samples are described in this chapter, along with methods for NMR-based structural studies and SPR-based functional assays.

2.1 DNA amplification, manipulation, cloning and transformation

2.1.1 Polymerase chain reaction

The polymerase chain reaction (PCR) was used to amplify the coding sequences for FH constructs 4-6, 10-11 and 14-15, for sequencing reactions and for screening colonies for the expected insert. The PCR was also used to effect insertion of a new stop codon in the case of a PspC construct that possessed originally four residues beyond the native C-terminus.

The FH constructs were amplified using primers designed by Dr C. Q. Schmidt (formerly of this lab) and are listed in Table 2.1. Primers used for the touchdown PCR reaction were purchased from Sigma-Genosys. M13 and alpha-factor/AOX1 primers were purchased from Invitrogen.

OLIGONUCLEOTIDE NAME	DNA SEQUENCE	T_m (^oC)	PURPOSE	RESTRICTION SITE
FH 10-*F*	AA**CTGCAG**GAG AAAGAGAATGC GAACTTCCTAA A	63.41	Touchdown PCR	***Pst*1**
FH 11-*R*	GC**TCTAGA**CTA CTCCACAATAC ACACTGGTAAA GTTG	64.27	Touchdown PCR	***Xba*1**
FH 4-*F*	AA**CTGCAG**GAG AAATTTCATGC AAATCCCC	62.86	Touchdown PCR	***Pst*1**
FH 6-*R*	GC**TCTAGA**CTA TTTTCTGAGGC ATGGTACTGC	64.03	Touchdown PCR	***Xba*1**
FH 14-*F*	AA**CTGCAG**GAA TACAATTATGC CCACCTCCAC	64.47	Touchdown PCR	***Pst*1**
FH 15-*R*	GC**TCTAGA**CTA AAGGCCTTCAC ACTGAGGTG	64.25	Touchdown PCR	***Xba*1**
pCR® 4Blunt-TOPO®, M13, *F*	GTAAAACGACG GCCAG	57.55	Sequencing/ Screening PCR	**n/a**
pCR® 4Blunt-TOPO®, M13, *R*	CAGGAAACAGC TATGAC	50.6	Sequencing/ Screening PCR	**n/a**
Alpha-factor *F*	GGGGATTTCGA TGTTGCTGTTTT G	70.77	Sequencing/ Screening PCR	**n/a**
AOX1 *R*	CCGGTCTTCTC GTAAGTGCC	65.82	Sequencing/ Screening PCR	**n/a**

Table 2.1 Primer oligonucleotide sequences. The restriction-enzyme sites are highlighted in bold. T_m corresponds to the calculated melting temperature of each primer without the restriction enzyme site. *F* and *R* refer to the forward or reverse direction of the primers, respectively. The M13 primers were used to screen colonies after cloning into TOPO vector and sequencing. The alpha-factor *F* and AOX1 *R* were used for screening after sub-cloning into the pPICZ*α*B vector and sequencing.

In the case of the PspCN construct with an N-terminal His-tag and a C-terminal GSGC tetrapeptide artifact (His-tag PspCN-GSGC), the primers used for site-directed mutagenesis and subsequent sequencing are listed in Table 2.2. All primers were purchased from Invitrogen. Melting temperatures were calculated using the Sigma-Genosys primer calculator: (http://www.sigma-genosys.com/calc/DNACalc.asp).

OLIGONUCLEOTIDE NAME	SEQUENCE	Tm (^{o}C)	PURPOSE
PspCN *F*	GAAACCGGGTGAAAAATAGTAGGGATCCGGCTGCTAACC	84	Site-directed mutagenesis
PspCN *R*	GTTAGCAGCCGGATCCCTACTATTTTTCACCCGGTTTC	84	Site-directed mutagenesis
T7 promoter	TAATACGACTCACTATAGGG	64	Sequencing
T7 terminator	GCTAGTTATTGCTCAGCGG	68	Sequencing

Table 2.2 Primer oligonucleotide sequences used for the site-directed mutagenesis and sequencing reactions for His-tag PspCN-GSGC. Melting temperatures are shown for each primer as calculated using the Sigma-Genosys online primer calculator.

2.1.1.1 Generation of coding sequences for the FH constructs

For generation and amplification of sequences coding for FH constructs, 1 µl aliquots of 10 µM solutions of forward and reversed primers were mixed with 2 µl of the template DNA (40 ng/µl), 5 µl Pfu reaction buffer 10x (Fermentas), 1 µl deoxynucleotide triphosphates (dNTPs) (2.5 mM, Invitrogen), and 1 µl Pfu polymerase (2.5 U/ml, Fermentas) in a microcentifuge (PCR) tube and then the volume was adjusted to 50 µl by adding 39 µl of molecular biology-grade H_2O. A gene for full-length FH (prepared from human cDNA by Dr C. Q. Schmidt) served as template DNA.

The PCR cycling parameters summarized in Table 2.3 were employed. In each case a subsequent PCR reaction was performed on the purified product of the first PCR, using an identical protocol, in order to obtain higher yields of the target DNA.

STEP	TEMPERATURE (^{o}C)	TIME (min:s)	REPETITIONS
Initialization	95	1:00	1x
Denaturation	95	0:30	14x
Touch-down annealing steps	T_m Gradient -0.5^{o} R =3^{o}/s	0:30 R =3^{o}/s	
Elongation	72	1.00 min/kb DNA	
Denaturation	95	00:30	20x
Annealing	T_m -4.86	00:30	
Elongation	72	1.00 min/kb DNA	
Final Elongation	72	5	-
Hold	4	hold	-

Table 2.3 PCR cycling parameters. Based on the melting temperatures, T_m, of the primers the annealing temperatures are determined after having taken into consideration the restriction sites. Because the resulting T_ms of forward and reverse primer sequences were different, the annealing temperature was set to that corresponding to the lower primer T_m, *i.e.* 62.86 ^{o}C.

2.1.1.2 Screening of *E. coli* colonies by PCR for the FH constructs

Following overnight growth, construct FH-10-11 and seven colonies for each of constructs FH-4-6 and FH-14-15 were picked and cells were transferred to microcentrifuge tubes that contained 20 μl EB buffer (10 mM Tris-HCl, pH 8.5) as well as to new plates. Samples were heated in the PCR machine to 95 ^{o}C for 3 min for the purpose of bacterial lysis and solubilisation of the plasmid DNA. In separate microcentrifuge tubes, 2 μl of the above samples were mixed with 1.5 μl of forward and reversed primers (10 μM, see Table 2.1) and 5 μl of PCR Master Mix (Promega) reagent. PCR was performed according to the scheme in Table 2.4.

STEP	TEMPERATURE (°C)	TIME (min:s)	REPETITIONS
Initialization	95.0	01:00	1x
Denaturation	96.0	00:30	29x
Annealing	50.0	00:30	
Elongation	60.0	01:00/kb DNA	
Hold	4.0		—

Table 2.4 PCR screening cycling parameters

2.1.1.3 Sequencing reactions for the FH constructs

Samples for DNA sequencing were submitted to the GenePool (Ashworth Laboratories, School of Biological Sciences, University of Edinburgh). For submission of the PCR product, 1 µl of DNA (200 µg/ml) was mixed with 1 µl (10 µM) of the appropriate forward or reverse primers (M13 *F* and *R* supplied with the TOPO cloning kit, or alpha-factor *F* and AOX1 *R* in the case of pPICZαB vector), 4 µl Big Dye (Applied Biosystems) and 14 µl molecular biology-grade H_2O (Sigma), to a final volume of 20 µl. The PCR was performed according to the scheme in Table 2.5 and resultant products submitted for sequencing.

STEP	TEMPERATURE (°C)	TIME (min:s)	REPETITIONS
Initialization	95	00:30	1x
Denaturation	96	01:00	24x
Annealing	50	00:30	
Elongation	60	04:00	
Hold	4	hold	—

Table 2.5 PCR sequencing cycling parameters

2.1.1.4 Site-directed mutagenesis for the PspC construct

For site-directed mutagenesis of His-tag PspCN-GSGC, the QuickChange Site-Directed Mutagenesis Kit (Stratagene) was used and the instructions that came with it were followed. In short, 1 μl of template DNA (coding for His-tag PspCN-GSGC, 125 ng/μl) was mixed with 5 μl Pfu buffer, 1 μl Pfu turbo polymerase (Agilent Technologies), 1.3 μl of forward and reverse primers (10 μM), 1 μl deoxynucleotide triphosphates (dNTPs), 2.5 mM (Invitrogen), 1.5 μl DMSO (Stratagene) and the final volume was adjusted by adding 37.9 μl molecular biology-grade H_2O. The PCR cycling parameters summarized in Table 2.6 were employed. The aim of this PCR was insertion of a stop codon at a specific point; therefore 18 cycles were repeated according to manufacturer's (Stratagene's) instructions. Following the PCR reaction, the products were digested with *Dpn*I (20,000 U/ml, NEB) in order to digest the parental DNA template (it targets 5'-Gm^{6}ATC-3' and is specific for methylated and hemimethylated DNA). 5 μl were added and the reaction was incubated for 2 h at 37 ^{o}C.

STEP	TEMPERATURE (^{o}C)	TIME (min:s)	REPETITIONS
Initialization	95	2:0	1x
Denaturation	95	1:0	18x
Annealing	55	1:0	
Elongation	68	10:00	
Hold	4	hold	–

Table 2.6 PCR cycling parameters for QuickChange site-directed mutagenesis method.

2.1.1.5 Sequencing of the PspC construct following site-directed mutagenesis

For sequencing of the His-tagged PspCN construct after site-directed mutagenesis, the following procedure was followed: The digested PCR product from the mutagenesis reaction was transformed into *E. coli* BL21(DE3) cells (Lucigen) and 4 colonies were subjected to purification of plasmid DNA after an overnight growth at 37 ^{o}C. Of the resultant DNA, 300 ng were mixed with 4 μl BigDye, 1 μl

T7 promoter or T7 terminator primer (to final concentration of 10 μM each, Invitrogen) and adjusted to a final volume of 20 μl with molecular-biology grade H_2O (Sigma). The PCR cycling parameters summarized in Table 2.7 were subsequently employed.

STEP	TEMPERATURE (oC)	TIME (min:s)	REPETITIONS
Initialization	95	1:00	1x
Denaturation	96	00:30	30x
Annealing	50	00:30	
Elongation	60	1:00	
Hold	4	hold	–

Table 2.7 PCR sequencing cycling parameters for the PspCN construct following site-directed mutagenesis.

2.1.2 Agarose gel electrophoresis for nucleic acids

A 1.0-1.2 % agarose solution in 50 ml Tris-acetate-EDTA (TAE) buffer was prepared (1-1.2 g agarose was weighed) and the solution was brought just to the boil until the agarose had completely dissolved, using a microwave oven. It was then cooled to 50-60 ^{o}C prior to adding 5 μl of SYBR-safe stain, purchased from Invitrogen®, to a final concentration of 0.5 μg/ml. The rack with the agarose gel was placed in a gel tank and was completely covered with 1x Tris-acetate-EDTA (TAE) buffer. The DNA samples were then mixed with loading buffer and placed in the wells. Alongside the samples, molecular-weight markers were loaded, covering the required size range. The gel was run at 125 V for approximately 40 min and then illuminated with an ultraviolet lamp to view the DNA bands.

2.1.3 DNA purification and plasmid DNA extraction

The QIAprep Miniprep or QIAfilter Maxiprep kits for plasmid extraction and the QIAquick Spin kit for extraction and purification of DNA from agarose gels were

all purchased from QIAGEN. All procedures were carried out following the supplied protocols.

2.1.4 DNA quantification - estimation of DNA concentration by UV spectroscopy

Nucleic acids have an absorption maximum at 260 nm therefore quantification of the DNA samples was achieved by UV spectroscopy. (Eppendorf BioSpectrometer, Eppendorf). Samples contaminated by protein or single-stranded DNA/RNA have a maximal absorption at 280 nm. In order to asses the purity of the DNA samples the ratios $A_{[260]}/A_{[280]}$ and $A_{[260]}/A_{[320]}$ were calculated. The higher the ratio the less contaminated the sample, and values of 1.8 for $A_{[260]}/A_{[280]}$ and 2.0-2.2 for $A_{[260]}/A_{[320]}$ signified adequately pure DNA samples.

2.1.5 TOPO cloning reaction for the FH constructs

To clone the PCR product into the plasmid vector, pCR 4Blunt-TOPO, the Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen) was used. For the cloning reaction, 2 µl of the fresh PCR product was mixed with 0.5 µl of the salt solution and 0.5 µl of vector solution (supplied with the cloning kit). The sample was centrifuged and left at room temperature for 15 min. The 3-µl resultant mixture was used to transform One Shot® Top 10 chemically competent *E. coli* cells (Invitrogen), following the protocol and utilizing the reagents provided in the TOPO cloning kit. From the transformed cells, aliquots of 50 and 100 µl were spread on Luria-Bertani (LB) (see Appendix section A for composition of media) agar plates containing kanamycin (kanamycin sulfate, Gibco) (100 µg/ml, w/v) and were left to grow overnight at 37 ^{o}C.

2.1.6 Restriction enzyme single- and double-digests for the FH constructs

Double digestion was performed on the TOPO vector carrying the required DNA insert to facilitate further sub-cloning of the insert into the pPICZaB vector.

Single digestion or, alternatively, linearization was performed on the pPICZaB vector containing the desired DNA insert to facilitate integration into the *P. pastoris* genome. The reactions were preformed using NEB enzymes and buffers as follows: For the double digests, 4.8 µl of NEB buffer 3 (10 % v/v) were mixed with 0.5 µl of BSA (1 % v/v), 48 µl of a solution of the TOPO vector containing the DNA insert, 1.2 µl restriction endonuclease 1, *Xba*I (20 U/µl), and 1.2 µl restriction endonuclease *Pst*I (20 U/µl) yielding a total reaction volume of 55.7 µl. For the single digest/linearization, 20 µl of NEB buffer 3 (10 % v/v) were mixed with 2 µl BSA (1 % v/v), 200 µl pPICZaB vector containing the DNA insert, 4 µl *Sac*I (20 U/µl) restriction endonuclease and 4 µl H_2O yielding a total reaction volume of 230 µl. Each solution was mixed and incubated at 37 ^oC for 3 h and 6 h, for the double- and single-digests, respectively. The restriction endonucleases were inactivated at the temperature and time indicated by NEB. The restriction digests were analyzed by agarose gel electrophoresis to ensure success of the digestion. Resultant DNA fragments were purified and concentrated by phenol-chloroform extraction followed by ethanol precipitation or by gel extraction using the QIAquick gel-extraction kit (QIAGEN).

2.1.7 Ethanol precipitation for the FH constructs

The volume of the DNA solution was estimated and the salt concentration was adjusted by adding 3 M sodium acetate, pH 5.2, to a final concentration of 0.3 M. The solution was mixed and 2-2.5 volumes of ice-cold ethanol (100%) were added, then the sample was mixed and placed on ice for 1 h. The DNA was recovered by centrifugation at 13000 rpm for between 15 and 30 min at 0 ^oC (Eppendorf Centrifuge 5415 R). The supernatant was carefully removed with an automatic pipette and the tube was filled halfway (750 µl) with 70 % aqueous ethanol. Then it was centrifuged (Eppendorf Centrifuge 5415 R) at 13000 rpm for 5 min at 4 ^oC and again the supernatant was carefully removed. Finally, the open tube was stored at room temperature until the pellet had dried. It was then resuspended in the desired volume of water.

2.1.8 Phenol chloroform extraction for the FH constructs

Phenol-chloroform extraction was performed as follows. An equal to the sample volume of phenol: chloroform: isoamyl alcohol (in a ratio 25:24:1, saturated with 10 mM Tris, pH 8.0, 1 mM EDTA) was added. The contents of the tube were mixed until an emulsion formed. The mixture was then centrifuged on a bench-top microfuge (Eppendorf Centrifuge 5415 R) at 13000 rpm for 2 min at room temperature. The aqueous phase was transferred into a new tube and the interface and the organic phase were discarded. The abovementioned steps were repeated until no protein was visible at the interface of the organic and aqueous phase. An equal volume of chloroform was added and steps 2-4 were repeated. Finally, the DNA was recovered by ethanol precipitation.

2.1.9 Ligation reactions for the FH constructs

Approximately 50 ng of cut pPICZαB vector were combined with a three-fold molar excess of DNA insert to a volume of 10 μl. Then 7 μl of molecular biology-grade H_2O were added as well as 2 μl 10x Quick ligation buffer, (NEB) and 1 μl T4 DNA ligase (400 U/μl, NEB) to a final volume of 20 μl. The solution was mixed, centrifuged briefly and incubated at room temperature for 15 min. Following this, transformation of *E. coli* Top10 cells was performed (see below) and cells were spread on LB zeocin plates (25 μg/ml) and left to grow overnight at 37 ^{o}C.

2.1.10 Transformations of plasmids into chemically competent *E. coli* cells and culture growth for the FH constructs

The chemically competent cells that were used were purchased from Invitrogen and were stored at -80 ^{o}C. A vial of frozen competent cells was thawed on ice then 3 μl of the TOPO cloning reaction mixture was added and the contents were gently mixed by swirling the tube. It was then incubated on ice for 30 min, placed in a pre-warmed water bath at 42 ^{o}C without shaking for 90 s and then returned to ice for 5 min. An aliquot of 250 μl of super optimal growth with catabolite repression

(SOC) medium (Invitrogen), pre-warmed to room temperature, was added and the tube was capped tightly then shaken in a horizontal position in an incubator set to 37 ^{o}C, 200 rpm, for 1 hour. Antibody-selective LB agar plates, containing either 25 μg/ml zeocin or 100 μg/ml kanamycin, were pre-warmed for 30 min at 37 ^{o}C. Then 100-150 μl aliquots of the transformed cell cultures were streaked onto the plates that were incubated overnight at 37 ^{o}C in a non-shaking incubator.

Subsequently, single colonies were used to inoculate 5 ml of the appropriate LB media containing zeocin or kanamycin at the above concentrations. These were incubated overnight at 37 ^{o}C with shaking at 250 rpm. The resulting cultures were used for any required DNA analysis.

2.1.11 Digestion of the pE-SUMOPro Kan vector and the genes of interest for the PspC constructs

Synthetic genes for PspCN and PspCNR1 were designed with an *Xba*I site (to allow cloning into the pE-SUMOPro Kan vector (LifeSensors) digested only with *Xba*I) and were supplied (codon optimized), cloned into the pMA-T vector, by GeneArt.

The pE-SUMOPro Kan vector was digested as follows: 2 μl of vector (1 μg) were mixed with 2 μl 10x NEB buffer number 4, 1 μl *BsaI* HF (for high-fidelity) (20,000 U/ml, NEB), 0.5 μl 100x BSA (100x, 10 mg/ml) and 15 μl molecular biology-grade H_2O. The reaction was incubated at 37 ^{o}C for 1.5 h followed by heat inactivation of the enzyme (65 ^{o}C for 20 min). The digestion of the two synthetic genes was carried out as follows: 17 μl of dissolved gene (1.7 μg) were mixed with 2 μl 10x NEB buffer number 4, 0.5 μl *BsaI HF* (NEB), 0.5 μl *Xba*I (20,000 U/ml, NEB) and 0.5 μl 100x BSA. The reaction was incubated at 37 ^{o}C for 1.5 h and the enzyme was subsequently heat inactivated by incubation at 65 ^{o}C for 20 min.

2.1.12 Ligation of the PspC inserts into the pE-SUMOPro Kan vector and transformation of *E. coli* Top10 cells

An aliquot of 1 μl of linearized pE-SUMOPro Kan vector was mixed with 3 μl of linearized synthetic gene plasmid, 1 μl 5x DNA-dilution buffer (Rapid DNA Dephos and Ligation Kit, Roche), 5 μl 2x T4 DNA ligation buffer (Roche) and 0.5 μl T4 DNA Ligase (5 U/ml, Roche). The reaction was incubated at room temperature for as long as required for the Top10 chemically competent cells to thaw (approximately 10-20 min). Of the ligation reaction, 2 μl were then used to transform the cells as follows: an aliquot of 2 μl of ligation reaction was added to the cells and mixed gently by swirling the tube. The cells were then incubated on ice for 30 min and were subsequently heat-shocked at 42 ^{o}C for 30 s. To the cells was added 250 μl SOC medium and then they were placed in a shaking incubator set at 37 ^{o}C for 1 h. Samples of 20 and 200 μl of the reaction were then plated out on antibody-selective LB agar plates containing 30 μg/ml kanamycin, and allowed to grow overnight at 37 ^{o}C.

2.1.13 Inspection of colonies for correct insert for the PspC constructs cloned into pE-SUMOPro Kan vector

Ten colonies for PspCN and 10 colonies for PspCNR1 were chosen from the LB agar plates to inoculate 5 ml of LB broth containing 30 μg/ml kanamycin and with the same loop (for each of the colonies) to inoculate 3 ml of LB broth containing 100 μg/ml carbenicillin (carbenicillin disodium salt, Apollo Scientific Ltd.). The cultures were left to grow overnight at 37 ^{o}C. Cultures that did not grow on carbenicillin then underwent purification of plasmid DNA. The resultant DNA was then digested as follows: 17 μl of the DNA were mixed with 2 μl of 10x NEB buffer number 4, 1 μl *Xba*I (20,000 U/ml) and 0.5 μl 100x BSA. The reaction was incubated at 37 ^{o}C for 90 min and was finally subjected to agarose gel electrophoresis to enable identification of the colony or colonies that contained the desired insert.

2.1.14 Transformation of constructs cloned into pET-15b in BL21 (DE3) *E. coli* cells

A glycerol stock of the DNA for the His-tag PspCN-GSGC version of PspCN transformed in the *E. coli* strain BL21 (DE3) prepared by Dr A. Herbert (the DNA was originally ordered from GeneArt), as well as the freshly prepared DNA for the His-tag PspCN construct (following site-directed mutagenesis using the DNA form the His-tag PspCN-GSGC construct as a template) were used for re-streaking on LB agar plates containing 50 μg/ml carbenicillin. The plates were left to grow overnight at 37 ^{o}C. Single colonies were used to inoculate 5 ml of LB broth containing 50 μg/ml carbenicillin and the cultures were left to grow overnight in a shaking incubator set at 37 ^{o}C. The plasmid DNA was subsequently purified and used to transform BL21 (DE3) *E. coli* cells (Lucigen). The transformation protocol followed (provided by Lucigen) was as follows: The cells were thawed on ice, 1 μl of the 15 and 47 ng/μl DNA stock were added to 40 μl of cells for each DNA concentration, and the whole was stirred briefly with the pipette tip. The cells were then incubated on ice for 30 min, heat shocked for 45 s at 42 ^{o}C and returned to ice for 2 min. A volume of 960 μl of express recovery medium was then added and the tubes were placed in a shaking incubator at 37 ^{o}C for 1 h. Subsequently, 50 and 100 μl aliquots of the transformation reaction were plated on LB agar plates containing 50 μg/ml carbenicillin and were left to grow overnight at 37 ^{o}C. Single colonies were used to inoculate LB media for the large-scale expressions.

2.1.15 Production of the recombinant FH constructs in *P. pastoris*

2.1.15.1 Overview of *P. pastoris* as an expression system

P. pastoris is a yeast expression host that was used to produce recombinant FH-derived fragments. For expression in this system, the insert is placed under the control of a very strong, inducible promoter (explained below). As a eukaryotic expression system, *P. pastoris* has important advantages over prokaryotic ones, such

as the abilities to carry out post-translational modifications, glycosylation, correct protein folding, proteolytic processing and disulfide bond formation (98), (99).

As a methylotrophic yeast, *P. pastoris* utilizes methanol as a sole carbon source through a methanol metabolic pathway comprising a set of enzymes. The initial step in the metabolism of methanol occurs when the enzyme alcohol oxidase (AOX) induces the oxidation of methanol to formaldehyde and the formation of hydrogen peroxide - a harmful by-product. To avoid the toxic effects of hydrogen peroxide this process takes place in a specialized organelle called the peroxisome that isolates this toxic compound from the rest of the cell. The second enzyme, catalase, is mostly located in the peroxisome and serves to rapidly catalyze decomposition of hydrogen peroxide to less reactive gaseous oxygen and water molecules. Alcohol oxidase has a poor affinity for O_2 and in order to compensate for that the yeast cells produce very high amounts of this enzyme. Consequently the promoters concerned are exceptionally strong and are exploited to obtain high yields of inducible recombinant proteins.

This organism has two genes that code for alcohol oxidase AOX, *AOX1* and *AOX2*. *AOX2* is 97 % similar to *AOX1*. However the resultant growth following uptake of methanol, and hence the activity of alcohol oxidase, is much faster when the *AOX1* gene is responsible. When methanol is the carbon source for the growth of *P. pastoris*, *AOX1* is expressed at very high levels while being regulated at the level of transcription. Regulation involves two mechanisms: repression/derepression and induction. In the first case, the presence of glucose represses transcription even if methanol is present. For this reason, growth on glycerol (derepression) is recommended during protein production work, which is performed before induction with methanol which is essential for *AOX1* expression. There are two different methanol-metabolizing *P. pastoris* phenotypes: MUTs and MUT^+. The strain which phenotypically is MUT^S (methanol utilization slow) results from the loss of the *AOX1* gene and hence the majority of alcohol oxidase activity. This strain relies on *AOX2* and grows on methanol slowly. The strain with phenotype MUT^+ (methanol utilization plus) refers to the wild-type ability of strains to grow on methanol as a sole carbon source. The *P. pastoris* strain used for this project was KM71H (Invitrogen) and belongs to the MUT^S type.

With *P. pastoris* as an expression system, heterologous proteins can either be expressed intracellularly or secreted. The fact that *P. pastoris* secretes low levels of native proteins, combined with the minimal media used for its growth, results in the majority of total protein in the media being the recombinant protein of interest, effectively serving as a first purification step. As far as post-translational modifications are concerned, *P. pastoris* has the potential to perform most modifications associated with higher eukaryotes. These include O- and N-linked glycosylation and disulfide bridge formation, the latter being of particular relevance to the current work since CCP modules have two disulfide bonds per CCP.

To utilize the strongly inducible *AOX1* promoter for gene expression, the plasmid pPICZαB was used as a vector. This plasmid carries the DNA encoding the alpha-factor secretion signal peptide that is fused to the insert to ensure the target protein is secreted. Using this vector ensure that once the heterologous protein-encoding sequence has been integrated into the genome of *P. pastoris*, the *AOX1* promoter is responsible for expression and expression is modulated through controlled methanol feeds during cell cultures (98), (99), (100), (101).

In some cases the N-terminus of the expressed protein may include, usually, one or two non-native Glu-Ala repeats. Two steps are required for the processing of the alpha-factor signal sequence. The first step involves cleavage by the *KEX2* gene product, with the final Kex2 cleavage occuring between Arg and Gln in the following sequence: Glu-Lys-Arg* Glu-Ala-Glu-Ala (* is the site of cleavage) (102). The second step involves cleavage of the Glu-Ala repeats by the *STE13* gene product (103). There are cases where Ste13 cleavage is not efficient thus leaving the aforementioned non-native N-terminal repeats.

2.1.15.2 Preparation of *P. pastoris* cells

The following procedure was employed: Starting from a re-streaked plate of *P. pastoris* strain KM71H cells, 10 ml of Yeast Extract Peptone Dextrose (YPD) medium were inoculated with a single colony and left to grow overnight at 30 ^{o}C. A batch of 250 ml of fresh YPD medium was poured into a 1-L flask, inoculated with

0.1-0.5 ml of the overnight culture and the cells were again left to grow overnight until they reached an OD_{600} of ~1.3. The cells were centrifuged for 5 min at 1500×g, 4 ^{o}C, and then resuspended in 250 ml of ice-cold sterile water. The cells were again centrifuged for 5 min, 1500×g, 4 ^{o}C, and resuspended in 125 ml of ice-cold sterile water. The cells were again centrifuged as above and resuspended in 10 ml of ice-cold 1 M sorbitol. Cells were again centrifuged as above, and resuspended in 0.5 ml of ice-cold 1 M sorbitol.

2.1.15.3 Transformation by electroporation

The following protocol was used: An 80-µl aliquot of the cells were mixed with 5-10 µg of linearized DNA, in 5-10 µl of sterile water, and transferred into an ice-cold 0.2-cm electroporation cuvette. The cuvette with the cells was incubated on ice for 5 min. The cells were pulsed for 5 s (settings on the BioRad gene pulser II: charging voltage 1500 V, capacitance 25 µF, resistance 200 mΩ). Immediately after pulsing, 1 ml of ice-cold 1 M sorbitol was added and the contents of the cuvette were transferred into a 15-ml tube. The tube was incubated for 3 h at 30 ^{o}C without shaking. Aliquots of 150 and 350 µl of the contents of the 15-ml tube were spread on Yeast Extract Peptone Dextrose Medium (plus Sorbitol) (YPDS) plates containing 100 µg/ml and 300 µg/ml zeocin and were left to grow at 30 ^{o}C for three days until colonies were visible.

2.1.15.4 FH constructs 8-9, 13-14, 10-15 and 8-15

P. pastoris clones (in strain KM71H) expressing FH-8-9, FH-13-14 and FH-10-15 were provided by Dr C. Q. Schmidt. These were expressed either in shaker flasks or in fermentors as outlined in the sections below (2.1.15.5 and 2.1.15.6) and were purified as described in section 2.2.6. In the case of unlabeled production of FH-10-15 and FH-8-15 crude *P. pastoris* supernatant also provided by Dr C. Q. Schmidt was the starting point and was thereafter purified. The results of all purification procedures are presented in Chapter 3.

2.1.15.5 Mini-scale expression (trial expression) of recombinant proteins in *P. pastoris*

This procedure was carried out as follows: A single colony was used to inoculate 5-10 ml buffered minimal glycerol (BMG) in a 50-ml tube and was placed in a shaking incubator set to 250-300 rpm, 30 ^{o}C for two days. The cells were spun at 1500×g for 10 min (Sorvall RC6 *PLUS*, rotor: SH-3000), and resuspended in 3 ml buffered minimal methanol (BMM) and the 50-ml tube was returned to the incubator once the temperature was set to 15 ^{o}C. During this induction period, 100% methanol was added every 24 h to maintain a final concentration of 0.5 % (v/v). This was carried out for two days. On the third day, methanol to 0.5 % (v/v) was added in the morning followed by methanol to 1.2 % (v/v) in the evening. The cells were harvested by centrifuging at 4000×g for 20 min (Sorvall RC6 *PLUS*, rotor: SH-3000). Prior to purification, phenylmethanesulfonyl fluoride (PMSF, 100 mM in isopropanol) and ethylenediaminetetraacetic acid (EDTA, 500 mM, pH 8.2) were added to final concentrations of 0.5 mM and 5 mM respectively.

2.1.15.6 *P. pastoris* recombinant protein expression in shaker flasks

The procedure was carried out as follows: A single colony was used to inoculate 25 ml BMG in a 500-ml sterile baffled flask and this was placed in a shaking incubator (250-300 rpm), 30 ^{o}C. After 24 h 500 ml of fresh BMG was added in an empty, sterile 2-L baffled flask and inoculated with 25 ml of the overnight culture, returned to the incubator, and left to grow for two-three days. The cells were spun for 5 min at 1500×g (Sorvall RC6 *PLUS*, rotor: SLA-3000), resuspended in 125 ml of BMM and the flask was returned to the incubator, once the temperature was set to 15 ^{o}C. During this induction period, 100% methanol was added every 24 h to maintain a final concentration of 0.5-1.2 % (v/v). The cells were harvested by centrifuging at 4000×g for 20 min (Sorvall RC6 *PLUS*, rotor: SLA-3000). Prior to purification PMSF and EDTA were added to final concentrations of 0.5 mM and 5 mM respectively.

2.1.15.7 *P. pastoris* recombinant protein production in fermentor

Expression in fermentor cultures, in ^{15}N- or ^{15}N and ^{13}C-enriched media, was needed for structural studies by NMR in the case of FH-10-11 and for recording ^{1}H, ^{15}N-HSQC/TROSY spectra in the case of FH-13-14 and FH-10-15. Compared to a shaker flask there are advantages that come with a fermentor growth. First, typically expression levels are higher in the controlled environment that fermentation provides, such as control of temperature, pH, agitation and oxygen, especially for proteins that are secreted. Second, the level of transcription from the *AOX1* promoter can be greater when the cells are fed with methanol at growth-limiting rates compared to excessive feeding in shaker flasks. Finally, when methanol is metabolized, high levels of oxygen are needed for the successful expression of the gene of interest (101). Hence, the possibility of monitoring and adjusting accordingly the oxygen in a fermentor culture is advantageous.

Two fermentation protocols were followed for the production of ^{15}N- and ^{15}N, ^{13}C-labeled protein samples, which differed in the initial buffer volume and composition and the isotopic labeled reagents added. All fermentations were carried out in two-liter cylindrical Bioflow 3000 (New Brunswick scientific) fermentor vessels.

2.1.15.7.1 Preparation of *P. pastoris* cells that contained the gene of interest

From the glycerol stock, under sterile conditions, a small amount of cells was spread on a YPD plate and left to grow at 30 ^{o}C until colonies appeared (this took approximately three-four days). A single colony was used to inoculate 10 ml of fresh BMG, and this was placed in a shaking incubator (30 ^{o}C, 250 rpm) for 24 h. A batch of 150 ml of fresh BMG were poured into a sterile 500-ml baffled flask, inoculated with the 10 ml and placed back in the incubator for approximately 24 h. The cells were spun for 5 min at 1500×g and resuspended in 20 ml of fermentor media or potassium phosphate buffer, 100 mM, for ^{15}N-labeled or ^{15}N, ^{13}C-labeled samples respectively.

2.1.15.7.2 Preparation of fermentor growth media; fermentor vessel; initial set-up of Bioflow 3000.

This preparatory step started on the day at which the 150 ml of BMG were inoculated with the 10-ml first culture, and was carried out as follows:
The compositions of the ^{15}N-enriched and ^{15}N, ^{13}C-enriched media are listed in Table 2.8.

	^{15}N-enriched media	**^{15}N, ^{13}C-enriched media**
Basal salts	0.75 g $CaSO_4$ x 7 H_2O 12 g MgSO4 x 7 H_2O 8 g K_2SO_4 20 ml 100 % glycerol 60 ml 1 M potassium phosphate buffer, pH 6.0 740 ml dH20	0.56 g $CaSO_4$ x 7 H_2O 9 g $MgSO_4$ x 7 H_2O 6 g K_2SO_4 – 45 ml 1 M potassium phosphate buffer pH 6.0 555 ml dH20
Final volume	800 ml	600 ml

Table 2.8 Basal salts and media used for the fermentation buffers.

The media were added into the fermentor vessel. Following this, the vessel was assembled, the probes for the dissolved oxygen (DO), pH, as well as the filter for the air, were attached and the whole was autoclaved. After having cooled down, the DO probe was attached to the main unit of the Bioflow 3000 and was left to charge overnight. In addition, the supply for air was turned on so that air flowed continuously through the sterile filter, and the impeller speed was set to 200 rpm. The pH probe was also attached, as was the temperature probe and a feeding line for base (2 M KOH) used to maintain pH levels stable at 5.0. To the vessel were added 0.5 ml Antifoam 204 (Sigma Aldrich) and 3.5 ml filter-sterilized (0.2-μm cut-off) fermentation trace salts (PTM1 salt, Amresco), and the temperature was set to 30 ^{o}C.

After the overnight period, for ^{15}N-enrichment, 7-8 g ^{15}N-$(NH_4)_2SO_4$ (*ISOTEC*, Sigma-Aldrich) were dissolved in roughly 20 ml dH_2O and sterile-filter injected (20 μm filter) into the vessel media. For ^{15}N, ^{13}C-enriched media, 6 g of ^{15}N-

$(NH_4)_2SO_4$ was dissolved in 20 ml distilled H_2O and 15 g ^{13}C-Glucose in about 45 ml distilled H_2O and both were again sterile-filter injected (20 μm filter) into the media.

Once temperature and pH conditions had stabilized, the relative percentage of dissolved oxygen was normalized to 100. The impeller speed was set (between 300 and 1000 rpm) so as to preserve a relative oxygen level of 40 % during the period of growth and induction. (Note that when the cells are metabolizing the carbon source, the dissolved oxygen levels drop below 40 % initially, then increases and remains relatively stable until the cells have consumed the food, at which point the dissolved oxygen level increase above 40 %. To compensate, the impeller speed increases to near 1000 rpm when dissolved oxygen levels are low but decreases and remains stable at 300 rpm when they are high.) After these preparations the vessel was inoculated with the cells.

For the ^{15}N-enriched fermentations: After inoculating and incubation for 16 h, based on inspection of the impeller speed and dissolved oxygen level, the cells were fed 40 ml autoclaved glycerol accompanied by 0.48 ml PTM1 trace salts (both were sterile-filter injected). Once the glycerol source had been exhausted, roughly after one day, the temperature was reduced to 15 ^{o}C and allowed to stabilize. Subsequently methanol feeds (accompanied by addition of PTM1 salts) were instigated by an initial feed of 0.5 % of culture volume. Induction was then promoted by the first of two 1.2-1.3 % feeds. At this point it was important not to overfeed with methanol so as to avoid poisoning the cells. The total induction period was about three days.

For the ^{15}N, ^{13}C-enriched fermentation: once having inoculated and after one-day growth period, cells were fed 1.0 g of ^{13}C-glycerol (*ISOTEC*, Sigma-Aldrich) diluted in 3 ml distilled H_2O (sterile-filter injected). The importance of this step derives from the fact that all of the glucose should have been consumed before the addition of methanol as glucose represses the *AOX1* promoter. Once glycerol had been consumed, ^{13}C-MeOH (*ISOTEC*, Sigma Aldrich) feeds followed, 0.5 % (of the culture volume) for the first feed, 1.25 % for the second, and 1.25 % for the third, with the timing based on careful observation of agitation rates and dissolved oxygen levels.

In both cases, the cell culture was subsequently harvested by centrifugation at 8000×g for 20 min. Then, the supernatant was decanted into clean centrifuge bottles

and was spun again at 8000×g for 30 min (Sorvall RC6 *PLUS*, rotor: F9S-4x1000y). Sterile filtration of the supernatant ensued and finally EDTA (500 mM, pH 8.2) and PMSF (100 mM in isopropanol) were added prior to purification (to final concentrations of 5 mM and 0.5 mM respectively).

2.1.16 Production of the recombinant PspC constructs in *E. coli*

2.1.16.1 Overview of *E. coli* as an expression system

E. coli is the most frequently used prokaryotic expression system as its fast growth rate; its good levels of protein production and its simple growth conditions make it an attractive option for production of heterologous proteins.

For the production of proteins a plasmid is transferred into an *E. coli* cell that contains a chromosomal copy of the gene that encodes T7 RNA polymerase. This allows regulated expression from the T7 promoter (104). The host strains are lysogens of bacteriophage DE3 and carry a DNA portion that includes the *lacI* gene, the *lacUV5* promoter and the gene for T7 RNA polymerase (105), (106). Once a DE3 lysogen is formed, the *lacUV5* promoter directs transcription of the T7 RNA polymerase. This occurs upon induction with isopropyl-β-D-thiogalactopyranoside (IPTG) (107). When IPTG is added to a bacterial culture that has reached mid-log phase (before the cells have reached saturation), T7 RNA polymerase production is induced. IPTG, which is a lactose analog, operates by binding to the *lacI* repressor. It induces its dissociation of repressor from the operator thus allowing transcription from the promoter. When sufficient amounts of the polymerase are produced, it binds to the T7 promoter and transcribes the gene of interest.

For the purpose of this study, two types of vectors were used, pET-15b and pE-SUMOPro Kan (LifeSensors). One of the features of the pET-15b vector is that it contains the highly efficient ribosome-binding site from the phage T7 protein which is used for expression of target genes that lack their own ribosomal site. The vector encodes an N-terminal His-tag sequence followed by a thrombin cleavage site and it has three cloning sites. In the present study the gene was provided by GeneArt and was cloned using the *Nde*I and *BamH*I site. The use of pE-SUMOPro Kan vector

results in the cloning and expression of the gene of interest as a fusion to SUMO. It has been shown that protein fusions enhance the expression and solubility of recombinant proteins (108). SUMO is an 11-kDa protein and a member of the ubiquitin-like protein family that regulates several cellular processes (109). SUMO will attach to Lys residue sidechains and its use results in change in protein function and activity. The advantages associated with fusing the target protein to SUMO are increased expression level, solubility, and the option of easy subsequent cleavage of SUMO which is facilitated by its recognition by SUMO protease (Ulp) (110), (111). This final option ultimately results in the production of native protein with no additional residues.

2.1.16.2 Optimization of *E. coli* strain and induction conditions for constructs cloned in pE-SUMOPro Kan vector

The DNA preparations that tested positive for the correct insert were individually used to transform different *E. coli* strains. Two *E. coli* strains were used, namely, BL21 (DE3) and BL21 (DE3) –pLysS. The procedure followed for the transformation of the BL21 (DE3) –pLysS strain was identical to that already described for the *E. coli* TOP10 cells in section 2.1.10. For the transformation of the BL21 (DE3) strain the protocol provided by Lucigen was followed (already described in section 2.1.14). Thus 20, 100 and 200 µl of BL21 (DE3) and 20, 200 µl of (DE3) –pLysS were plated on LB agar plates containing 30 µg/ml kanamycin and were left to grow overnight at 37 ^{o}C. The test inductions were performed in an identical manner for both strains and both PspC constructs. For the first set of conditions which were 3 h induction periods at 37 ^{o}C, single colonies from the transformations described above were used to inoculate 5 ml of LB broth containing 30 µg/ml kanamycin, and the cultures were placed in a shaking incubator set at 37 ^{o}C until the OD_{600} reached ~0.8. Samples of 0.5 ml were taken at this point. The cultures were induced with IPTG (Promega) to a final concentration of 1 mM and left to grow for 3 h at 37 ^{o}C. Following this, two 0.5-ml samples were taken for testing of total, and soluble or insoluble, recombinant protein products. The 0.5-ml samples taken prior to induction and one of the two 0.5-ml samples taken after induction for testing

the total fraction of expressed proteins were processed as follows: the samples were spun at 13000 rpm for 5 min and the pellets were resuspended in 250 μl SDS-containing PAGE loading buffer. The samples were then incubated at 95 ^{o}C for 5 min during which time they were mixed at 1-min intervals. They were then spun for 5 min at 13000 rpm and loaded on an SDS-PAGE gel. The remaining 0.5-ml samples were subjected to the following procedure: They were spun for 5 min at 13000 rpm and the supernatants were discarded. The pellets were resuspended in 125 μl BugBuster MasterMix (Novagen), the tubes were placed in a shaking incubator set at 37 ^{o}C for 10 min, they were then spun again and the supernatants were transferred to separate tubes to which 0.5 ml of SDS-PAGE loading buffer was added for loading on an SDS-PAGE gel in order to test the soluble fraction of the expression. The pellets from that stage were resuspended in 0.5 ml BugBuster MasterMix (Novagen), were incubated for 5 min at 95 ^{o}C, while being mixed at 1-min intervals. They were subsequently spun for 5 min at 13000 rpm, the supernatants were discarded and the procedure for lysing pellet was repeated. The final pellets were resuspended in 250 μl SDS-PAGE loading buffer for loading on an SDS-PAGE gel in order to determine the insoluble fraction of expressed protein.

The second set of induction conditions were overnight induction at 18 ^{o}C. The procedure followed was identical to that of the first set of conditions with the apparent exception of temperature and induction time, and the IPTG which was added at final concentration of 0.25 mM.

2.1.16.3 *E. coli* recombinant protein expression in shaker flasks

Single colonies form the transformations of the BL21 (DE3) cells were used to inoculate 10 ml LB broth containing 30 μg/ml kanamycin or 50 μg/ml carbenicillin. The cultures were left to grow overnight in a shaking incubator set at 37 ^{o}C. Aliquots of 1-2.5 ml of culture were then used to inoculate 500 ml of LB broth, which contained 30 μg/ml kanamycin or 50 μg/ml carbenicillin. The cultures were left to grow in a shaking incubator at 37 ^{o}C. Once the OD_{600} reached ~0.6, the temperature was reduced to ~20 ^{o}C and the cells were induced with IPTG to final concentrations of 0.25 mM (in case of the constructs having been cloned in the pE-

SUMOPro Kan vector) and 0.5 mM (in the case of the constructs having been cloned in pET-15b vector). The cultures were left to grow overnight, were then harvested by centrifugation at 6000×g for 15 min and the supernatant was discarded.

After harvesting the *E. coli* cultures, the cell pellets were weighed. BugBuster MasterMix (Novagen) was then added at a ratio of 5 ml/g pellet accompanied by EDTA-free protease inhibitors cocktail (protease inhibitor cocktail set V, EDTA-Free, Calbiochem). The cells were mixed until a uniform resuspension was achieved and were then placed on rolling platforms (Roller SRT1, Stuart Scientific) at 4 ^{o}C for 2 h. They were subsequently sonicated at 15-s intervals for a total of 5 min (Soniprep 150, Sanyo, at amplitude of 10 microns). They were finally spun at 16000×g for 30 min and the soluble lysate was diluted and subjected to the first purification step (Co^{2+}-affinity chromatography).

2.2 Manipulation and purification of the protein samples

2.2.1 Estimation of protein concentrations by UV Spectroscopy

Protein samples have a maximum UV absorbance at 278-280 nm, hence UV spectroscopy was utilized for this purpose (Eppendorf BioSpectrometer, Eppendorf). The ratios $A_{[280]}$ /$A_{[260]}$, and $A_{[280]}$/$A_{[320]}$ were used to asses the purity of the samples. The protein concentrations were calculated according to the Beer-Lambert's law equation: A= εbc, were ε is the molar extinction coefficient (calculated by ExPASy, ProtPram tool), b the pathlength and c the concentration.

2.2.2 Concentration of protein samples

2.2.2.1 Concentration by trichloroacetic acid (TCA) precipitation (for SDS-PAGE)

A 20 % (v/v) TCA solution was added to the sample, at an equal volume. The mixture was incubated on ice for 1 h. It was then spun in a bench-top micro-centrifuge (Eppendorf Centrifuge 5415 R) at 13000 rpm for 30 min at 4 ^{o}C. The supernatant was discarded. An aliquot of 300 μl of ice-cold acetone was added and

the whole was mixed. It was then spun at 13000 rpm for 5 min at 4 ^{o}C. The pellet was air-dried at 95 ^{o}C for 5-10 min. The sample was finally resuspended in SDS-containing gel-loading buffer in preparation for loading onto an SDS-PAGE gel.

2.2.2.2 Concentration by membrane concentrators and buffer exchange

With respect to all the single and double-isotopically labeled protein samples, concentration was achieved solely by this method. It was also used for the various buffer exchanges and was exclusively used for all the PspC constructs. The sample was carefully decanted in 6- or 20-ml Vivaspin concentrators (Sartorius Mechatronics UK Ltd) with the appropriate molecular weight cut-off membrane (5000-10000) Da. In the case of concentration it was then centrifuged at 4000×g, until the volume was ~400 μl at which point the sample was transferred into a microcentrifuge tube. Finally, the membrane of the concentrator was washed with ~100-200 μl of the same buffer, and this volume was added to the same tube. In the case of buffer-exchange, the sample was again centrifuged at 4000×g to a volume of 500-600 μl. Buffer was then added up to a volume of 5-6 ml and the sample was again centrifuged under the same conditions. This was repeated three times.

2.2.3 Sodium dodecyl –sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gels (4-12 % Bis-Tris Gel) were purchased from (NuPAGE® Novex Bis-Tris Mini Gels). Depending on the various numbers of wells and gel thicknesses, samples were mixed with the appropriate volumes of loading buffer (4x), and reducing agent (10x) and were heated at 95 ^{o}C for 2 min (in a heating block) prior to gel loading. Protein markers with molecular weights in the 10-150 kDa range (Precision Plus Protein Standards, Kaleidoscope, Bio-Rad) were run alongside the samples. The gel was run in 1x NuPAGE MES-buffer (Invitrogen) at 180 V for approximately 50 min. The gel was then washed with water. Visualization was achieved by staining the gels with BioSafe Coomassie stain (Bio-

Rad®, CA, USA) for at least 1 h. Following this, the stain was discarded, and the gel was left to destain for a minimum of 3-6 h or overnight.

2.2.4 Enzymatic cleavage of N-linked glycans for the FH constructs

Endo H_f enzyme (New England BioLabs) (a recombinant fusion of endoglycosidase H and maltose binding protein) was used for the removal of the mannose-rich N-linked carbohydrate residues from the proteins. Endo H_f can cut between the two first GlcNAc molecules, leaving the protein with one GlcNAc moiety per glycosylation site. Variable numbers of units of Endo H_f (10^6 U/ml) were used at 37 ^oC for 1-3 h, depending on the amount of sample to be treated.

2.2.5 Cleavage of SUMO from the SUMO-ylated proteins

Cleavage of SUMO-PspCN and SUMO-PspCNR1 was performed by adding 100 μl of SUMO protease (or ULP1-Ubiquitin like protease, 10 U/μl, produced in the lab by Dr A. Herbert). The protease was added in a dialysis tube containing the protein to be cleaved and was left at 4^o C overnight in 20 mM potassium phosphate buffer, 500 mM NaCl, 20 mM imidazole buffer, pH 7.0.

2.2.6 Protein Chromatography

The solvents and buffer reagents used for the purification of the protein constructs were purchased from Sigma-Aldrich and Fisher Scientific. All columns used were purchased from GE Healthcare. The gel-filtration column (for size-exclusion chromatography) was pre-packed. For the initial purification step of the majority of the FH constructs, cation-exchange chromatography was employed.
4-5 ml SP-Sepharose were packed into graduated Econo-Pac polypropylene columns (BioRad Laboratories), or 30 ml SP- or Q-Sepharose in an XK 26 column (Pharmacia Biotech) for the cation and anion exchange chromatography respectively. Prior to loading on the ion-exchange column the supernatant was diluted accordingly and pH adjusted appropriately (see below) so as to achieve binding to the column.

For size-exclusion chromatography, the sample was concentrated and buffer-exchanged using membrane concentrators.

For ion-exchange chromatography, based on the pI of the proteins as calculated by ExPASY ProtParam tool, the pH of the supernatant and the column were set ~2-3 units lower or higher than the pI in the case of cation and anion exchange, respectively. This was to ensure the protein would be positively or negatively charged and, through electrostatic interactions, would bind to the negatively charged SP-Sepharose or positively charged Q-Sepharose. The packed columns were equilibrated with between 2 and 4 column volumes of the appropriate binding buffer. After harvesting (prior to loading) the supernatant was diluted either with the same buffer or distilled water and was then applied to the column (this reduced the salt concentration so as not to weaken the interactions between sample and resin). Elution of the protein was then achieved by applying the same buffer but containing 1 M NaCl. This was done either on the bench or using the ÄKTA-design FPLC system (pump P-920, UV detector unit UPC-900). In the latter case a gradient of the buffer was applied, from 100 % binding buffer to 100 % buffer with 1 M salt (linear gradient). The flow-rate used was 3.5 ml/min and the volume of the fractions collected was 2.5 ml.

Immobilized metal-affinity chromatography was employed as the first purification step in the case of the PspC constructs, which were expressed with a His-tag. A 5-ml HiTrap IMAC HP column (GE Healthcare) pre-packed with IMAC Sepharose High-Performance resin charged with Co^{2+} was used. The binding buffer consisted of 20 mM potassium phosphate buffer, pH 7.0, supplemented with 0.5 M NaCl to inhibit non-specific binding and 20 mM imidazole to minimize the binding of unwanted host cell proteins. This buffer was used to equilibrate the column and to dilute the lysate obtained after harvesting. The elution buffer was made by adding imidazole to the binding buffer to a final concentration of 0.5 M. The lysate was applied on the pre-equilibrated column manually using a syringe and the protein was then eluted using the ÄKTA-design FPLC system (pump P-920, UV detector unit UPC-900), and a linear gradient from 20 mM to 500 mM imidazole. The flow-rate used was 5 ml/min and the volume of the fractions collected was 2.5 ml.

Size-exclusion chromatography was employed as a subsequent purification step using the gel-filtration chromatography column, HiLoad 16/60
(16 mm x 600 mm) Superdex 75, 120 ml, prep grade. Prior to loading the column was equilibrated with 1.1-1.6 column volumes of the appropriate buffer. Based on the SDS of the fractions from the SP/Q-Sepharose elutions as well as the Co^{2+}-affinity chromatography, selected fractions were chosen and concentrated in membrane concentrators. They were then buffer-exchanged into the appropriate buffer and 0.5-1 ml samples were applied to the column using the ÄKTA-design™ FPLC system (pump P-920, UV detector unit UPC-900). The flow-rate used was
1 ml/min and the volume of the fractions collected was 1.5 ml.

It is noted that due to a problem that occurred regarding the autozero function of the FPLC system that was used for the purification of the proteins, some of the chromatograms presented in Chapter 3 display measurements on the *y-axis* that have negative values instead of zero.

2.2.7 Mass spectrometry

The Scottish Instrumentation and Resource Center for advanced Mass Spectrometry (SIRCAMS) facility of the University of Edinburgh provided a mass analysis service. Prior to submission, the sample was prepared as follows: A minimum of 10 μM concentration of the protein sample in ~20 μl of 20 mM potassium phosphate buffer or phosphate-buffered saline (PBS) was transferred to a microcentrifuge tube and was briefly centrifuged (13000 rpm). Mass analysis was achieved using LC-MS, LC-MS Dionex U300 HPLC 500 μM PSDVB monolyth column-Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS). The results served as a verification of the mass of the proteins. In addition, in the case of proteins produced in *P. pastoris*, it provided information regarding the formation or not of disulfide bonds and whether none, one or two expression artifacts (*i.e.* Glu-Ala from the *P. pastoris* expression system) were present. Finally, apart form LC-MS accurate mass analysis, trypsin digest was performed where considered necessary (again by SIRCAMS) in order to identify matching peptides within the

protein samples (trypsin potentially cuts to the C-terminal side of Lys and Arg unless the next residue is Pro).

2.3 NMR studies

All of the spectra reported in this thesis were collected on a Bruker AVANCE 18.8-Tesla (800 MHz) spectrometer with the exception of the ^{1}H, ^{15}N-HSQC spectrum for ^{13}C, ^{15}N-FH-11-12 which was acquired on a Bruker AVANCE 14.1-Tesla (600 MHz) spectrometer. Both instruments were fitted with 5-mm triple-resonance probes/cryoprobes. All 1D, 2-D and 3-D spectra were recorded at 25 ^{o}C (298 K). All the experiments were setup with the guidance of Mr. Juraj Bella with the exception of several 1-D and 2-D ^{15}N-HSQC experiments that were setup by the author unassisted. The pulse sequences employed in this study were written by Dr Dušan Uhrin. Experimental parameters and procedures for setting up experiments were taken from http://nmrlinux.chem.ed.ac.uk/highfield/highfield.html.

2.3.1 Data processing and analysis

Topspin (Bruker SpectroSpin) was used for experimental set-up, data acquisitions and initial data processing on the instrument. The data were subsequently transferred onto the Linux-based computer cluster and processed offline using software provided by the CCPNMR (Collaborative Computational Project for the NMR community) (112). Thus data were processed using AZARA (W. Boucher, Department of Biochemistry, University of Cambridge, UK), after which spectra were visualized and assignments achieved using ANALYSIS (version 2.0.7 initially and version 2.1.1 subsequently) (112).

2.3.2 Processing of the data

Within AZARA, a parameter file called *ser.ref* and a script input file called *scr* were produced. The *scr* file contains various commands such as Fourier transformation, phasing and baseline correction. Depending on the specific

requirements of each experiment, the appropriate alterations were made in the *scr* file. This was subsequently applied to process experiments and the resultant spectrum visualized using the commands "*plot2"* in the case of the 2D spectra and the command "*planes"* in the case of the 3-D spectra. When a satisfactory result was achieved the output spectrum resulting from the processing, called *spc*, was loaded into ANALYSIS for assignment.

2.3.3 NMR experimental parameters for the FH and PspCN constructs

Table 2.9 summarizes the NMR experimental parameters for the spectra presented in Chapters 3 and 4 for constructs FH-11-12, FH-13-14, FH-13, FH-10-15, PspCN and PspCNR1.

Construct	Experiment	Dim 1	Dim 2	Number of scans	Size of FID[&] (Dim 1)	Size of FID (Dim 2)	Spectral width (ppm)	Spectral width (Hz)	Spectral resolution (Hz)	Water suppression
FH-10-11	^{1}H	^{1}H	–	32	32768	–	Dim 1: 16.94	Dim 1: 13550.14	Dim 1: 0.41	1
^{13}C, ^{15}N-FH-11-12[$]	^{1}H, ^{15}N-HSQC (113)	^{1}H	^{15}N	4	–	512	Dim 1: 13.98 Dim 2: 24.80	Dim 1: 8389.26 Dim 2: 1507.84	Dim 1: 4.096 Dim 2: 2.945	2
FH-13-14	^{1}H	^{1}H	–	32	32768	–	Dim 1: 15.02	Dim 1: 12019.23	Dim 1: 0.37	1
^{15}N-FH-13-14	^{1}H, ^{15}N-HSQC	^{1}H	^{15}N	8	–	512	Dim 1: 11.00 Dim 2: 28.00	Dim 1: 8802.82 Dim 2: 2270.15	Dim 1: 4.30 Dim 2: 4.43	2
^{15}N-FH-13[#]	^{1}H, ^{15}N-HSQC	^{1}H	^{15}N	48	–	256	Dim 1: 16.93 Dim 2: 35.99	Dim 1: 13550.14 Dim 2: 2918.86	Dim 1: 6.62 Dim 2: 11.40	2
^{15}N-FH-10-15	^{1}H	^{1}H	–	32	32768	–	Dim 1: 15.03	Dim 1: 12019.23	Dim 1: 0.37	1
^{15}N-FH-10-15	^{1}H, ^{15}N-HSQC	^{1}H	^{15}N	48	–	256	Dim 1: 13.95 Dim 2: 60.00	Dim 1: 11160.71 Dim 2: 4863.81	Dim 1: 5.45 Dim 2: 18.99	2
^{15}N-FH-10-15	^{1}H, ^{15}N-TROSY	^{1}H	^{15}N	192	–	512	Dim 1: 13.95 Dim 2: 60.00	Dim 1: 11160.71 Dim 2: 4863.81	Dim 1: 5.45 Dim 2: 9.45	3
PspCN	^{1}H	^{1}H	–	32	32768	–	Dim 1: 15.03	Dim 1: 12019.23	Dim 1: 0.37	1
PspCN	^{1}H, ^{15}N-HSQC	^{1}H	^{15}N	256	–	512	Dim 1: 11.00 Dim 2: 60.00	Dim 1: 8802.82 Dim 2: 4863.81	Dim 1: 4.30 Dim 2: 9.50	1
PspCNR1	^{1}H	^{1}H	–	32	32768	–	Dim 1: 15.03	Dim 1: 12019.23	Dim 1: 0.37	1

Table 2.9 NMR experimental parameters for the FH and PspC constructs

[$] From Ilias Matis
[#] From Dr C. Q. Schmidt
[&] free induction decay (FID)
1 Double-pulsed field gradient spin-echo
2 Watergate
3 Dephasing of water magnetization by coherence selection pulsed field gradients

2.3.4 NMR structural studies

The construct FH-10-11 was the most experimentally tractable and was used for structural determination by NMR. 2-D and 3-D double and triple-resonance NMR experiments were recorded on the double-labeled sample as a basis for structure determination. As this sample degraded slowly over time, a total of three ^{15}N, ^{13}C-labeled samples were prepared (using the same methods). The NMR experiments required for the structural studies (Tables 2.9 and 2.10) were therefore recorded in three sets, each on a recently expressed and purified sample.

In addition to FH-10-11, NMR was also used to solve the solution structure of FH-11-12. This construct was produced by Ilias Matis (in this group) who also collected the NMR data and carried out the majority of the assignments as part of his MSc project. However due to lack of time on that project, the remaining steps needed to produce the 3-D structure were undertaken by the author. Specifically, corrections were made where necessary in the assignments and cross peaks in the NOESY spectra were re-picked and re-assigned.

The information provided hereafter regarding the assignment strategy applies to FH-10-11 while the same approaches were used - as required - for the corrections of the assignments and the picking of the NOESY spectra in the case of FH-11-12. For FH-10-11 and FH-11-12, CYANA and CNS-based structure calculations were performed.

2.3.4.1 Sample preparation for FH-10-11

The construct FH-10-11 was produced from *P. pastoris* and purified as described above (and reported in Chapter 3). The samples were concentrated in

centrifugal concentrators and were buffer-exchanged into 20 mM potassium phosphate buffer, pH 6.6. Aliquots of D_2O and NaN_3 solutions were added to final concentrations of 5-10 % (v/v) and 0.01 % (w/v) respectively. For the ^{13}C D_2O NOESY and the slowly exchanging amide experiments, the samples were freeze-dried overnight and subsequently dissolved in 99.6 % D_2O.

2.3.4.2 Collection of the NMR data

Tables 2.10 (for 2-D spectra) and 2.11 (for 3-D spectra), list the experiments that were used for the structural studies of FH-10-11.

Experiment	Dim 1	Dim 2	Number of scans	Size of FID (Dim 2)	Spectral width (ppm)	Spectral width (Hz)	Spectral resolution (Hz)	Water suppression
^{1}H, ^{15}N-HSQC (113) (Sample 1)	^{1}H	^{15}N	4	512	Dim 1: 11.00 Dim 2: 28.00	Dim 1: 8802.82 Dim 2: 2270.15	Dim 1: 4.43 Dim 2: 4.30	2
^{1}H, ^{13}C-HSQC (114) (Sample 2)	^{1}H	^{13}C	12	1024	Dim 1: 11.00 Dim 2: 34.21	Dim 1: 8802.82 Dim 2: 6882.31	Dim 1: 4.30 Dim 2: 6.72	3
(HB)CB(CGCD)HD (115) (Sample 3)	^{1}H	$^{13}C_{arom}$	256	256	Dim 1: 11.00 Dim 2: 40.02	Dim 1: 8802.82 Dim 2: 8051.53	Dim 1: 4.30 Dim 2: 31.45	4
(HB)CB(CGCDCE)HE (115) (Sample 3)	^{1}H	$^{13}C_{arom}$	256	256	Dim 1: 11.00 Dim 2: 40.02	Dim 1: 8802.82 Dim 2: 8051.53	Dim 1: 4.30 Dim 2: 31.45	4
Aromatic ^{1}H, ^{13}C-HSQC (114) (Sample 3)	^{1}H	$^{13}C_{arom}$	16	1024	Dim 1: 11.00 Dim 2: 40.02	Dim 1: 8802.82 Dim 2: 8051.30	Dim 1: 8.60 Dim 2: 7.86	3

Table 2.10 Summary of the 2D acquired spectra utilized for the structural studies.

2 Watergate
3 Dephasing of water magnetization by coherence selection pulsed field gradients
4 Watergate using a composite off-resonance pulse

Experiment	Dim 1	Dim 2	Dim 3	Number of scans	Size of FID (Dim 2)	Size of FID (Dim 3)	Spectral width (ppm)	Spectral width (Hz)	Spectral resolution (Hz)	Water suppression
HBHA(CO)NH (116) (Sample 1)	^{1}H	$^{1}H\alpha/\beta$	^{15}N	4	256	128	Dim 1: 11.00 Dim 2: 6.67 Dim 3: 27.99	Dim 1: 8802.82 Dim 2: 5336.18 Dim 3: 2269.63	Dim 1: 4.30 Dim 2: 20.84 Dim 3: 17.73	3
CBCANH (117) (Sample 1)	^{1}H	$^{13}C\alpha/\beta$	^{15}N	16	512	1024	Dim 1: 11.00 Dim 2: 63.00 Dim 3: 27.99	Dim 1: 8802.82 Dim 2: 12674.27 Dim 3: 2269.63	Dim 1: 5.26 Dim 2: 8.42 Dim 3: 2.44	3
CBCA(CO)NH (118) (Sample 1)	^{1}H	$^{13}C\alpha/\beta$	^{15}N	4	512	256	Dim 1: 11.00 Dim 2: 63.00 Dim 3: 27.99	Dim 1: 8802.82 Dim 2: 12674.27 Dim 3: 2269.63	Dim 1: 0.33 Dim 2: 8.42 Dim 3: 9.76	2
HNCO (119) (Sample 1)	^{1}H	^{13}CO	^{15}N	4	512	256	Dim 1: 11.00 Dim 2: 10.50 Dim 3: 27.99	Dim 1: 8802.82 Dim 2: 2112.38 Dim 3: 2269.63	Dim 1: 5.26 Dim 2: 8.42 Dim 3: 9.76	3
HN(CA)CO (120) (Sample 1)	^{1}H	^{13}CO	^{15}N	16	512	256	Dim 1: 11.00 Dim 2: 10.50 Dim 3: 27.99	Dim 1: 8802.82 Dim 2: 2112.38 Dim 3: 2269.63	Dim 1: 5.26 Dim 2: 8.42 Dim 3: 9.76	3
HCCH-TOCSY (121) (Sample 2)	^{1}H	^{1}H	^{13}C	16	128	256	Dim 1: 11.02 Dim 2: 6.67 Dim 3: 34.20	Dim 1: 6613.76 Dim 2: 4000 Dim 3: 5159.96	Dim 1: 5.26 Dim 2: 33.67 Dim 3: 9.76	5
H(C)(CO)NH-	^{1}H	$^{1}H_{sidech}$	^{15}N	16	128	1024	Dim 1: 11.02	Dim 1: 6613.76	Dim 1:7.60	2

TOCSY (122) (Sample 2)		ain					Dim 2: 6.67 Dim 3: 28.00	Dim 2: 4001.60 Dim 3: 1702.42	Dim 2: 33.67 Dim 3: 1.42	
(H)C(CO)NH-TOCSY (122) (Sample 2)	^{1}H	$^{13}C_{sidec}$ hain	^{15}N	8	128	256	Dim 1: 11.02 Dim 2: 63.01 Dim 3: 28.00	Dim 1: 6613.76 Dim 2: 9505.70 Dim 3: 1702.42	Dim 1: 5.26 Dim 2: 33.67 Dim 3: 9.76	2
^{13}C- HSQC NOESY (123) (Sample 3)	^{1}H	^{1}H	^{13}C	16	512	512	Dim 1: 22.01 Dim 2: 10.50 Dim 3: 34.21	Dim 1: 17605.63 Dim 2: 8403.36 Dim 3: 6882.31	Dim 1: 4.10 Dim 2: 8.42 Dim 3: 11.78	5
^{15}N-HSQC NOESY (124) (Sample 3)	^{1}H	^{1}H	^{15}N	16	512	256	Dim 1: 11.00 Dim 2: 10.54 Dim 3: 28.00	Dim 1: 8802.82 Dim 2: 8403.36 Dim 3: 2270.15	Dim 1: 4.10 Dim 2: 8.42 Dim 3: 8.55	2

Table 2.11 Summary of the 3D acquired spectra utilized for the structural studies.

2 Watergate
3 Dephasing of water magnetization by coherence selection pulsed field gradients
5 Purging gradients plus storing the water magnetization along the z axis just before the acquisition

2.3.4.3 Overview of resonance assignment

The assignment of ^{1}H, ^{13}C and ^{15}N resonances in a ^{13}C, ^{15}N-labeled protein is achieved on the basis of a suite of 3-D NMR experiments. Several 3-D experiments use the 2-D ^{15}N-HSQC as a reference or root spectrum. This experiment correlates the amide nitrogen of each amino acid residue with its attached proton, hence yielding a two-dimensional spectrum with each cross-peak occurring at the intersection of an amide nitrogen chemical shift and the chemical shift of its attached proton (Fig. 2.1). The 3-D experiments exploit their third dimensions to allow these correlations to be extended, through-bond, to other atoms or indeed to a group of atoms that corresponds to a spin system. Hence by combining sets of three-dimensional experiments, overlapping strips (each corresponding to correlations with a particular amide group) can allow for a string of spin systems to be arranged in order, which will then be matched to a sequence of residues in the polypeptide chain, as described in more detail below. Following this, TOCSY experiments are useful for extending the assignment to the sidechain of every amino acid residue. For the HCCH-TOCSY, which was the experiment solely used for this purpose, the reference experiment is the ^{13}C-HSQC, which correlates each carbon with its attached aliphatic proton/protons both in the backbone and the sidechain. A special set of NMR experiments is used to assign aromatic atoms since these lie within their own spin systems. Finally, the assignments from the ^{15}N-HSQC and ^{13}C-HSQC can be transferred to the ^{15}N-NOESY and ^{13}C-NOESY respectively. Peaks in these experiments arise based on the nuclear Overhauser effect (NOE), and are used to identify pairs of protons that are close in space.

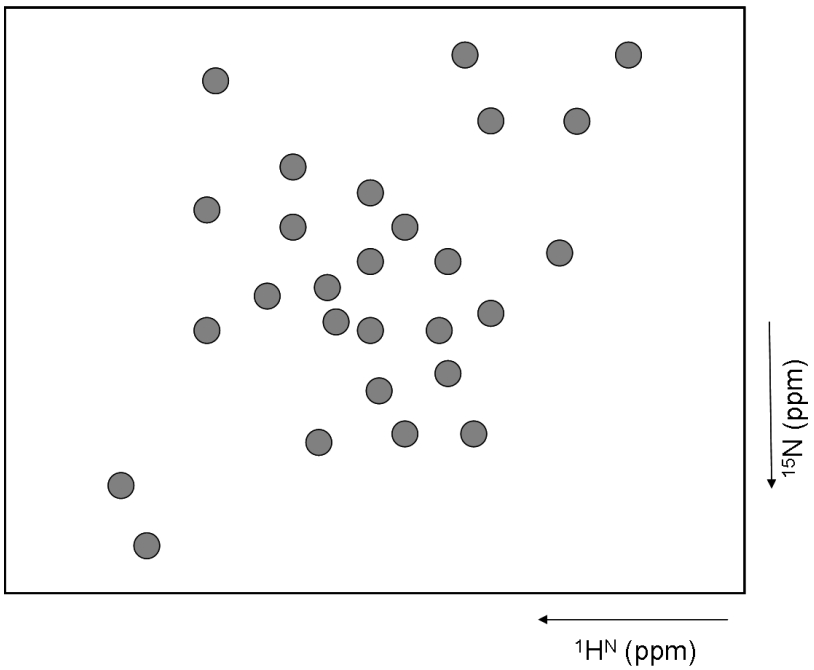


Figure 2.1 Depiction of a ^{1}H, ^{15}N-HSQC spectrum. Each cross-peak (grey dot) represents an amide proton.

2.3.4.4 Assignment of backbone resonances for ^{13}C, ^{15}N-labeled FH-10-11

The backbone assignment was carried out by using two complementary pairs of experiments: CBCANH and CBCA(CO)NH, and HNCO and HN(CA)CO. Both pairs correlate the i-1 residue with the i residue (Fig. 2.2). As a result, the atomic assignment was made for the Cα, Cβ and the CO resonances as well as the amide nitrogen and proton. The method by which the i-1 residue is correlated with the i residue for the two pairs of experiments is explained below.

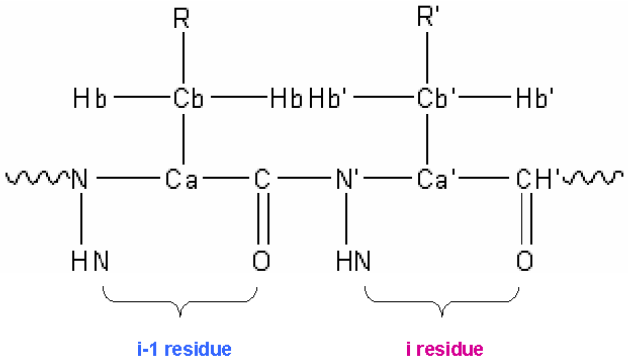


Figure 2.2 Nomenclature referring to sequential amino acid residues.

In the case of the CBCANH-CBCA(CO)NH pair, in the first experiment, magnetization is transferred to $^{15}N_i$ from both $^{13}C\alpha/\beta_i$ and $^{13}C\alpha/\beta_{i-1}$, hence for each NH_i group there are Cα and Cβ peaks from both residue i and i-1. The stronger ones usually correspond to the Cα and Cβ of the i residue, whereas the weaker generally come from the i-1 residue. In the second experiment, the magnetization is transferred to $^{15}N_i$ from $^{13}C\alpha/_{\beta i-1}$ only, resulting in a strip where only the Cα and Cβ of the preceding residue are visible. Comparisons of the positions of peaks in overlaid strips with those in other overlaid strips thus allowed sequential assignments to be made quite rapidly (Fig. 2.4).

With respect to the HNCO and HN(CA)CO pair, in an analogous manner as described above, sequential links can be set for the CO of the i-1 and the i residues because in the first case only the CO of the preceding residue is visible in a strip and in the second case two CO peaks are visible, of the i and the i-1 residue. This pair of experiments exhibits higher resolution and less overlapped spectra thus allowing potential ambiguities associated with the CBCANH and CBCA(CO)NH experiments to be resolved.

Figure 2.3 shows a sequential "walk" between Pro621 and Glu625 of FH-10-11. Each strip contains information about which carbons are in $residue_i$ and which are in $residue_{i-1}$ as found by overlapping peaks arising in the two spectra for the pair of experiments CBCANH and CBCA(CO)NH.

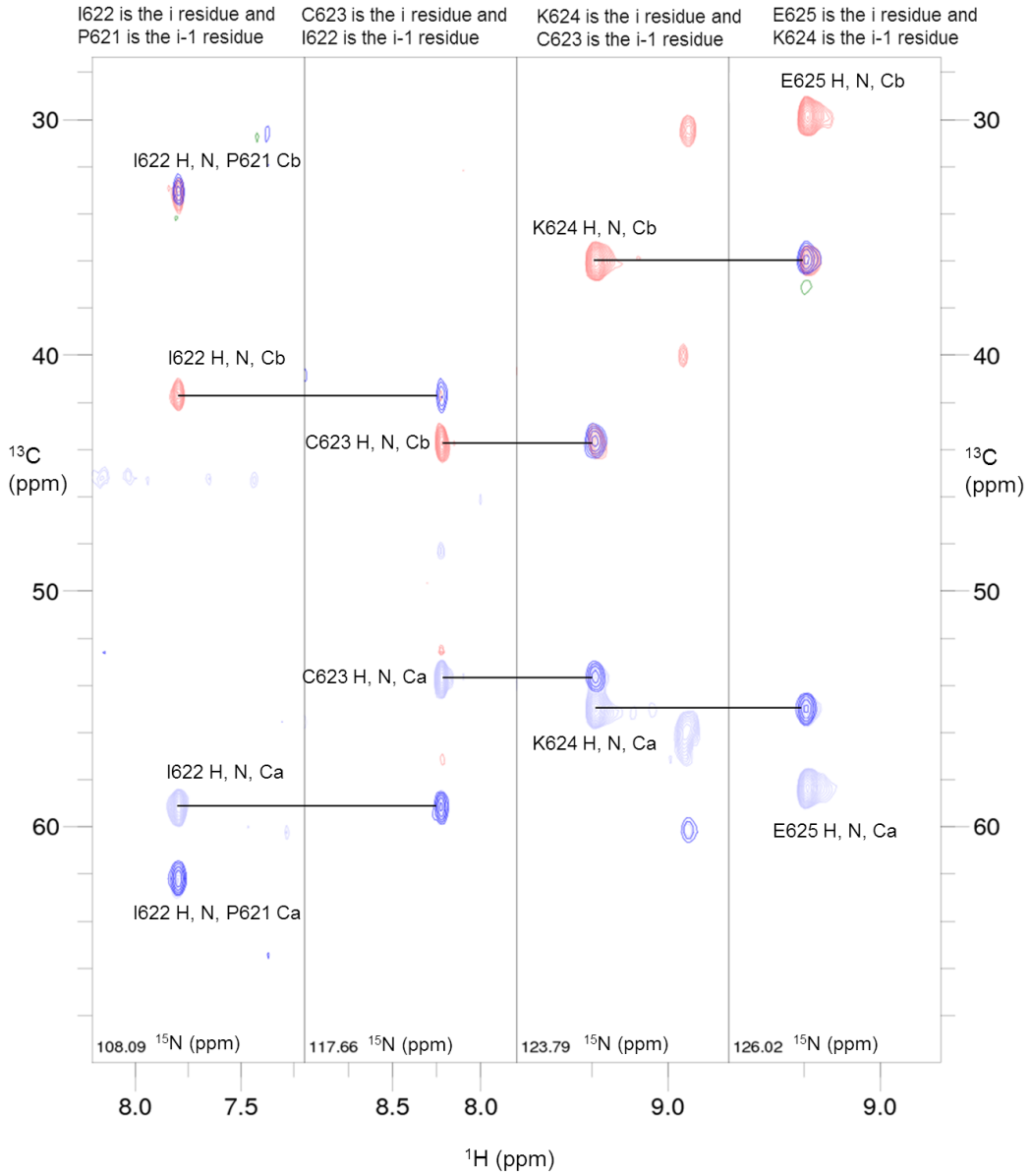


Figure 2.3 Strips that are placed in sequential order and correspond to Pro621-Glu625 within the sequence of FH-10-11. Peaks that are light pink (positive phase) or light blue (negative phase) correspond to the CBCANH experiment, whereas peaks that are blue (positive phase) or green (negative phase) arise from the CBCA(CO)NH experiment. Horizontal lines show the sequential backbone-connectivities between the two complementary set of experiments. Residues and sequence numbers are shown.

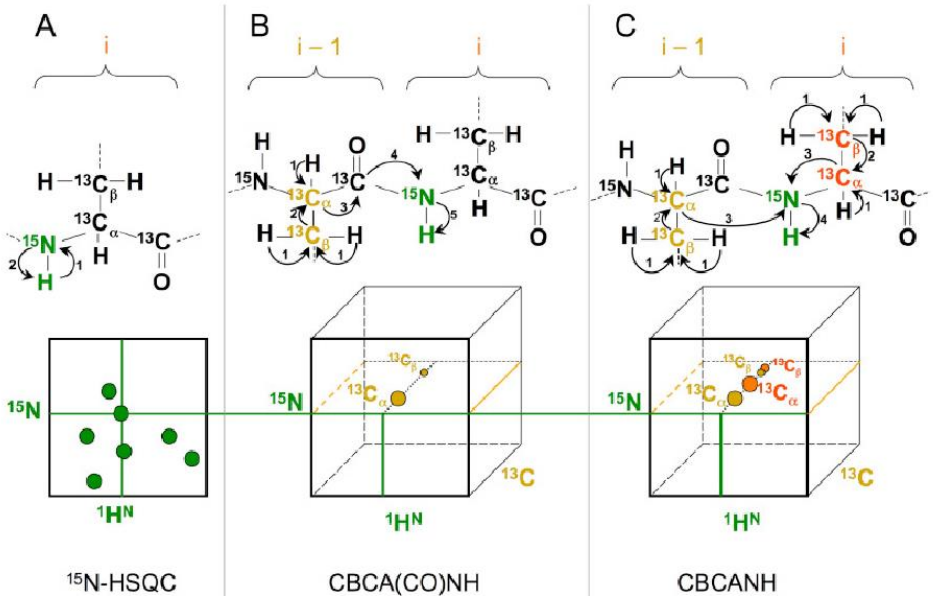


Figure 2.4 Schematic representation of magnetization transfer and the corresponding ^{15}N-HSQC (A), CBCANH (C) and CBCA(CO)NH (B) 3D experiments. The magnetization transfer throughout the steps of the experiments is indicated by arrows. One strip corresponding to a ^{15}N ^{13}C plane is shown and those coordinates correspond to specific ^{1}H ^{15}N shifts-cross peaks in the ^{15}N-HSQC.
***Figure is reproduced from Ph.D. thesis: "The structure of an active N-terminal fragment of human complement factor H" by Hocking, HG, 2008, University of Edinburgh.**

2.3.4.5 Assignment of sidechain resonances

TOCSY (Total Correlation Spectroscopy) was employed for the purpose of sidechain ^{1}H and ^{13}C assignment. The ^{15}N-TOCSY correlates aliphatic sidechain proton resonances from residue i, with root resonances HN and NH of the same residue. The H(C)(CO)NH-TOCSY and (H)C(CO)NH-TOCSY experiments correlate aliphatic proton and carbon shifts of the residue i-1 residue with the root resonances of residue i. Due to insufficient data obtained in these three experiments the majority of the sidechain assignment was carried out using the HCCH-TOCSY experiment. This experiment correlates the aliphatic proton shifts with the root resonances of the same residue. The experiments CBCANH, CBCA(CO)NH and HBHA(CO)NH were used in order to find the Cβ/Hα and Cβ/Hβ peaks in the ^{13}C-HSQC as described above. These peaks were then used to navigate in the HCCH-TOCSY spectra, ending up at the root resonance of that peak. There were occasions when peaks could not be found in the ^{13}C-HSQC using the above method. A solution

to this was the following: From a candidate peak on a strip from the HCCH-TOCSY, navigation back to the ^{13}C-HSQC ensued and this was overlaid with the HCCH-TOCSY. The resultant specific location should indicate the ^{13}C and ^{1}H chemical shift of that peak in the ^{13}C-HSQC spectrum (Fig. 2.5).

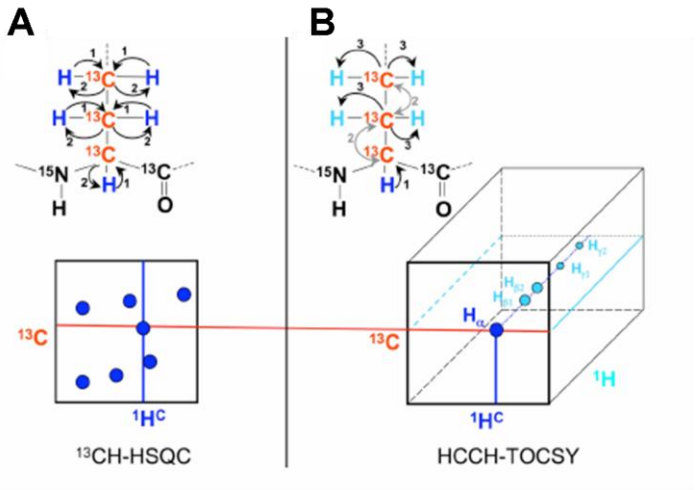


Figure 2.5 Schematic representation of the magnetization transfer (indicated by arrows) and the corresponding 3D spectra used for the assignment of aliphatic sidechain residues. Correlations are shown between the ^{13}C-HSQC (A) and the HCCH-TOCSY (B).
***Figure is reproduced from Ph.D. thesis: "The structure of an active N-terminal fragment of human complement factor H" by Hocking, HG, 2008, University of Edinburgh.**

2.3.4.6 Assignment of resonances of aromatic residues: tryptophan, tyrosine, phenylalanine and histidine

The protons of tryptophan, tyrosine, phenylalanine and histidine residues are subject to the aromatic ring-current effects and have higher chemical shifts, typically between 6-8 ppm, thus making them distinguishable from most other (non-amide) protons. In addition, protons from the aromatic rings of the corresponding residues occupy separate spin systems (from the remainder of the residue). Three specialized 2-D experiments were used for the aromatic chemical shift assignment. First, an aromatic ^{13}C-HSQC spectrum was collected which has Cδ and Cε chemical shifts in one dimension and the corresponding Hδ and Hε in the second dimension. The other two experiments, (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE correlate the Cβ chemical shifts with shifts from the protons of the rings: Hδ and Hε. In cases where insufficient data were detected in those experiments, the ^{13}C-NOESY-HSQC was

used to identify missing shifts. In most cases the aromatic sidechains are buried within the interior of the protein thus giving clear NOESY strips that can provide useful information about the chemical shifts of the corresponding aromatic protons.

2.3.4.7 Prediction of proline *cis-trans* peptide bond conformation

The construct FH-10-11 contains twelve proline residues while FH-11-12 contains ten. In general, in proteins the *trans*- conformation of the peptide bond is energetically more favorable than the the *cis*- conformation, due to steric repulsion that arises between the peptide bond and the sidechain. However, due to the cyclic nature of proline sidechains there is an appreciable incidence of *cis*-conformations of X-Pro peptide bonds. Based on statistical analysis, the chemical shift difference $\Delta_{\beta\gamma}=\delta\ [^{13}C\beta]-\delta\ [^{13}C\gamma]$ has been proposed as an indicator of *cis*- and *trans*-conformation (124). The statistical values of this chemical shift difference have been shown to be 4.51 ppm ± 1.17 ppm and 9.64 ppm ± 1.27 ppm, for *trans*- and *cis*-conformations respectively (125). The tables below show the correspondent results for FH-10-11 and FH-11-12 (Tables 2.11 and 2.12). The resultant conformations were cross-verified by using the ^{13}C-NOESY spectra. In *cis*- isomers the Hα of a Pro residue will have a strong cross-peak with the Hα of the preceding residue while in a *trans*- isomer the Hδ of a Pro residue will have a strong cross-peak with the Hα of the preceding residue.

Sequence number of Proline residue	^{13}Cβ (ppm)	^{13}Cγ (ppm)	δ[13Cβ]-δ[13Cγ] (ppm)	Conformation
572	32.190	27.177	5.013	*trans*
580	31.672	27.202	4.470	*trans*
599	31.710	27.898	3.812	*trans*
606	32.297	27.732	4.565	*trans*
618	33.532	25.322	8.210	*cis*
621	33.030	27.016	6.014	*trans*
632	30.838	27.985	2.853	*trans*
633	30.060	26.954	3.106	*trans*
634	32.390	27.219	5.171	*trans*
661	31.717	27.879	3.838	*trans*
668	32.326	27.604	4.722	*trans*
682	33.033	26.188	6.845	*trans*

Table 2.12 Summary of chemical shifts and prediction of proline peptide bond conformation for FH-10-11.

Sequence number of Proline residue	^{13}Cβ (ppm)	^{13}Cγ (ppm)	δ[13Cβ]-δ[13Cγ] (ppm)	Conformation
632	30.755	28.179	2.576	*trans*
633	29.960	27.024	2.936	*trans*
634	32.586	27.127	5.469	*trans*
661	31.956	28.350	3.606	*trans*
668	32.596	27.295	5.301	*trans*
682	33.235	26.013	7.222	*trans*
695	32.419	27.325	5.094	*trans*
707	30.621	26.288	4.333	*trans*
708	34.857	24.772	10.085	*cis*
742	32.631	26.150	6.481	*trans*

Table 2.13 Summary of chemical shifts and prediction of proline peptide bond conformation for FH-11-12.

2.3.4.8 Picking and assignment of NOE cross-peaks

It should be noted here that in the case of FH-10-11 there was insufficient signal-to-noise in the ^{13}C-NOESY spectrum recorded in the standard (10% v/v D_2O) NMR buffer. Hence, the ^{13}C-NOESY resulting from a more concentrated protein sample in 99.6% v/v D_2O was exclusively used for picking and assignment.

The reference spectrum for the ^{15}N-NOESY was, again, the ^{15}N-HSQC. From each peak in the latter, navigation in the ^{15}N-NOESY yielded a strip for each root resonance (amide). All the non-noise peaks in that strip were picked (except those corresponding to water) and assigned to that root resonance, by overlaying planes of the ^{15}N-NOESY with its reference ^{15}N-HSQC. This results in assignment in

two dimensions. In a similar manner the picking and assignment for the ^{13}C-NOESY was done, only in this case the reference spectrum was the ^{13}C-HSQC and the identification of intra-residue cross peaks was done through the overlay of the ^{1}H-^{1}H planes with the HCCH-TOCSY. Figure 2.6 shows an example of how a NOE cross-peak between close in space amide protons is confirmed.

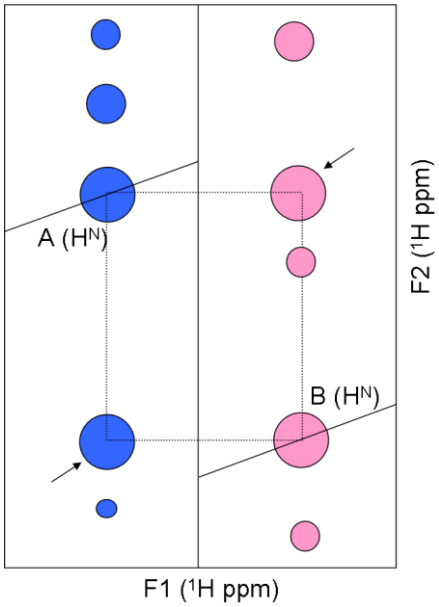


Figure 2.6 Representation of two symmetry-related amide proton cross-peaks in two hypothetical strips of the NOESY spectrum. Proton A experiences "through space" dipolar coupling with proton B resulting in a NOE cross-peak at the proton frequency of B. Proton B also experiences a corresponding effect resulting in a NOE cross-peak at the proton frequency of A. The two symmetry related cross-peaks are indicated by an arrow.

For the assignment of the second dimension proton shifts, when two symmetry-related NOE cross peaks could be identified unambiguously, the assignment was regarded as firm. In case of lack of clarity regarding assignment, the peaks were left unassigned. Peak lists for the two NOESY experiments and chemical shift tables for all the atoms were exported, alongside the amino acid sequence, from the appropriate CCPNMR ANALYSIS project and used for the structure calculations in CYANA (126). This required a set of pre-written scripts to be correctly amended which would then be implemented in the structure calculations.

2.3.4.9 Structure calculations

2.3.4.9.1 Overview

The software CYANA (126) was employed for the initial structure calculations for FH-10-11 and FH-11-12. CYANA combines the automated assignment of NOESY spectra with subsequent structure calculations performed in torsion-angle space. It has been shown however that refinement of protein structures in a thin layer of water solvent significantly improves quality indices, important for the overall assessment of the final structures. Due to the absence of that feature in CYANA, the software CNS was employed for the intermediate and final structure calculations. Structure calculations in CNS are performed in cartesian space. The water-refined structures from CNS yielded the final NMR ensembles for FH-10-11 and FH-11-12. Figure 2.7 schematically illustrates the basic steps followed for the aforementioned structure calculations.

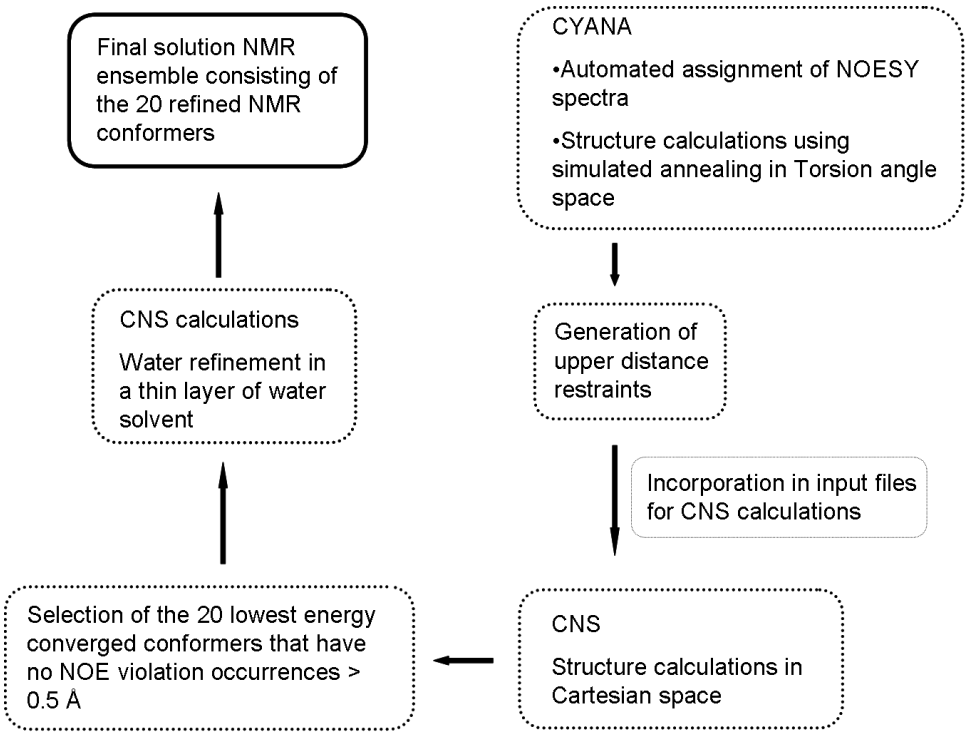


Figure 2.7 Flow-chart of the most important steps of structure calculations in CYANA and subsequently in CNS in order to obtain the NMR ensembles for FH-10-11 and FH-11-12.

2.3.4.9.2 Structure calculations in CYANA

For the initial structure calculations and automated NOE assignment of the NOESY spectra the structure-calculation software CYANA version 2.1 was used for FH-10-11 and FH-11-12 (126). The CYANA 2.1 software, is based on DYANA (127) and embeds the CANDID algorithm (for Combined automated NOE assignment and structure determination) for automated NOESY assignment (128). This ambiguous assignment is then followed by structure calculations using torsion-angle dynamics.

2.3.4.9.2.1 Automated NOESY assignment

Within CYANA (126) seven cycles of NOE cross peak assignment that proceeded in an iterative approach resulted in the assignment of an initially partially assigned NOESY spectrum. The initial NOE assignments created were weighed with respect to several filtering criteria, such as self-consistency with the entire NOE network. Spurious cross-peaks were eliminated and a 3-D structure was calculated. The impact of incomplete and/or erroneous input data was effectively minimized by the presence of two "safeguard" features incorporated in the program: constraint combination and network anchoring. In constraint combination, distance constraints are generated with combined assignments from different, in general unrelated, cross peaks thus minimizing the potential occurrence of erroneous restraints originating from NOE artifacts; this is applied only in the two first cycles. Furthermore, the concept of ambiguous distance constraints that exist in CYANA is important for dealing with the preliminary chemical shift-based NOESY cross-peaks assignments. Briefly, every NOESY peak is handled as a superposition of the signals from each of its multiple assignments. The resultant summed distance between a pair of atoms is always shorter than any of the individual distances used in the sum. Therefore when the correct assignment exists an ambiguous constraint is not misinterpreted when incorrect assignment possibilities may also be present. This property of ambiguous distance constraints further implies that (with constraint combination) the combined constraints will have lower probabilities of being incorrect. In network anchoring,

each preliminary chemical shift-based assignment is rated on the basis of how well the corresponding NOE-connectivity fits in a consistent set of neighboring NOE-based networks. This assessment of the self-consistency of the NOE assignments is independent of the 3-D structure that is generated after every cycle. The network anchoring filter can detect false restraints that could potentially result in artificial protein constraints.

2.3.4.9.2.2 Structure calculations

The structure calculations are performed in such a way that the potential energy target function is minimized whilst satisfying the experimental distance restraints. A conformation that satisfies the restraints more closely than another one will lead to a lower target function. The energy minimization function uses simulated annealing in torsion-angle space, ultimately resulting in a global energy minimum for the polypeptide. In torsion angle dynamics empirical restraints such as bond lengths, bond angles and chiralities, are maintained at their optimal value. In all cycles except for cycle seven, swapping of pairs of diastereoscopic atoms was performed during the simulated annealing for the unassigned stereospecific atom pairs (e.g. methylene groups).

After the completion of the seven-cycle NOE assignment and structure calculations, several quality-control criteria are provided that allow assessment of the resulting structures and will be presented in the NMR chapter. Figure 2.8 illustrates the seven-cycle automated NOE cross-peak assignment and 3D structure calculation performed in CYANA.

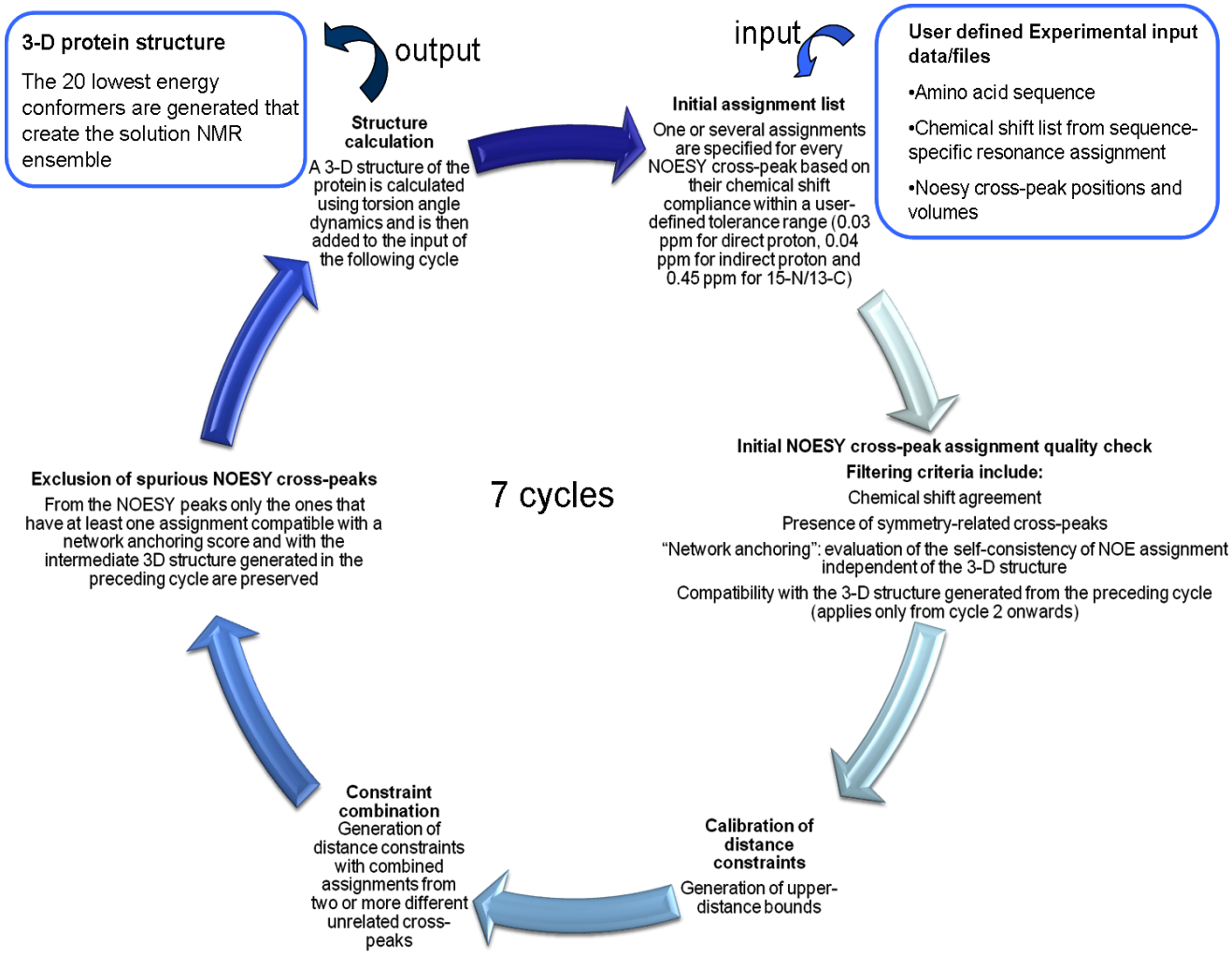


Figure 2.8 Automated NOE cross-peak assignment and 3D structure calculation performed by CYANA (adapted from (126)).

2.3.4.9.3 Structure calculations in CNS

The calculations performed using CYANA resulted in the assignment of the NOESY spectra, the generation of distance restraints and the calculation of 3-D NMR-derived structures. Several validation programs were routinely employed to assess quality features of the resultant NMR ensembles. Those features include Ramachandran plot statistics and packing quality. To avoid unsatisfactory quality results, arising when electrostatic and Van der Waals interactions are treated unrealistically, refinement of 3-D protein structures in explicit solvent is recommended as this has been shown to significantly improve structural quality criteria (129). This refinement approach was not an option in CYANA; however it has been implemented in the CNS structure calculation software. Therefore CNS version 1.2 (130) was used in this study for the final structure calculations. Due to its lack of automated NOE assignment capability (as implemented in the Edinburgh Biomolecular NMR Unit) however, it was not used for the initial calculations. The restraint output file generated from the CYANA calculations was therefore imported (through FormatConverter, a facility of ANALYSIS) into ANALYSIS, and converted into CNS format and then exported for implementation in the CNS calculations.

The structure calculations in CNS are performed in cartesian space during which a global minimum region is searched for the total energy function F_{tot}. F_{tot} consists of the following energy terms: $F_{tot}=F_{covalent} + F_{repel} + F_{NOE}$. In this equation F_{tot} represents the effective potential energy, $F_{covalent}$ maintains correct bond lengths, angles, planarities, chiralities, F_{repel} describes the non-bonded interactions and is a repulsive function that prevents overly close contacts between atoms (131), (132) and F_{NOE} refer to the experimental NOE restraints. The basic features of the protocols used in the CNS calculations comprised a generation of a molecular topology file which essentially included information on the primary structure of the proteins, a template/random starting structure, prochiral swapping (swapping of methylene, methyl and amide protons belonging to sidechains so as to prevent energy minimization from getting trapped in local energy minima) and finally refinement.

In every round of CNS calculations, 100 structures were calculated in total for FH-10-11 and FH-11-12. These structures were subsequently ranked according to overall and NOE energies. The converged structures were checked for violated NOE restraints (NOE violations > 0.5 Å). When a satisfactory output was accomplished in which no violations occurred > 0.5 Å - in this case for the 20 lowest energy structures (by using a script that as an output indicates violations > 0.5 Å that have arisen between specific sets of restraints within the corresponding member/s of the 20 conformers), and a substantial number of structures showed convergence (around 20 structures, by plotting the overall and NOE-derived energy against all 100 structure numbers within "gnuplot"), the selected structures were subjected to water refinement within CNS. For that purpose the RECOORD protocols (http://www.ebi.ac.uk/msd/recoord) were used. During the calculations, a 7 Å thin layer water shell was created around the proteins while the refinement took place in a full molecular dynamics force field including electrostatic and Lennard-Jones non-bonded potentials (129).

2.3.4.10 Dynamics

Two relaxation pathways are involved in the return of the spins to the equilibrium state following an NMR excitation pulse: the spin-lattice (or longitudinal) and spin-spin (or transverse). Insight into the backbone dynamics of a protein can be obtained by measuring time constants: T_1 for spin-lattice and T_2 for spin-spin, since relaxation depends on the molecular motion. In this study, in addition to the T_1 and T_2 rates steady-state $\{^1H\}$, ^{15}N-NOE enhancements were measured for FH-10-11 and FH-11-12 (using a 600 MHz spectrometer for both).
For T_1, a series of interleaved ^{15}N-HSQC, one for each value of the relaxation delay time τ was measured. These τ values were used to obtain the T_1 values and for FH-10-11 were 51.2 ms, 301.2 ms, 501.2 ms, 601.2 ms, 751.2 ms, 851.2 ms and
901.2 ms while for FH-11-12 51.2 ms, 301.2 ms, 501.2 ms, 601.2 ms, 701.2 ms, 801.2 ms and 901.2 ms. All amide peaks from each ^{15}N-HSQC spectrum were picked in ANALYSIS while peaks with very weak signals or peaks located in overlapped regions were excluded. The Rates Analysis function in ANALYSIS was used to

determine the T_1 values by an exponential fit of the cross-peak height as a function of relaxation delay times (133).

As for T_1, for T_2 a series of interleaved ^{15}N-HSQC experiments were recorded for each τ value. The τ values used for FH-10-11 were 16.96 ms, 33.92 ms, 67.84 ms, 101.76 ms, 118.72 ms, 135.68 ms and 152.64 ms while for FH-11-12 16.96 ms, 33.92 ms, 67.84 ms, 84.8 ms, 101.76 ms, 118.82 ms and 135.68 ms. A similar procedure as for T_1 was subsequently followed within ANALYSIS (133).

The heteronuclear {^{1}H}, ^{15}N-NOE experiment is a more sensitive measure of the dynamics than the T_1 and T_2 relaxation rates measurements and probes mobility on the ns timescale. For the experiment a reference 2-D experiment was recorded under non-$^1H^N$ saturation conditions followed by an experiment under steady-state saturation conditions (134), (135). The {^{1}H}, ^{15}N-NOE build up occurs during the steady-state saturation. Spectra were once again processed in AZARA and assigned in ANALYSIS. The function Data Analysis/ Heteronuclear NOE was used in this case. The more mobile the amide bond (on the ns scale), the lower the cross-peak intensity, as spins that exhibit fast motion relative to that of the overall molecule will show a decreased NOE intensity compared to the reference spectrum.

2.3.4.11 Programs for visualization, acquisition of structural characteristics and assessment of the NMR structures

The program MOLMOL (136) 2k.2 was used for initial calculation of the RMSD values for the structures of FH-10-11 and FH-11-12. The molecular visualization system PyMol (137) was used to visualize coordinate files (in PDB format) and to generate molecular structure images for this thesis. Secondary structural elements were calculated using STRIDE (138). Electrostatic analyses and the corresponding molecular structure images were generated using GRASP (139). Lipophilic and hydrophilic representations (and images) were generated using MOLCAD (140) module of SYBYL v6.9. The solvent-accessible surface area and the buried surface area at the intermodular junctions were calculated using the server GETAREA (141) as well as VADAR (142). Comparisons of the closest-to-mean CCP module structures of FH-10, 11 and 12 against all CCPs with known structures

within the complement system were performed using the structural alignment program Combinatorial Extension (143). Categorization of the CCP modules 10, 11 and 12 to a specific sequence cluster was determined according to Soares *et al.* (144). Intermodular angles for the ensembles of the NMR structures were determined as previously described (44), (145). Dr D. Soares performed the following analyses (and provided the corresponding images): Inter-modular angles, VADAR, GRASP, MOLCAD, overlay with other CCPs. Assessment of the final ensembles for FH-10-11 and FH-11-12 was achieved by using PROCHECK (146).

2.4 SAXS data collection and analysis

Synchrotron radiation X-ray scattering data were collected on the X33 beamline of the EMBL (DESY, Hamburg) by Dr H. D. T. Mertens with the assistance of the author, using a 1M PILATUS pixel detector (Dectris, Switzerland) and eight frames of 15-s exposure time. Samples were analysed at 20 °C, using protein concentrations (for FH-10-11, FH-11-12 and FH-10-12) of 0.3 – 10.7 mg/ml in PBS buffer. The sample-to-detector distance was 2.7 m covering a range of momentum transfer $0.08 \leq s \geq 6.0$ nm^{-1} (where $s=4\pi\sin(\theta)/\lambda$ with 2θ being the scattering angle and $\lambda=0.15$ nm being the wavelength). Dr H. D. T. Mertens performed the processing and analysis of the data (fit of the NMR structures of FH fragments to the SAXS data, analysis of flexibility, *ab initio* shape determination and molecular modelling). A purified sample of FH-10-12 was provided by Dr C. Q. Schmidt.

2.5 Studying the interaction between PspC and FH

2.5.1 Overview

The PspC-derived fragments (PspCN and PspCNR1) were produced in *E. coli* by recombinant means as described in Section 2.1.16. The sequences for both PspCN and PspCNR1 are based on *S. pneumoniae* strain D39, serotype 2. FH and C3b were purchased, as plasma-purified preparations, from Complement Technology, Texas. A

series of purified FH fragments were produced from recombinant strains of *P. pastoris* as described in Section 2.1.15. Purified samples of FH-6-8Y and FH-19-20 were provided by Dr C. Q. Schmidt and Dr A. Herbert respectively.

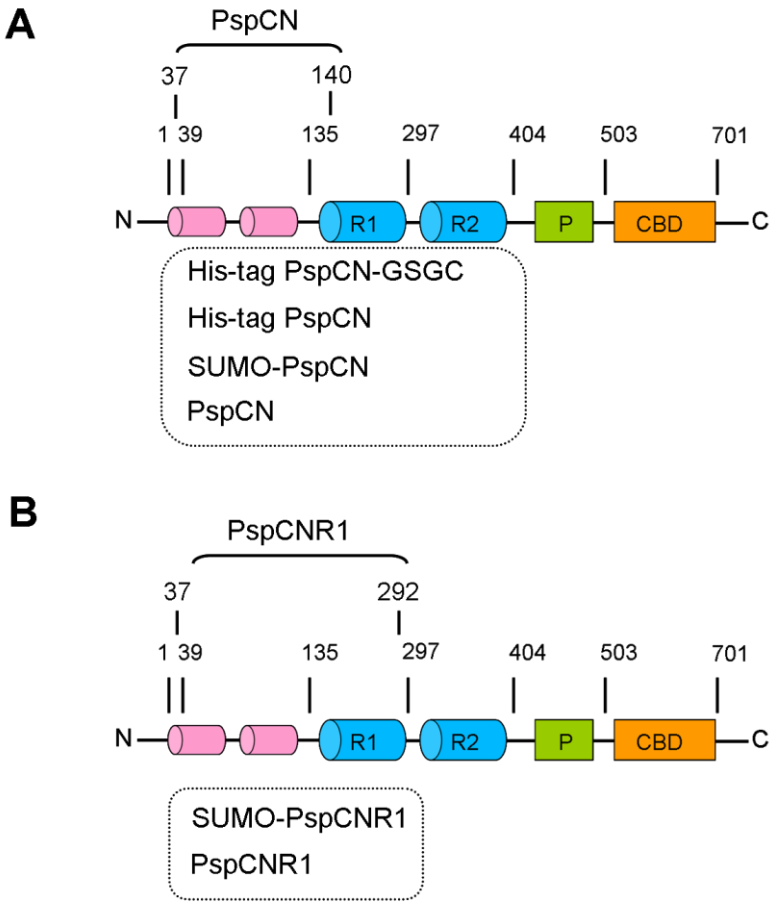


Figure 2.9 Summary of proteins employed in current study of interactions between (fragments of): PspC, FH and C3. Representations of PspC in which the N-terminal region, the R1 and R2 repeats, the Proline-rich region (P) and the Choline-binding domain (CBD) are highlighted and sequence numbers are shown. (A) The fragments of the N-terminal region, designated PspCN, (B) The fragments designated PspCNR1.

Surface plasmon resonance (SPR) was carried out using a Biacore T200 sensitivity-enhanced instrument (GE Healthcare). For the first tranche of measurements – regarded as a preliminary study - His-tagged PspCN-GSGC was used. This construct encompassed a C-terminal non-native extension, GSGC. For the second tranche a His-tagged PspCN construct was used in which the GSGC sequence is absent. For the third tranche – which was regarded as providing the most definitive data - versions of both PspCN and PspCNR1 were used. These either contained only

native sequence or were produced as fusions with SUMO. These three tranches of experiments were carried out using the buffers provided by GE Healthcare. These, which were provided as 10x concentrated stocks and subsequently diluted to 1x solutions prior to use, were either PBS (containing 0.1 M phosphate buffer with 27 mM KCl and 1.37 M NaCl which when diluted yields a pH 7.4) or HBS-P^+ (containing 0.1 M HEPES, 1.5 M NaCl and 0.5 % surfactant P20 which also yields pH 7.4 when diluted). All samples were filtered prior to use (Ultrafree - MC, centrifugal filter devices, Millipore). Data were processed using BIACORE T200 evaluation software version 1.1.

An NTA chip (*i.e.* with a matrix consisting of carboxymethylated dextran pre-immobilized with nitrilotriacetic acid (NTA) (Series S Sensor chip NTA, GE Healthcare) was used. For conditioning of the chip, prior to immobilization of the His-tagged PspC constructs and commencement of the experiments, the following procedure was performed: first 50 mM NaOH was injected repeatedly to remove any non-bonded dextran until a relatively stable baseline was observed which was then followed by 350 mM EDTA in 1 M NaCl pH ~8 to prevent any Ni^{2+}-dependent precipitation event on the chip. This was followed by between five and ten injections with 500 μM $NiCl_2$ (for a healthy NTA chip and saturated capture of Ni^{2+} typically 40-60 Response Units, (RU), are expected), followed by the His-tagged protein (PspCN-GSGC, PspCN, SUMO-PspCN, SUMO-PspCNR1 or SUMO) followed by 350 mM EDTA in 1 M NaCl in order to obtain a more stable surface.

The C1 chip was also used (Series S carboxymethylated matrix-free C1 sensor chips, GE Healthcare). Sensor surfaces were prepared by immobilizing human C3b, C3c and C3d in three of the four flow cells using standard amine coupling. Flow cell 1 served as the reference and was prepared by performing a control coupling reaction, *i.e.* one in the absence of any proteins.

2.5.2 First tranche

For the first tranche of experiments, the construct His-tag PspCN-GSGC was immobilized on the NTA chip. Once the conditioning was completed and a stable baseline was consistently observed the experiments commenced. The chip was

charged with Ni^{2+} - typically ~55 RUs were observed – meaning this was the maximum response obtained. PspCN was flowed at a concentration of 500 μM and ~ 400 RUs were immobilized on flow cell 2 while flow cell 1 was the reference surface and was prepared in the absence of proteins. All experiments were performed at 25 ^{o}C using PBS, 0.01 % (v/v) surfactant P20, supplemented with 50 μM EDTA. 0.5 μM solutions of FH constructs 10-11, 10-15, 8-15, 19-20 as well as full-length FH were injected in duplicates over all flow cells using a contact time of 90 s, a dissociation time of 900 s and a flow rate of 30 μl/min. After every injection the chip was stripped with 350 mM EDTA in 1 M NaCl (180 s injection at 30 μl/min) followed by 50 mM NaOH (30 s at 30 μl/min), followed by two consecutive injections of 0.5 % w/v SDS (30 s at 30 μl/min) and finally a 350 mM EDTA in 1 M NaCl injection (30 s at 30 μl/min) before applying the next sample.

2.5.3 Second tranche

For the second set of experiments the construct His-tagged PspCN (*i.e.* no GSGC C-terminal extension) was used. Once again the His-tag was utilized for immobilization on a NTA sensor chip. Initially, the NTA chip was prepared as previously described and PspCN was immobilized at a level of 90 RUs on flow cell 2 while flow cell 1 served as a no-protein reference. All experiments were performed at 25 ^{o}C using PBS (pH 7.4), 0.01 % (v/v) surfactant P20, supplemented with 50 μM EDTA. Solutions containing 0.5 μM full-length FH and FH constructs 6-8Y, 8-9, 10-15, 8-15, and 19-20 were injected in duplicates over all flow cells using a contact time of 90 s, a dissociation time of 1200 s and a flow rate of 30 μl/min. After every injection the chip was stripped as previously described in section 2.4.2.

2.5.4 Third tranche

For this part of the experiments immobilization of SUMO-PspCN, SUMO-PspCNR1 and SUMO alone (which was the control) – all of these proteins being His-tagged - was performed on an NTA sensor chip. The NTA chip was conditioned as previously described. It was then charged with Ni^{2+} (typically ~55 RUs were

observed). SUMO-PspCN (15 nM) was immobilized on flow cell 2, SUMO-PspCNR1 (15 nM) and SUMO (50 nM) was immobilized on flow cell 3. Flow cell 1 was the reference and was prepared in the absence of proteins. All experiments were performed at 25 ^{o}C using HEPES-buffered 150 mM Saline (pH 7.4), 0.05 % (v/v) surfactant P20, supplemented with 50 μM EDTA.

2.5.4.1 Binding of the various complement proteins to the SUMO-ylated proteins

FH constructs 6-8Y, 8-9, 10-15, 8-15, 19-20 as well as full-length FH and C3b were injected in duplicates over all flow cells using a contact time of 180 s, a dissociation time of 900 s and a flow rate of 50 μl/min. The concentrations used were 50 and 500 nM for all proteins with the exception of full-length FH which was injected only at 50 nM. After every injection the chip was stripped as already described before applying the next sample.

2.5.4.1.1 FH-8-9

FH-8-9 was injected at increasing concentrations; 0, 5, 10, 25, 50, 100, 250 and 2x500 nM. The contact time was 240 s, the dissociation time 1200 s and the flow-rate 50 μl/min. After every injection the chip was stripped as already described before applying the next sample. Dissociation constants were determined by fitting kinetic parameters (from background-subtracted traces) to a one-to-one Langmuir binding model.

2.5.4.1.2 FH-8-15

The procedure was carried out in similarly way to that used for FH-8-9. The FH-8-15 concentrations used were 0, 0.25, 0.5, 1, 5, 10, 25, 50, 75, 2x 100 nM. The contact/ dissociation times, flow rates, regeneration strategies/conditions and determination of dissociation constants were those already mentioned for FH-8-9.

2.5.4.1.3 Full length FH

The procedure was carried out in a similar way as for FH-8-9 and FH-8-15. The concentrations used were 0, 0.25, 0.5, 1, 5, 10, 25, 50, and 2 x 75 nM. The contact/ dissociation times, flow rates, regeneration strategies/conditions and determination of dissociation constants were those already mentioned for FH-8-9 and FH-8-15.

Another series of experiments was aimed at determining whether the PspCN: FH complex, or indeed PspCN on its own, interact directly with C3 or its fragments. For these experiments, the native-sequence PspC constructs, PspCN and PspCNR1, were prepared by proteolytic cleavage of the N-terminal SUMO tag. Steady-state affinity experiments were performed using C1 sensor chips. Full-length FH, PspCN, PspCNR1 and the complexes FH-PspCN and FH-PspCNR1 were injected at increasing concentrations, as described in detail below. All experiments were performed at 25 ^{o}C using HEPES-buffered 150 mM saline, 0.05 % (v/v) surfactant P20, supplemented with 50 μM EDTA.

For full-length FH solutions of 0, 0.02, 0.05, 0.1, 0.2, 0.5, 1 and 2 μM were injected. A contact time of 120 s was used, followed by a dissociation time of 600 s using a flow rate of 30 μl/min, over all flow cells. The chip was regenerated between sample injections by two injections of 1 M NaCl for contact times of 30 s. Dissociation constants were calculated by fitting steady-state binding levels derived from the background-subtracted traces to a one-to-one binding model.

PspCN and PspCNR1 were each injected separately at concentrations of 0, 0.022, 0.055, 0.11, 0.22, 0.55, 1.1 and 2.1 μM. The procedure for measuring dissociation constants was carried out in an identical way to that employed for full-length FH.

For the study of the interactions of the FH-PspCN and FH-PspCNR1 complexes with C3b, C3c, C3d, a stoichiometry of 1:1.1 FH: PspC construct was used. The small excess was intended to ensure that nearly all of FH was engaged in complex formation with PspCN or PspCNR1. The concentration series used consisted of the following mixtures of FH and PspC constructs: 0 μM and 0 μM,

0.02 μM and 0.022 μM, 0.05 μM and 0.055 μM, 0.1 μM and 0.011 μM, 0.2 and 0.22 μM, 0.5 μM and 0.55 μM, 1 μM and 1.1 μM and 2 μM and 2.2 μM. An identical procedure to that used for FH and PspCN alone was used to measure K_D values.

CHAPTER 3

Production, purification and characterization of recombinant proteins

Overview

A full explanation of the rationale behind the selection of protein domains for analysis in the current study, and the choice of expression hosts, was provided in Chapters 1 and 2. Initial work entailed recombinant production, and assessment of suitability for structural studies, of the following proteins: FH-4-6, FH-10-11 and FH-14-15 (Fig. 3.1). Subsequently, ^{15}N-labeled samples of FH-10-11 were prepared for further exploratory NMR studies along with ^{15}N-labeled FH-13-14 and FH-10-15. Ultimately, FH-10-11 was produced under ^{15}N and ^{15}N, ^{13}C-labeling conditions in order to allow determination of 3-D solution structure by NMR. Finally, FH-8-9 and 8-15 were produced for SPR-based binding studies.

The fragments PspCN and PspCNR1 from *S. pneumoniae* were produced and purified at a later stage of the project because they had been reported to interact with the central region of FH (see Chapter 1). Thus these constructs were intended for binding studies. As summarized in figure 3.1, PspCN corresponds to the N-terminal domain of PspC, while PspCNR1 also includes the first of the R repeats, R1. Several versions of these proteins were produced as summarized in Table 3.5, (section 3.3): (i) His-tag PspCN-GSGC which encompassed a His N-terminal tag and a C-terminal non-native amino acid addition (*i.e.* GSGC); (ii) His-tag PspCN, which was the same as (i) but lacked the C-terminal amino acid addition; (iii) the SUMO-PspCN and (iv) the SUMO-PspCNR1 fragments, fusions with SUMO that were employed for the final set of definitive binding experiments.

Pure and properly folded proteins were required for structural studies (using NMR and SAXS) and the binding analyses (using SPR) carried out in this project. Moreover, multiple-milligram quantities of isotopically labeled protein were needed for the NMR work. With this in mind, two expression hosts - one eukaryotic and the other prokaryotic – were exploited for the production of recombinant proteins.

The methylotrophic yeast *P. pastoris* has, for many years, been the expression host of choice in this laboratory for production of proteins containing multiple CCP modules, each of which has two disulfide bonds. Thus it was the obvious choice for production, in the present study, of the FH fragments. As explained more fully in Chapter 2, this eukaryotic expression system provides the

option of fusing the target protein sequence to a cleavable secretory signal, thus ensuring secretion into the media from which it may be purified easily. Yet like a prokaryote, *P. pastoris* requires only minimal nutritional media. On the downside, *P. pastoris* has a tendency to attach high-mannose sugars to N-glycosylation sites. These are normally subjected to post-production enzymatic removal. For example incubation with $EndoH_f$ of a protein bearing high-mannose N-linked glycans results in near-complete deglycosylation, leaving a single GlcNAc at each N-glycosylated site.

E. coli was the logical choice for production of the N-terminal region of the bacterial protein and virulence factor PspC that appears not to require any post-translational modifications. The use of *E. coli* also allowed facile production of fusion proteins to ease subsequent purification. As outlined above, in the current work, SUMO and His tags were utilized as fusion partners.

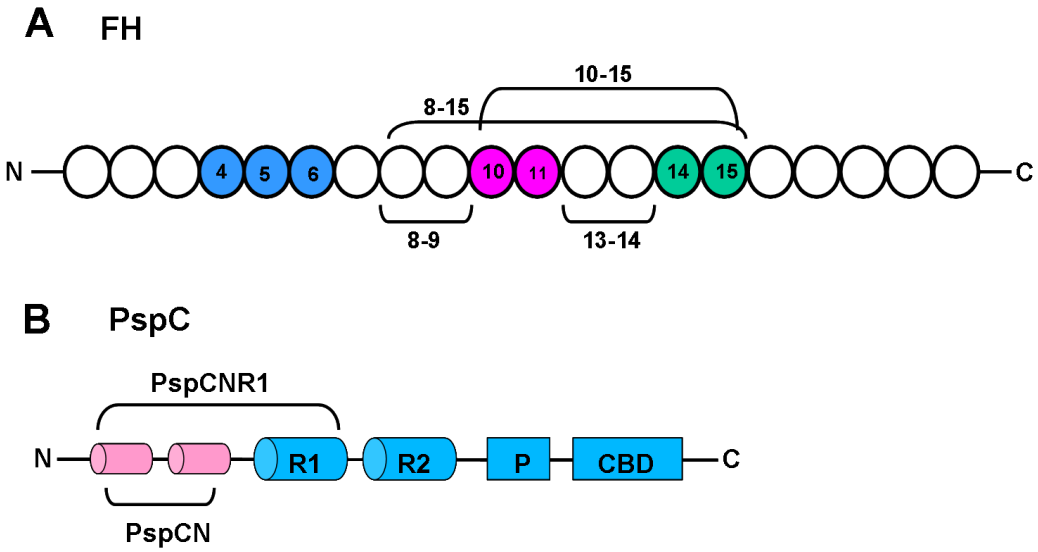


Figure 3.1 Summary of the FH and PspC segments employed in this study. (A) Representation of the FH molecule. Constructs produced by the author are highlighted in color while those that were not are indicated with brackets. (B) Representation of the PspC molecule in which the domains chosen for this study are indicated.

3.1 Production, purification and characterization of the FH constructs.

The table below (Table 3.1) summarizes the FH constructs produced and/or purified in this study, their N-terminal non-native sequences, and the numerical position of each construct within the sequence of native FH.

Construct	*Position in FH*	*Non-native sequence (N-terminal)*
FH-4-6	Residues 209-385	AG
FH-8-9	Residues 447-565	AG
FH-10-11	Residues 568-685	AG
FH-13-14	Residues 752-865	AG
FH-14-15	Residues 810-927	AG
FH-10-15	Residues 568-927	AG
FH-8-15	Residues 447-927	AG

Table 3.1 Summary of the FH constructs produced in this study. The numbering is based on the encoded protein sequence including the 18 amino acid residues of the natural signal sequence. Sequences are from UniprotKB/Swiss-Prot entry P08603.

Table 3.2 (parts 1-3) present a summary of the protein production efforts undertaken in this study for the FH constructs, while Table 3.3 summarizes the expected and experimentally determined – where available - molecular weight for each of the constructs.

FH construct	FH-4-6	FH-8-9 1st batch	FH-8-9 2nd batch	FH-10-11	^{15}N-FH-10-11 1st batch
Expression system	*P. pastoris*	*P. pastoris*	*P. pastoris*	*P. pastoris*	*P. pastoris*
Starting material	FH DNA	Overexpressing clone	Overexpressing clone	FH DNA	Overexpressing clone
Shaker flask/fermentor	Shaker flask	Shaker flask	Shaker flask	Shaker flask	Fermentor
EndoH$_f$ treated	No	Yes	Yes	No	No
Purification step 1	SP-Sepharose	Q-Sepharose	Q-Sepharose	SP-Sepharose	SP-Sepharose
Purification step 2	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion
Purification step 3	-	Size-exclusion	Size-exclusion	-	-
Validation of primary sequence	-	Accurate mass determination by MS	Accurate mass determination by MS	Accurate mass determination by MS	-
Assessment of folding	-	-	-	1D ^{1}H-NMR	^{1}H, ^{15}N-HSQC
Comments	Very low expression	-	Used for SPR	-	Used for NMR structural studies

Table 3.2 Overview of the FH constructs produced in this study (part 1).

FH construct	^{15}N-FH-10-11 2nd batch	^{13}C,^{15}N-FH-10-11 1st batch	^{13}C,^{15}N-FH-10-11 2nd batch	^{13}C,^{15}N-FH-10-11 3rd batch	FH-13-14
Expression system	*P. pastoris*	*P. pastoris*	*P. pastoris*	*P. pastoris*	*P. pastoris*
Starting material	Overexpressing clone	Overexpressing clone	Overexpressing clone	Overexpressing clone	Overexpressing clone
Shaker flask/fermentor	Fermentor	Fermentor	Fermentor	Fermentor	Shaker flask
EndoH$_f$ treated	No	No	No	No	Yes
Purification step 1	SP-Sepharose	SP-Sepharose	SP-Sepharose	SP-Sepharose	SP-Sepharose
Purification step 2	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion
Validation of primary sequence	-	-	-	-	-
Assessment of folding	^{1}H, ^{15}N-HSQC	^{1}H, ^{15}N-HSQC	^{1}H, ^{15}N-HSQC	^{1}H, ^{15}N-HSQC	1D ^{1}H-NMR
Comments	Used for NMR structural studies, SPR	Used for NMR structural studies	Used for NMR structural studies	Used for NMR structural studies	Degraded/ tendency to oligomerize

Table 3.2 Overview of the FH constructs produced in this study (Part 2).

FH construct	^{15}N-FH-13-14	FH-14-15	FH-10-15	^{15}N-FH-10-15	FH-8-15
Expression system	*P. pastoris*	*P. pastoris*	*P. pastoris*	*P. pastoris*	*P. pastoris*
Starting material	Overexpressing clone	FH DNA	Crude culture supernatant	Overexpressing clone	Crude culture supernatant
Shaker flask/fermentor	Fermentor	Shaker flask	-	Fermentor	-
EndoH$_f$ treated	Yes	Yes	Yes	Yes	Yes
Purification step 1	SP-Sepharose	SP-Sepharose	Size-exclusion	Q-Sepharose	Size-exclusion
Purification step 2	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion
Purification step 3	-	-	-	Size-exclusion	-
Validation of primary sequence	-	-	-	-	-
Assessment of folding	^{1}H, ^{15}N-HSQC	-	^{1}H, ^{15}N-HSQC	^{1}H, ^{15}N-HSQC/ ^{1}H, ^{15}N-TROSY	-
Comments	Partially aggregated/unfolded	Very low expression	Used for SPR	-	Used for SPR

Table 3.2 Overview of the FH constructs produced in this study (Part 3).

FH construct	Expected molecular weight (kDa) #$	Measured molecular weight (kDa)
FH-4-6	21,079	—
FH-8-9	13,261	13,864
FH-10-11	13,960	13,960 (No EA), 14,160 (1 EA)
FH-13-14	13,591	—
FH-14-15	13,874	—
FH-10-15	41,253	—
FH-8-15	54,310	—

Table 3.3 Summary of the expected and experimentally determined molecular weights for the FH constructs produced in this study.
 # The expected molecular weight, in kDa, is calculated based on the amino acid sequence including the N-terminal AG cloning artifact.

[$] **An additional 203 Da may be attributed to the single GlcNAc moiety presumed to be left attached at each of the utilized N-glycosylation sites following $EndoH_f$ treatment (FH-8-9 has one glycosylation site, FH-13-14 has two, FH-14-15 has three, FH-10-15 has five and FH-8-15 has six). In addition there might be one or two EA (Glu-Ala) N-terminal dipeptide units (each corresponding to 200.2 Da) originating from the *P. pastoris* expression system present in all constructs.**

3.1.1 Construct FH-4-6

3.1.1.1 Mini-scale expression

Six colonies of this construct were tested in mini-scale expression trials (see section 2.1.15.5 for details and figure 3.2 for results).

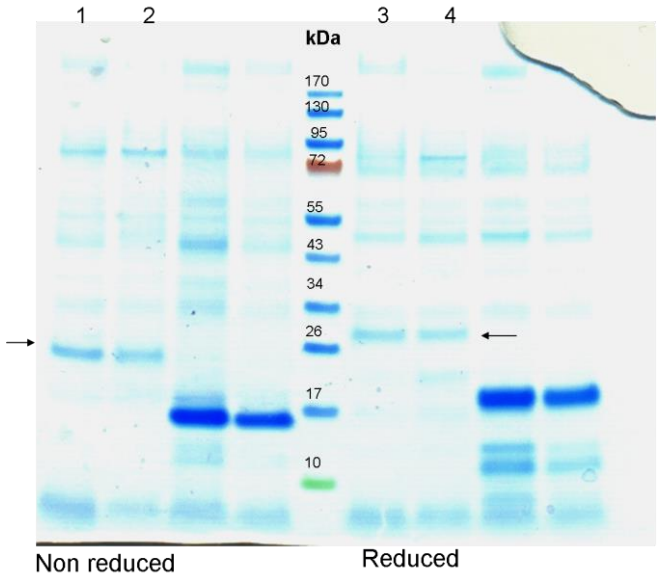


Figure 3.2 Gradient SDS-PAGE of spin-concentrated (initial volume prior to concentration: 500 µl) protein samples. Lanes 1, 2, 3, and 4 correspond to two colonies under reduced and non-reduced conditions. An arrow indicates the band that probably corresponds to FH-4-6. Remaining lanes correspond to other samples, not relevant to this section.

According to SDS-PAGE, all lanes contain bands at about the expected molecular weight for FH-4-6, ~21 kDa. Production levels were judged to be poor based on the strength of the stained bands but protein of an appropriate size was detectable in the SDS-PAGE of spin-concentrated samples. This outcome encouraged a scale-up to medium-scale production of protein in shaker flasks.

3.1.1.2 Medium-scale protein production

The supernatant of a shaker-flask culture (volumes of media: ~500 ml BMG followed by 125 ml BMM) was harvested, PMSF and EDTA were added (unless stated otherwise these were added to the crude supernatants of all the FH constructs- to inhibit proteolysis- presented hereafter) then subjected to cation-exchange chromatography at pH 4.5 (see Materials and Methods section 2.2.6). Protein-containing fractions were concentrated, buffer-exchanged into PBS and loaded onto a size-exclusion chromatography column (Fig. 3.3).

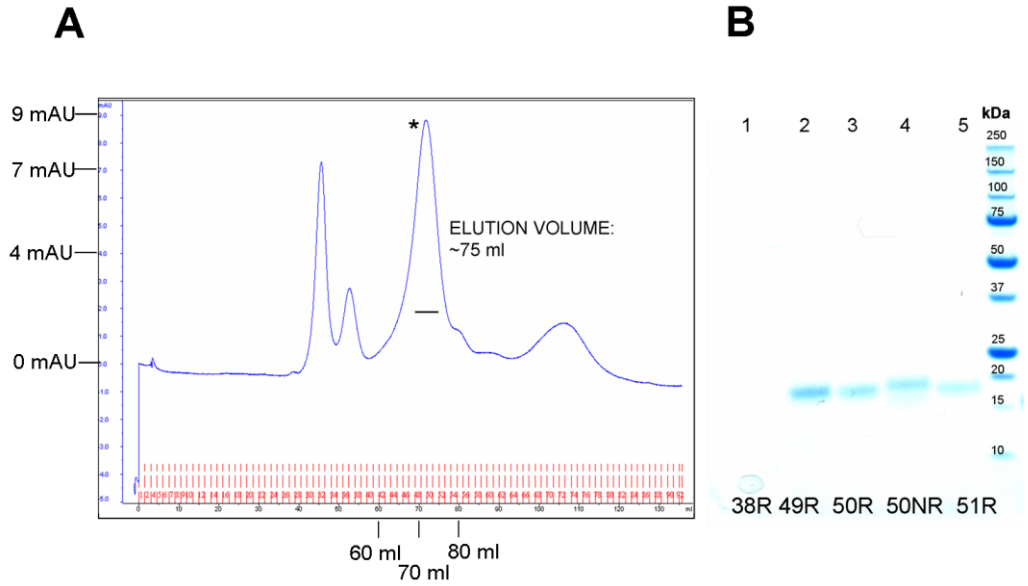


Figure 3.3 Purification of FH-4-6 using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram of the size-exclusion chromatography. (B) Gradient SDS-PAGE gel of the fractions corresponding to the higher peak in (A). Lanes 2-5 were loaded with aliquots from the fractions indicated by a black bar on the chromatogram in (A). R and NR indicate reducing and non-reducing conditions, respectively, and unless stated otherwise this applies to all the SDS-PAGE gels presented henceforth.

The size-exclusion chromatography profile revealed several peaks. For the triple module (~ 22 kDa), elution is expected at ~70 ml, corresponding to the peak in this chromatogram marked with an asterisk. Very small quantities are present, however, as reflected in the low absorbance readings. Subsequent SDS-PAGE of the fractions indicated with a black bar (in Fig. 3.3) all contain a dominant band which appears at ~20 kDa, the appropriate size for the triple CCP module. It was, however, decided not to pursue this investigation further due to the low yield and given the

multiple samples required (with deuteration *etc.*) for solving directly the structure of a triple module.

3.1.2 Construct FH-8-9

A glycerol stock of a *P. pastoris* KM71H clone expressing FH-8-9 was re-streaked on a YPD agar plate containing 100 μg/ml zeocin and single colonies were chosen to inoculate BMG/BMM minimal media for medium-scale production in shaker flasks, as described in Materials and Methods (section 2.1.15.6) (the total volume of BMM for the induction was 750 ml). The purification procedure is described in Materials and Methods (section 2.2.6) and summarized in figure 3.4. The first step was anion-exchange chromatography using a Q-Sepharose column while the second step was size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column. Where a third purification step was required size-exclusion chromatography was performed again. This entire procedure was undertaken twice (*i.e.* in two batches) in order to obtain the desired amount of protein, and representative results are shown below for batch 2 only (while the data for batch 1 may be found in the supplementary CD (section 1.1)).

A point to note regarding the purification strategy was the serial re-loading (3 in total) of the flow-through onto the Q-Sepharose column to recover adequate material. Most of the protein appeared to be in the flow-through of Q-Sepharose loadings 1 and 2 (see Fig. 3.4). Presumably this occurred due to overloading of the column (for example if many contaminants were present) therefore exceeding its capacity. Alternatively, the supernatant may have contained a significant amount of salt therefore the protein of interest did not successfully bind to the column (initially a three-fold dilution of the crude supernatant was performed, loading 1). Data will be presented only for the final loading (loading 3, see Fig. 3.4) on the Q-Sepharose column which corresponded to a final six-fold dilution at pH 7.0 of the initial supernatant. The results of loadings 1 and 2 were samples that did not contain pure FH-8-9 and a pure 80 μM sample in a volume of 0.5 ml respectively.

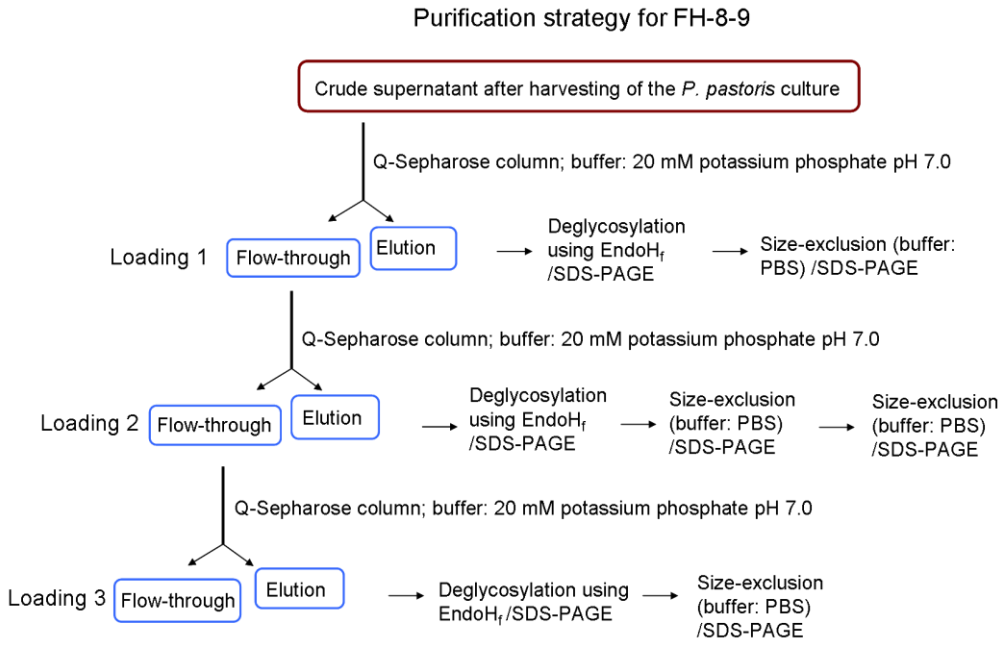


Figure 3.4 Schematic summarizing the strategy followed for the purification of FH-8-9.

3.1.2.1 Representative results for batch 2, loading 3

The Q-Sepharose column was eluted over a linear gradient from 0 to 1 M NaCl (Fig. 3.5 A). Fractions from the elution, as well as samples from the crude supernatant were deglycosylated and the product was analyzed by SDS-PAGE (Fig. 3.5 B).

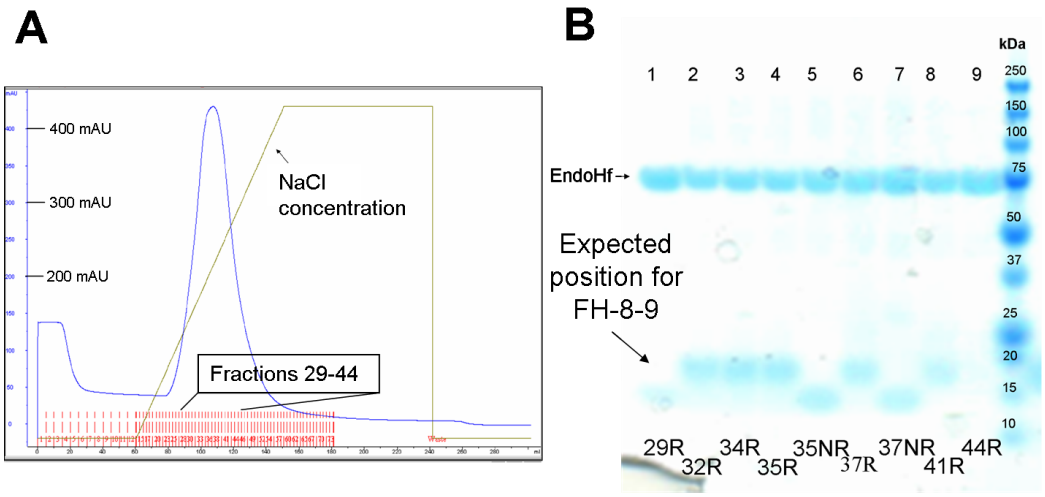


Figure 3.5 Q-Sepharose for loading 3 of FH-8-9. (A) Chromatogram of the elution. (B) Gradient SDS-PAGE of the fractions from the elution of (A) (lanes 1-9).

A single major peak appears in the chromatogram, while the profile of the SDS-PAGE gel showed what appeared to be single bands for FH-8-9 under both reducing and non-reducing conditions at 15 kDa. Fractions 35-40 were combined, $EndoH_f$ treated, and loaded on a size-exclusion chromatography column (Fig. 3.6 A). Fractions from this step were analyzed by SDS-PAGE (Fig. 3.6 B).

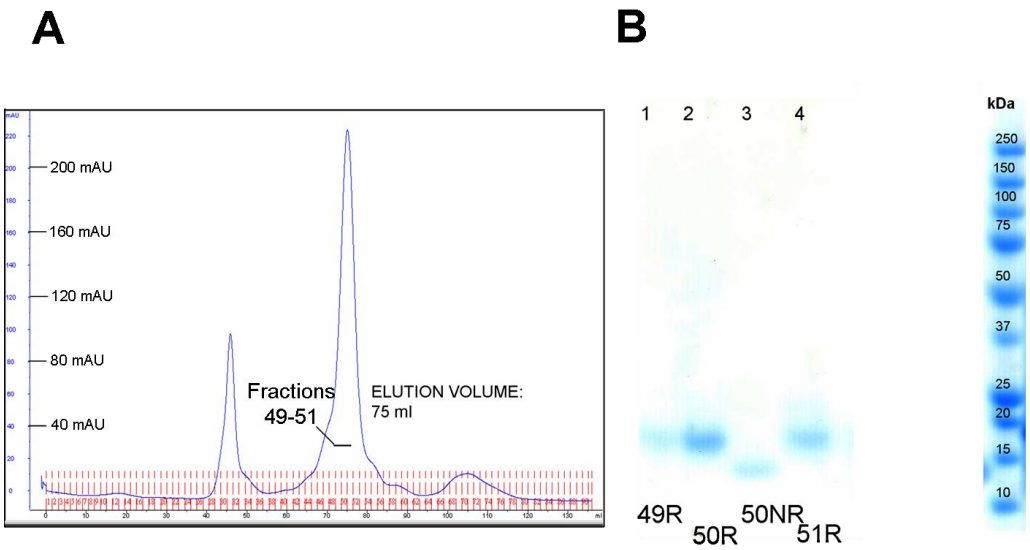


Figure 3.6 Final purification of FH-8-9. (A) Chromatogram of the size-exclusion chromatography run (B) Gradient SDS-PAGE of the fractions from the elution in (A) which are indicated with a black bar (lanes 1-4). The lanes between lane 4 and the molecular weight markers have been masked with a white box, as they correspond to samples not relevant to this section.

The peak indicated with a black bar in figure 3.6, that eluted around 75 ml corresponds to the expected one for a double CCP module of ~ 15 kDa. Fractions from the middle of the peak (that appear as single bands at the expected size of 15 kDa in SDS-PAGE in 3.6 B) shown in figure 3.6 A were combined to afford a final sample at a concentration of 80 μM in 0.5 ml. LC-MS was used for validation of the accurate mass (see Fig. 3.7).

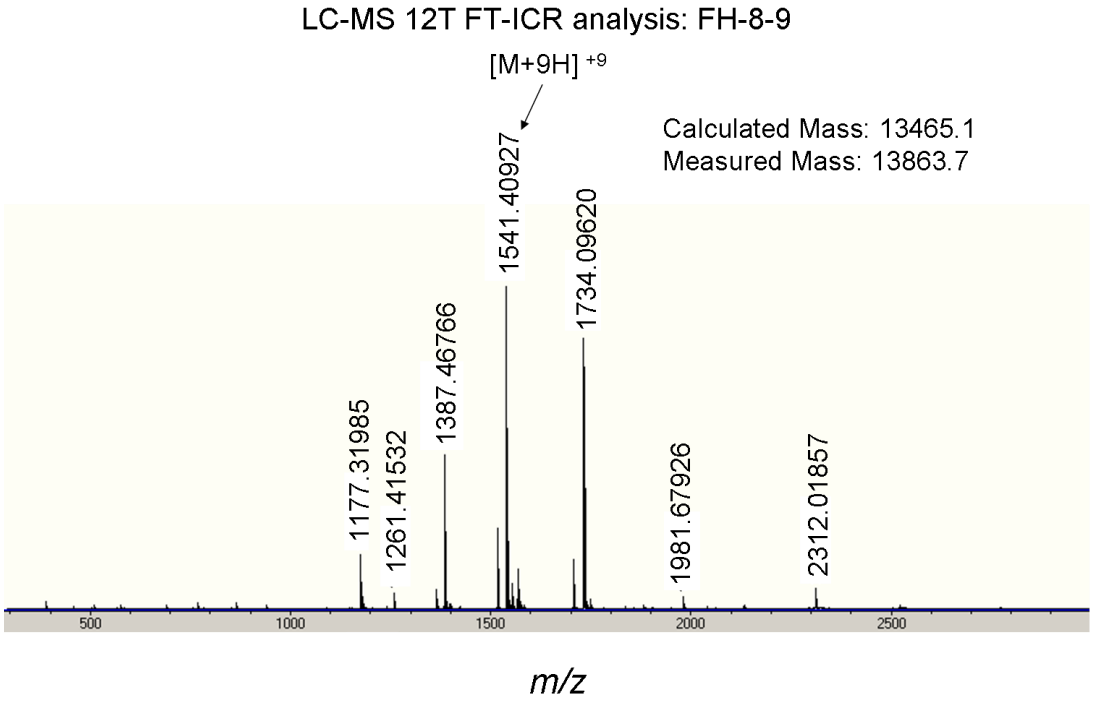


Figure 3.7 LC-MS spectrum of FH-8-9.

The calculated mass for FH-8-9 of 13465.1 Da indicated on the spectrum corresponds to the mass of FH-8-9 which encompasses a GlcNAc attached to the glycosylation site. The measured mass of 13863.7, also indicated on the spectrum, revealed that the protein also carried two N-terminal EA amino acid repeats that remained from the expression system.

In conclusion, despite major technical issues with the ion-exchange step that were not fully understood, work-up of this batch produced pure FH-8-9 according to SDS-PAGE and LC-MS analysis. This was used for the SPR experiments.

3.1.3 Construct FH-10-11

3.1.3.1 Mini-scale and medium-scale (shaker-flask) protein production

This construct was produced with the objective of determining its 3-D solution structure using NMR. Following very promising mini-scale protein production trials (Fig. 3.8) shaker-flask expression (volumes of media: 500 ml BMG followed by 125 ml BMM for the induction) was performed. After harvesting, the supernatant was diluted four-fold with sodium acetate buffer pH 4.5, and the pH was adjusted to 4.5. It was then loaded onto a 5-ml self-poured SP-Sepharose chromatography column and was eluted by gravity using the same buffer as above but containing 1 M NaCl. Fractions were analyzed by SDS-PAGE (Fig. 3.9).

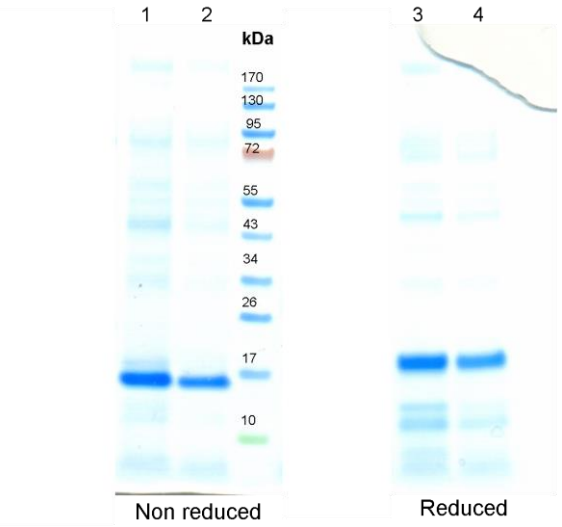


Figure 3.8 Gradient SDS-PAGE of spin-concentrated samples (initial volume prior to concentration was 500 µl) of cell-depleted media from mini-scale protein production trials from two colonies of the FH-10-11 construct (lanes 1-4). The two lanes between the molecular weight markers and lane 3 have been masked with a white box, as these correspond to samples not relevant to this section.

All samples showed bands running at the expected molecular weight for FH-10-11, ~15 kDa (as expected the protein migrates more slowly under reducing conditions). Note that FH-10-11 contains no N-glycosylation sites.

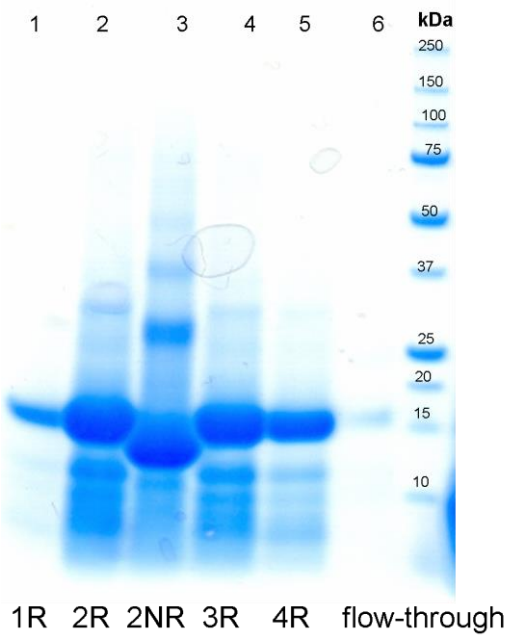


Figure 3.9 Gradient SDS-PAGE of fractions from cation-exchange chromatography performed on the supernatant from a shaker-flask FH-10-11 protein production trial. Lanes 1-5 were each loaded with 15 microlitres from a fraction of 0.5 ml. The band at lane 6 indicates that only a small proportion of the protein was present in the flow-through.

Fractions 1-4 from this elution (shown in figure 3.9) were pooled, buffer-exchanged into PBS, concentrated to 0.5 ml, and loaded onto the size-exclusion column (see also Materials and Methods section 2.2.6). The profile of the elution was typical for a double-CCP module. The dominant peak, which presumably corresponds to FH-10-11, eluted at ~75 ml PBS (Fig. 3.10 A). Fractions 48-53 from this peak were analyzed by SDS-PAGE (Fig. 3.10 B).

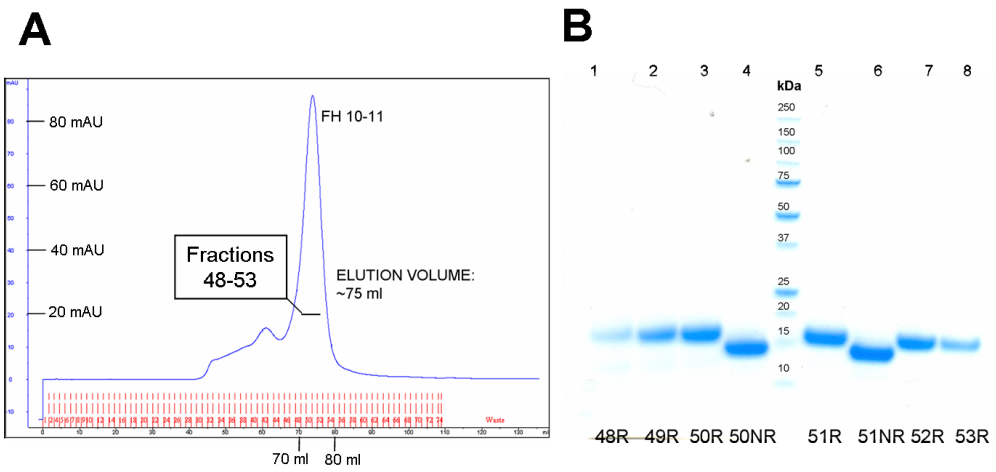


Figure 3.10 Second step in purification of construct FH-10-11 from shaker-flask protein production trial, using a HiLoad 16/60 Superdex 75 column. (A) The higher peak on the chromatogram probably corresponds to FH-10-11 based on the expected elution volume of a 15-kDa protein. (B) Gradient SDS-PAGE of fractions from the size-exclusion chromatography shown in (A): the peak indicated with a black bar, (lanes 1-8).

Fractions 50-53, which yielded single bands at the expected size of 15 kDa in SDS-PAGE (Fig. 3.10 B), were chosen, combined, concentrated to a volume of 0.5 ml and then buffer-exchanged into 20 mM potassium phosphate buffer, pH 6.6. NaN_3 and D_2O were added at 0.01 % w/v and 10 % v/v of the final volume (which was ~ 450 µl), respectively. The final sample which had a concentration of ~80 µM FH-10-11, was used to record a 1-D ^1H NMR spectrum, and for mass analysis (Figs. 3.11 and 3.12).

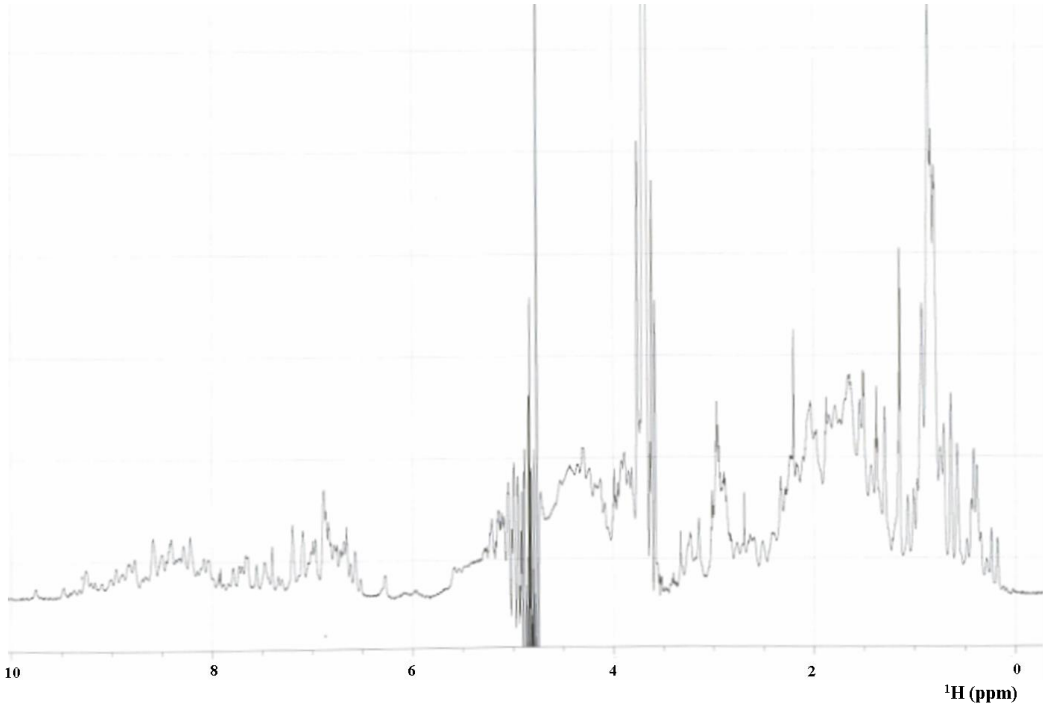


Figure 3.11 1D ^{1}H-NMR spectrum of FH-10-11 from shaker-flask trial.

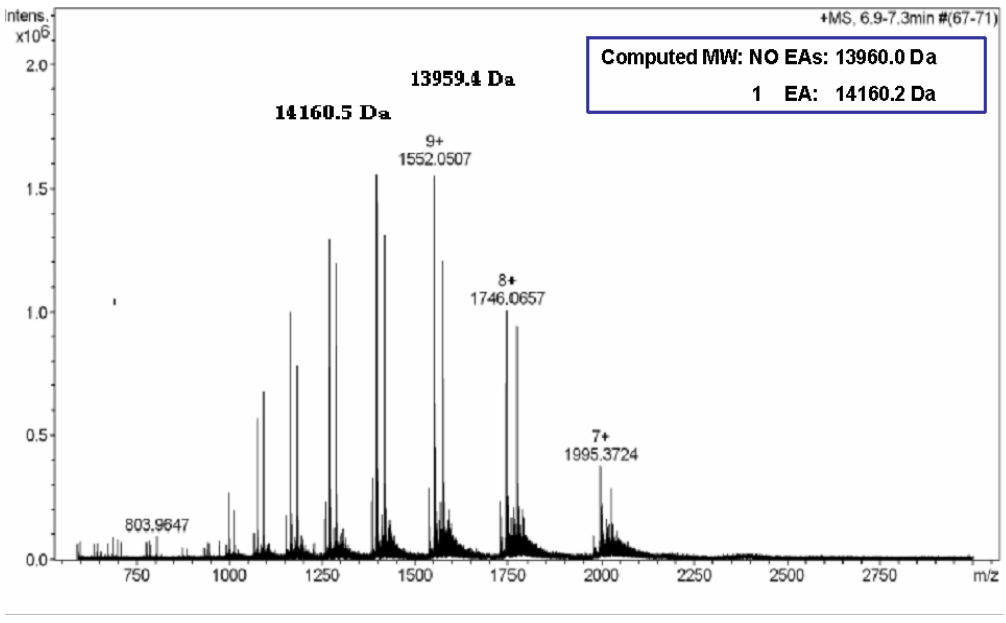


Figure 3.12 LC-MS spectrum for FH-10-11.

The NMR spectrum showed well-dispersed peaks, indicative of proper folding of the protein (Fig. 3.11). From mass spectrometry it can be clearly seen that there are two main species (peaks at charge state 9+). The molecular weight of

14160.5 Da corresponds to one EA addition (~200.2 Da) (compared to theoretical, 14160.2, assuming four disulfide bonds); whereas 13959.4 Da is close to the theoretical mass (13960.0, again assuming four disulfide bonds) expected if there are no EA additions (Fig. 3.12).

3.1.3.2 ^{15}N-enriched expression in a fermentor

Two ^{15}N-labeled samples were produced in total in order to ensure an adequate supply of protein for all the necessary NMR experiments. The protocol used for the fermentation is detailed in the Materials and Methods chapter, section 2.1.15.7. Data will be presented only for batch 1 (while the corresponding data for batch 2 may be found in the supplementary CD section 1.2).

3.1.3.2.1 Representative results for batch 1

After fermentation the cell-depleted growth media (~900 ml) was filtered. It was then diluted five-fold with distilled water, and the pH was adjusted to 4.5 prior to cation-exchange chromatography (see also Materials and Methods section 2.2.6) on a SP-Sepharose column. The protein was eluted by using sodium acetate buffer, pH 4.5, with a gradient from 0 M NaCl to 1 M NaCl. A major peak started eluting at ~45 % of the 1 M NaCl solution (~ 450 mM NaCl, Fig. 3.13 A). Corresponding fractions from the elution, as well as sample from the flow-through, were analyzed by SDS-PAGE (Fig. 3.13 B and C).

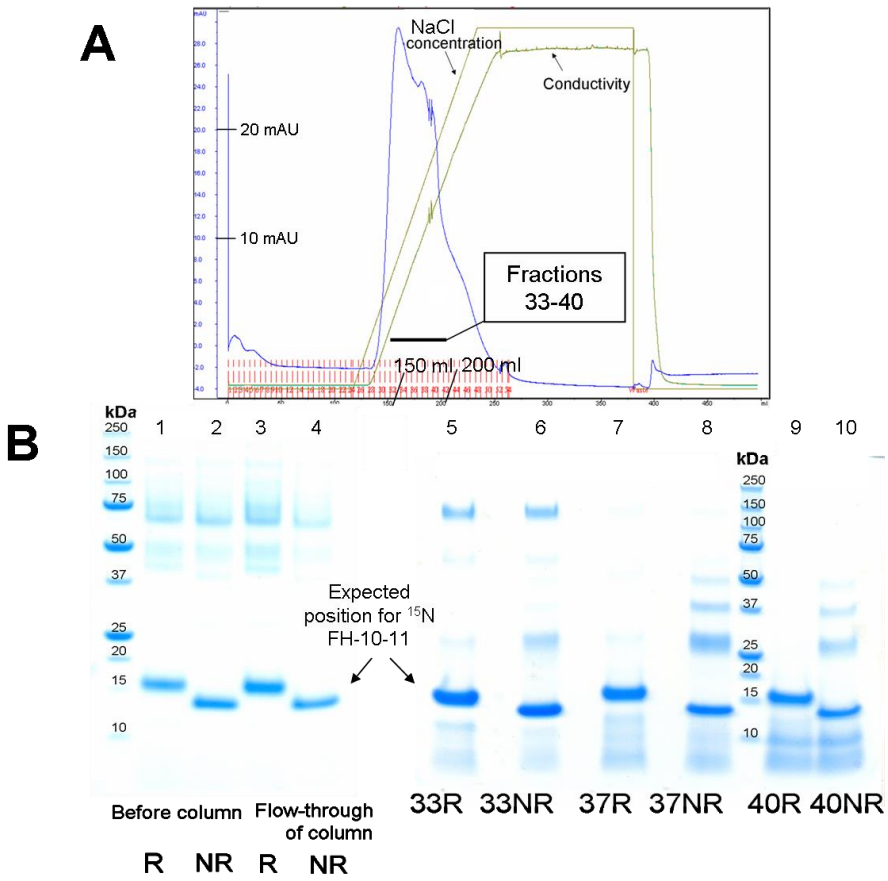


Figure 3.13 Chromatography of ^{15}N-FH-10-11 on SP-Sepharose. (A) Chromatogram (B) Gradient SDS-PAGE of: the sample prior to loading on column (lanes 1 and 2), the flow-through (lanes 3 and 4), and the collected fractions (lanes 5-10) corresponding to the peak indicated with a black bar shown in (A).

The chromatogram showed two main peaks and fractions from the region indicated with a black bar were analyzed by SDS-PAGE. In fractions from the crude supernatant bands at ~15 kDa presumably correspond to FH-10-11. In the samples from the flow-through, it is evident that protein was still present. In the ten-fold concentrated aliquots from the indicated fractions bands at ~15 kDa are visible in all lanes corresponding to the expected size for FH-10-11.

Surprisingly most of the protein was in the flow-through while the UV absorbance of the dominant peak eluting during the gradient was relatively low. Hence, the flow-through (~4800 ml) was further diluted by adding 1.5 L of distilled water and the pH was adjusted to 4.0. All subsequent cation-exchange chromatography steps on FH-10-11 were conducted at this lower pH using sodium acetate buffer. The diluted sample was reloaded onto the SP-Sepharose column and

eluted with a NaCl gradient as before. The fractions (Fig. 3.14 A) were analyzed by SDS-PAGE (Fig. 3.14 B and C).

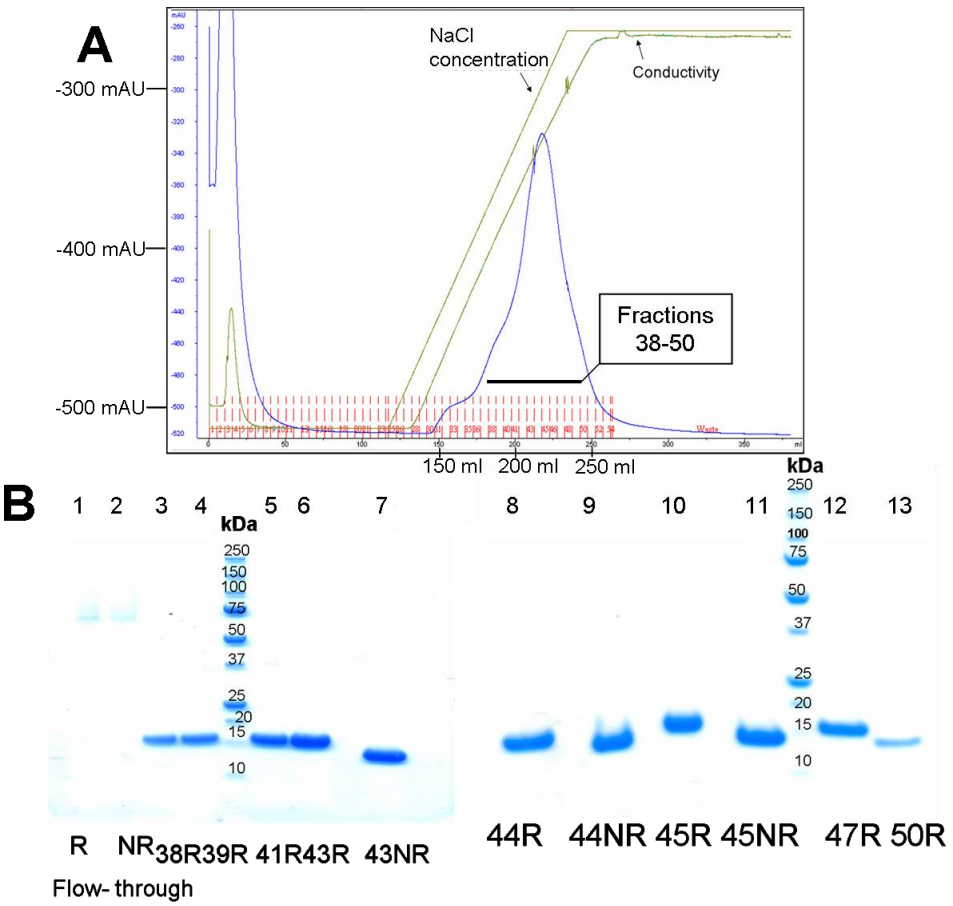


Figure 3.14 Reloading of the flow-through from first SP-Sepharose run. (A) Chromatogram. (B) The outcome of gradient SDS-PAGE on a sample from the flow-through (lanes 1 and 2) as well as fractions from the region indicated with a black bar in (A) as denoted below the gel (lanes 3-13) .

Fractions from the elution (shown above) corresponding to the region on the chromatogram indicated with a black bar were analyzed by SDS-PAGE. No bands of FH-10-11 are visible for the samples corresponding to the flow-through of the elution. Fractions from the elution contain bands at the expected molecular weight for the target protein, ~15 kDa.

Fractions 41-51 from this re-loaded SP-Sepharose run were combined, concentrated to a volume of 1 ml, buffer-exchanged into PBS (pH 7.6), filtered and applied to the size-exclusion chromatography column for a final polishing step (Fig. 3.15 A). Fractions from the elution were analyzed by SDS-PAGE (Fig. 3.15 B).

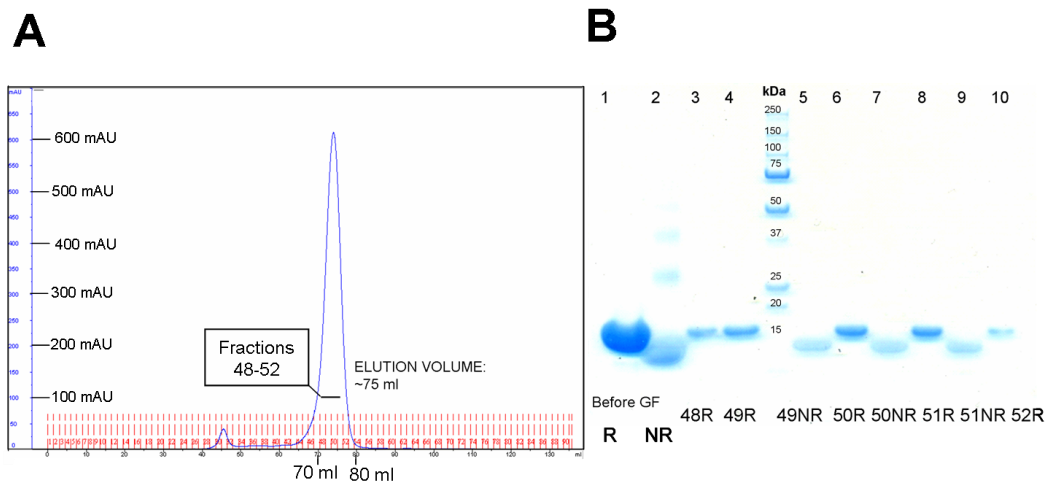


Figure 3.15 Size-exclusion chromatography for the FH-10-11 samples from the re-run of the SP-Sepharose step, using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram. (B) Gradient SDS-PAGE of a sample before the gel-filtration (lanes 1 and 2) as well as fractions from the peak indicated with a black bar in (A) and as indicated below the gel (lanes 3-10).

The SDS-PAGE indicated that small extra bands appearing, in non-reducing conditions, prior to loading on the gel-filtration column had been removed from fractions corresponding to the main peak in the chromatogram.

Fractions 47-52 from the gel-filtration step were concentrated in spin-concentrators to a volume of ~0.4 ml and buffer-exchanged into 20 mM potassium phosphate buffer, pH 6.6. The final sample was 610 µM.

3.1.3.3 ^{15}N, ^{13}C-enriched expression in a fermentor

Due to unanticipated degradation problems (evidenced by the state of the ^{15}N-labeled sample following the NMR experiments, data not shown) with the double-labeled FH-10-11 sample the entire expression and purification procedure had to be performed three times in order to obtain sufficient material for the NMR experiments necessary for the structural studies. The protocol used for fermentations is detailed in the Materials and Methods (section 2.1.15.7). The purification procedure for all three batches of double-labeled samples was identical to that followed for the ^{15}N-labeled sample. Representative results will be shown for batch 1 (data on batches 2 and 3 can be found in the supplementary CD sections 1.3 and 1.4).Batches 2 and 3 resulted in purified final samples at concentrations of 150 and 500 µM respectively (in volumes of ~ 450 µl).

3.1.3.3.1 Representative results for batch 1

The first purification step was cation-exchange chromatography at pH 4.0, of the diluted cell-depleted growth media (volume: ~700 ml), using an SP-Sepharose column (Fig. 3.16 A). Fractions from the elution of the column were analyzed by SDS-PAGE (Fig. 3.16 B and C).

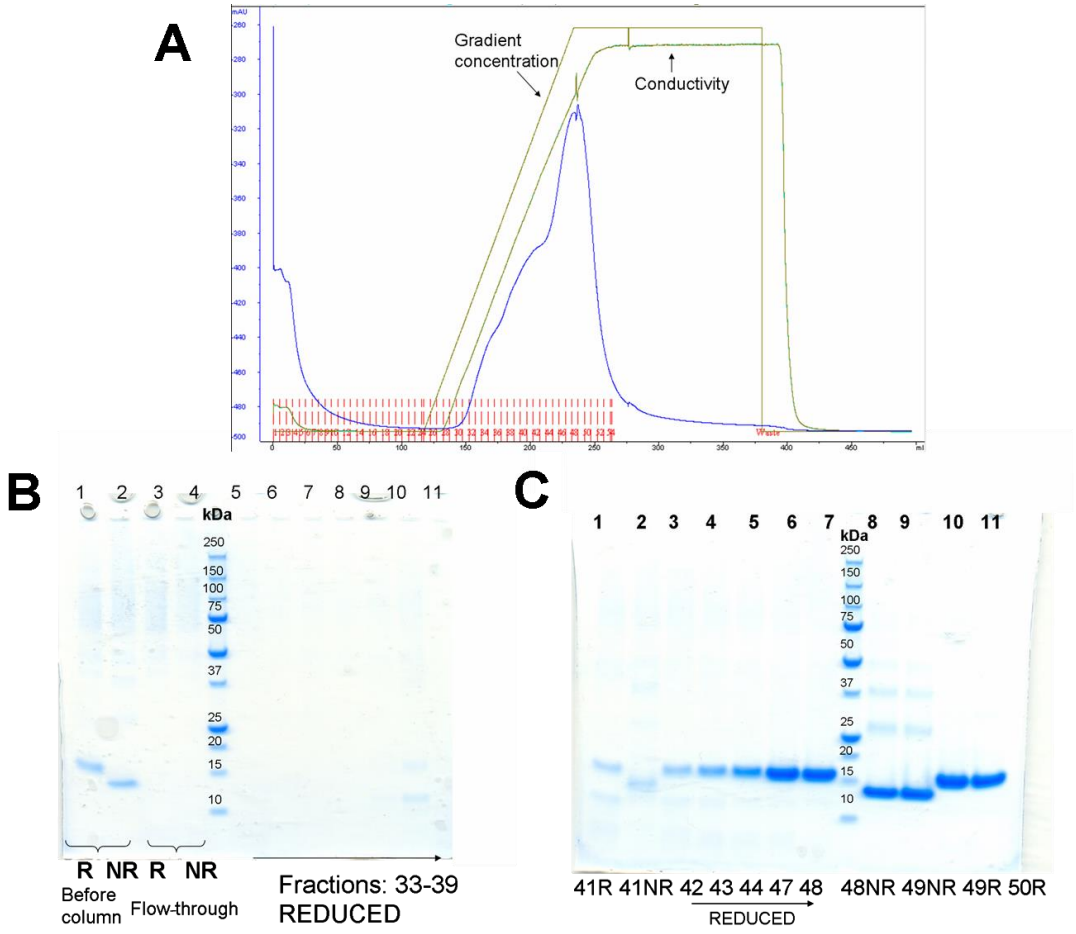


Figure 3.16 Purification of the double-labeled FH-10-11 sample, batch 1, on SP-Sepharose (A) Chromatogram showing which fractions were pooled for SDS-PAGE (B) Gradient SDS-PAGE of: Lanes 1, 2 - sample before loading onto the column; Lanes 3, 4 – ten-fold concentrated sample taken from the flow-through of the column; lanes 5-11 - fractions 33-39 corresponding to peak 1 as indicated in (A). (C) In this gel, lanes 1-5 are for fractions 41-44 corresponding to peak 2 as indicated in (A). Lanes 6-11 - fractions from the higher peak, peak 3 as shown in (A).

The chromatogram contained at least three overlapped peaks. According to SDS-PAGE, fractions from peak 3 (Fig. 3.16 C) contained stronger bands, compared to those from other peaks, at the expected size for FH-10-11. Fractions 47-53 from

that peak (see Fig. 3.16 A) were concentrated in spin concentrators and buffer-exchanged into PBS for size-exclusion chromatography (Fig. 3.17 A). The fractions from the elution of the size-exclusion column were analyzed by SDS-PAGE (Fig. 3.17 B).

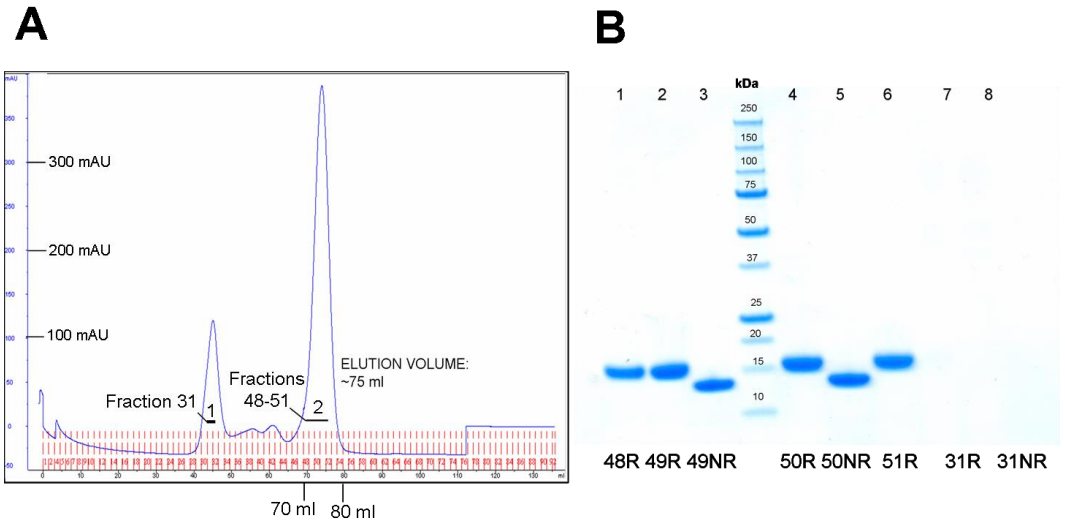


Figure 3.17 Size-exclusion chromatography of fractions 47-53 from the SP-Sepharose run, using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram. (B) Gradient SDS-PAGE. Lanes 1-8 correspond to fractions 48-51 from Peak 2. Lanes 7 and 8 correspond to fraction 31 from Peak 1.

Fractions from peak 2 in the chromatogram (Fig. 3.17) produced single strong bands of the expected size for FH-10-11 in SDS-PAGE (Fig. 3.17). Fractions 48-53 from the centre of this peak were concentrated and buffer-exchanged three times into 20 mM potassium phosphate, pH 6.6. This yielded a 400 µM sample in ~450 µl.

3.1.4 Construct FH-13-14

Overview

This construct was produced in order to asses its suitability for NMR-based structural analysis. A glycerol stock of a *P. pastoris* KM71H clone expressing FH-13-14 was re-streaked on a YPD agar plate containing 100 µg/ml zeocin and single colonies were used to inoculate BMG media. Unlabeled protein was produced in

shaker flask while ^{15}N-enriched protein was produced in a fermentor. Identical purification procedures were followed in both cases: cation-exchange chromatography on a SP-Sepharose column (sodium acetate buffer, pH 5.0 and eluted with a gradient from 0 - 1 M NaCl) was followed by size-exclusion chromatography on a HiLoad Superdex 75 column.

3.1.4.1 Shaker-flask production of unlabeled protein

The cell-depleted growth medium (375 ml) was diluted six-fold and the pH was adjusted to 5.0. It was then applied to the pre-equilibrated SP-Sepharose column (Fig. 3.18 A). Fractions were deglycosylated by incubation with EndoH$_f$ and analyzed by SDS-PAGE (Fig. 3.18 B and C).

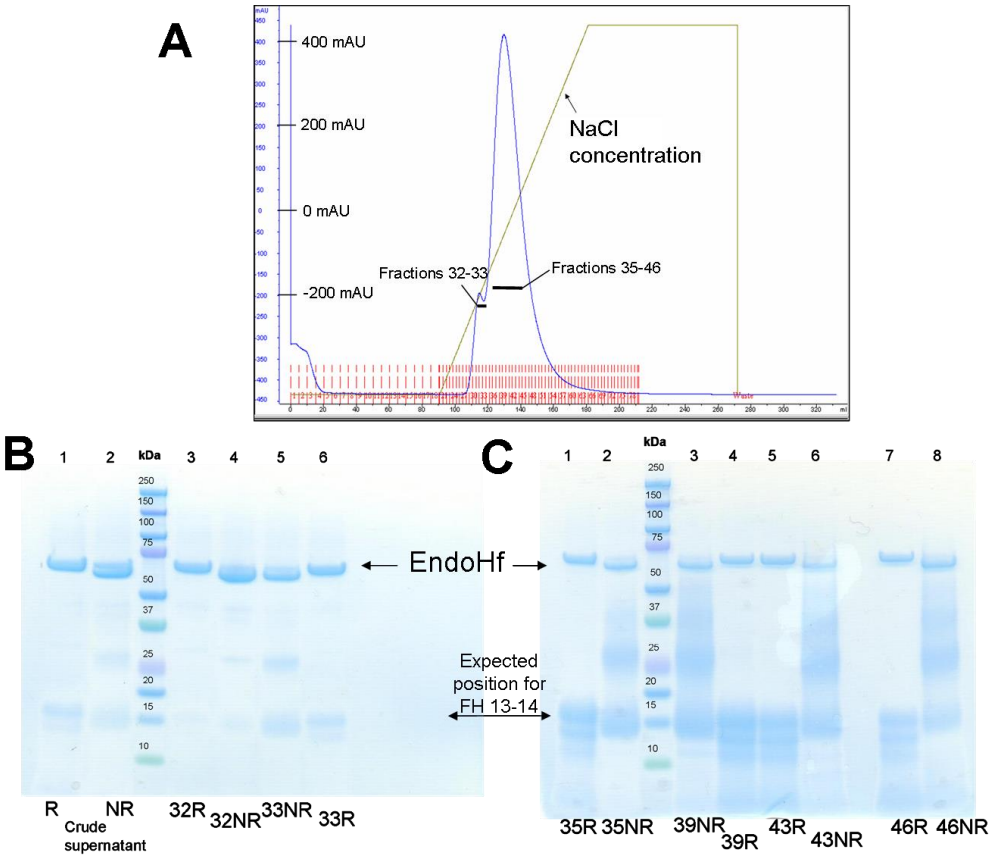


Figure 3.18 Cation-exchange chromatography of unlabeled FH-13-14. (A) Chromatogram from the elution of the SP-Sepharose column. (B), (C) Gradient SDS-PAGE of fractions from the elution shown in (A). (B) Lanes 1, 2: samples from crude supernatant immediately after harvesting. Lanes 3-6: samples from fractions as indicated below the gel. (C) Samples from the fractions indicated below the gel.

The chromatogram contained a small peak on the leading shoulder of a much larger main peak. The SDS-PAGE analysis revealed that samples (from fractions 35-46) prepared under non-reducing conditions contained multiple bands with the smallest running as expected for FH-13-14. On the other hand, under reducing conditions multiple bands appeared below a band of the expected molecular weight. This could be explained by the presence in the samples of proteolytically generated polypeptides held together (under non-reducing conditions) by disulfide bonds. Fractions 35-46 were combined, Endo H_f-treated, concentrated and buffer-exchanged into PBS. The sample was then divided into two equal portions and each was loaded, separately, onto a size-exclusion chromatography column (Fig. 3.19 A and C). Fractions were analyzed by SDS-PAGE (Fig. 3.19 B and D).

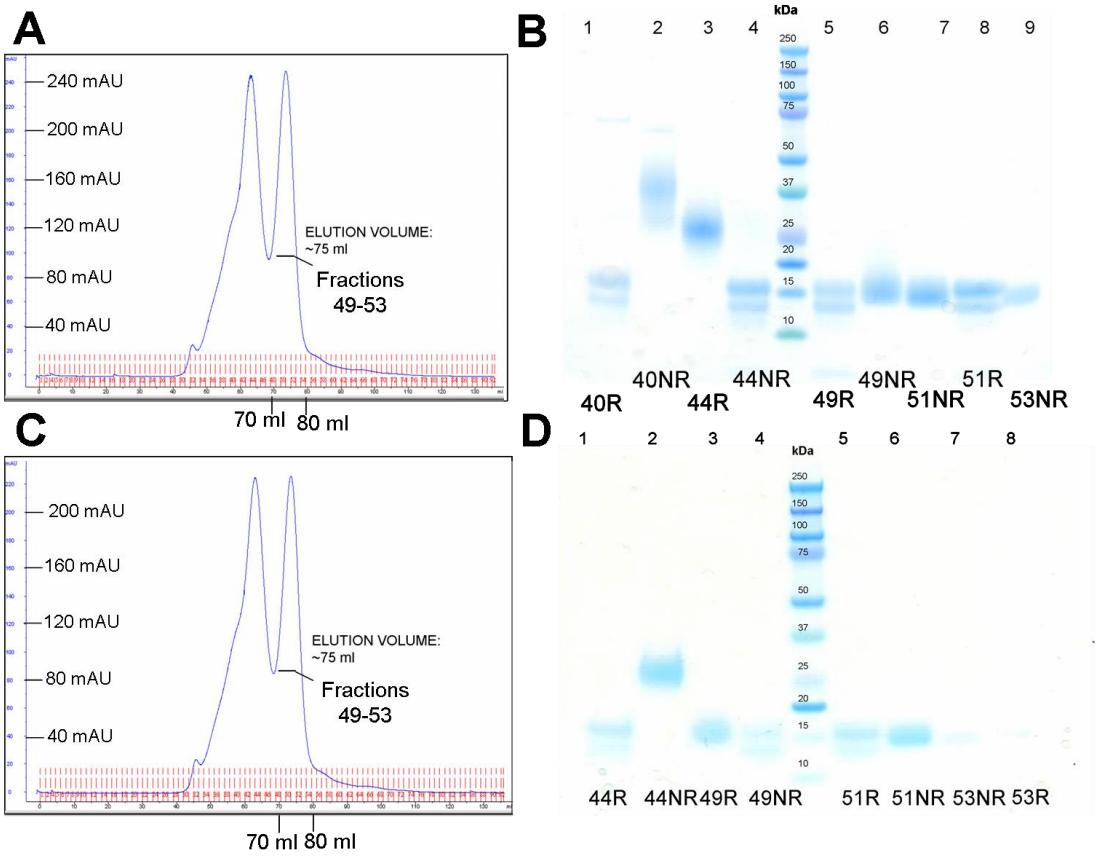


Figure 3.19 Size-exclusion chromatography of FH-13-14, using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram of the first batch of two. (B) Gradient SDS-PAGE of fractions from (A). Lanes 1-9: samples from the fractions indicated below the gel. (C) Chromatogram of the second batch of two. (D) Gradient SDS-PAGE of fractions from (C). Lanes 1-8: sampled from the fractions indicated below the gel. The peaks that probably correspond to FH-13-14 are indicated with a black bar.

The two size-exclusion chromatography runs performed showed identical elution profiles. The first peak probably contains a dimer (or oligomer(s)) formed by FH-13-14 that persist under denaturing conditions according to the corresponding SDS-PAGE. The larger peak eluted at the volume expected for a double-CCP module. It yields a single band on SDS-PAGE under non-reducing conditions, but there is evidence of degradation in SDS-PAGE loaded under reducing conditions.

Fractions 49-53 were pooled separately for each size-exclusion chromatography run. Following spin-concentration the first batch of fractions yielded a sample of ~100 μM protein that was stored at -80 ^{o}C. The second batch produced a sample of 150 μM protein in ~450 μl. This was buffer-exchanged in 20 mM potassium phosphate, pH 6.6, the pH was adjusted to 6.7, and this sample was used to record a ^{1}H-NMR spectrum (Fig. 3.20). The spectrum shows neither a very good dispersion nor well resolved peaks probably due to the presence of partially unfolded material. In addition, it revealed the presence of carbohydrate contaminants (multiple sharp peaks at around 4.8 ppm, (147)). These non-N-linked glycans are most likely non-specifically associated with the protein as was observed in the case of, for example, β-Lactoglobulin produced in *P. pastoris* (147). The sample was therefore applied to a ConA-lectin resin column in order to remove any remaining non-covalently attached sugar moieties. SDS-PAGE was subsequently performed on a sample prior to application on the column as well as fractions from the elution (Fig. 3.21).

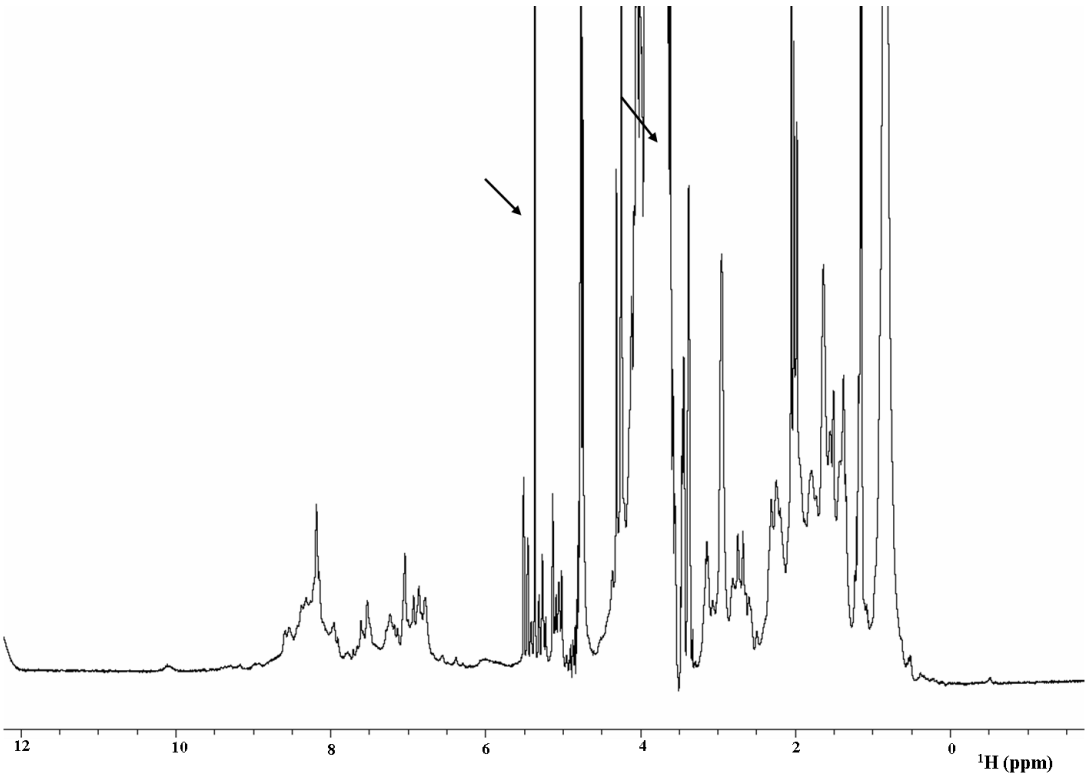


Figure 3.20 1D ^{1}H-NMR spectrum of FH-13-14. The sample was 150 µM in 20 mM potassium phosphate buffer, pH 6.7. Arrows indicate signals arising from contaminating carbohydrates.

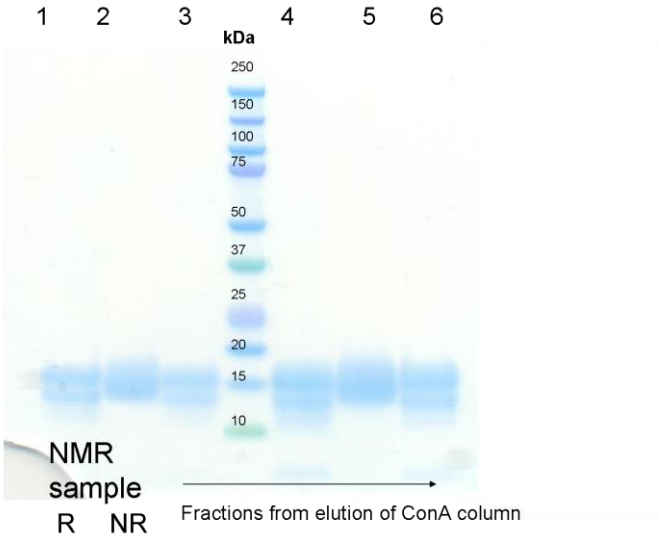


Figure 3.21 Gradient SDS-PAGE performed before and after elution from the ConA column. Lanes 1 and 2 contain the NMR sample (the sample originating from the second batch of size-exclusion chromatography, prior to loading on the ConA column). Lanes 3-6 contain sequential fractions from the elution of the column.

Fractions from the elution of the ConA column showed bands consistent with the profile already observed for FH-13-14. This exercise verified that the recombinant protein was not N-glycosylated presumably as a result of $EndoH_f$ treatment. Hence the polysaccharides evident in the ^{1}H-NMR spectrum shown in figure 3.20 do not correspond to N-linked glycans. However, whether carbohydrate moieties were still present was not tested further by NMR.

3.1.4.2. ^{15}N-enriched expression in a fermentor

The sample described above appeared to be highly susceptible to proteolytic degradation according to SDS-PAGE run under reducing conditions as well as improperly or incompletely folded according to ^{1}H-NMR. Therefore FH-13-14 behaves differently to the vast majority of CCP module pairs that have been studied in this laboratory. Since FH-13 is compactly folded and stable as an individually expressed module (30), it was hypothesized that FH-14 must be the source of these problems. To test this, production of a ^{15}N sample for FH-13-14 was undertaken.

Fermentation, with ^{15}N enrichment, was performed as described (Materials and Methods, section 2.1.15.7). The culture was harvested and the supernatant (~900 ml) was diluted tenfold (rather than threefold, in an effort to reduce the loss of protein in the flow-through from ion-exchange chromatography, as observed for FH-8-9 and FH-10-11 – see above). The supernatant was applied to a SP-Sepharose column equilibrated with 20 mM sodium acetate buffer pH 5.0. The protein was eluted over a linear gradient from 0 to 1 M NaCl (Fig. 3.22 A). Fractions from this capture step were $EndoH_f$-treated and analyzed by SDS-PAGE (Fig. 3.22 B).

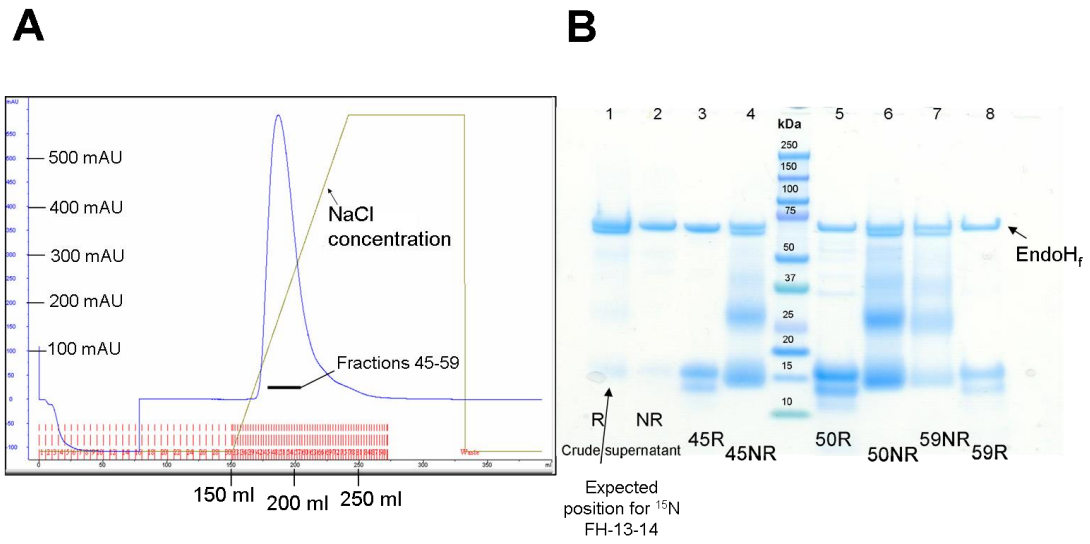


Figure 3.22 Cation exchange chromatography of ^{15}N-FH-13-14 on SP-Sepharose column. (A) Chromatogram (B) Gradient SDS-PAGE Lanes 1-2; before loading. Lanes 3-8: fractions as indicated.

Protein eluted as a single peak early on in the gradient. SDS-PAGE revealed a similar pattern of bands to that observed with the non-labeled sample; oligomers under non-reducing conditions and degradation products when the disulfide bonds had been reduced. There was little variation between fractions (taken from across the peak) in terms of the relative quantities of oligomers/degradation products detected so it was concluded that these all co-eluted with the target material. In an attempt to enrich for the target protein, fractions 43-62 were combined, EndoH$_f$-treated, concentrated and buffer exchanged into PBS for size-exclusion chromatography. The pooled sample was then split into two equal portions and loaded separately (Fig. 3.23 A and C). The resultant fractions were monitored by SDS-PAGE (Fig. 3.23 B and D).

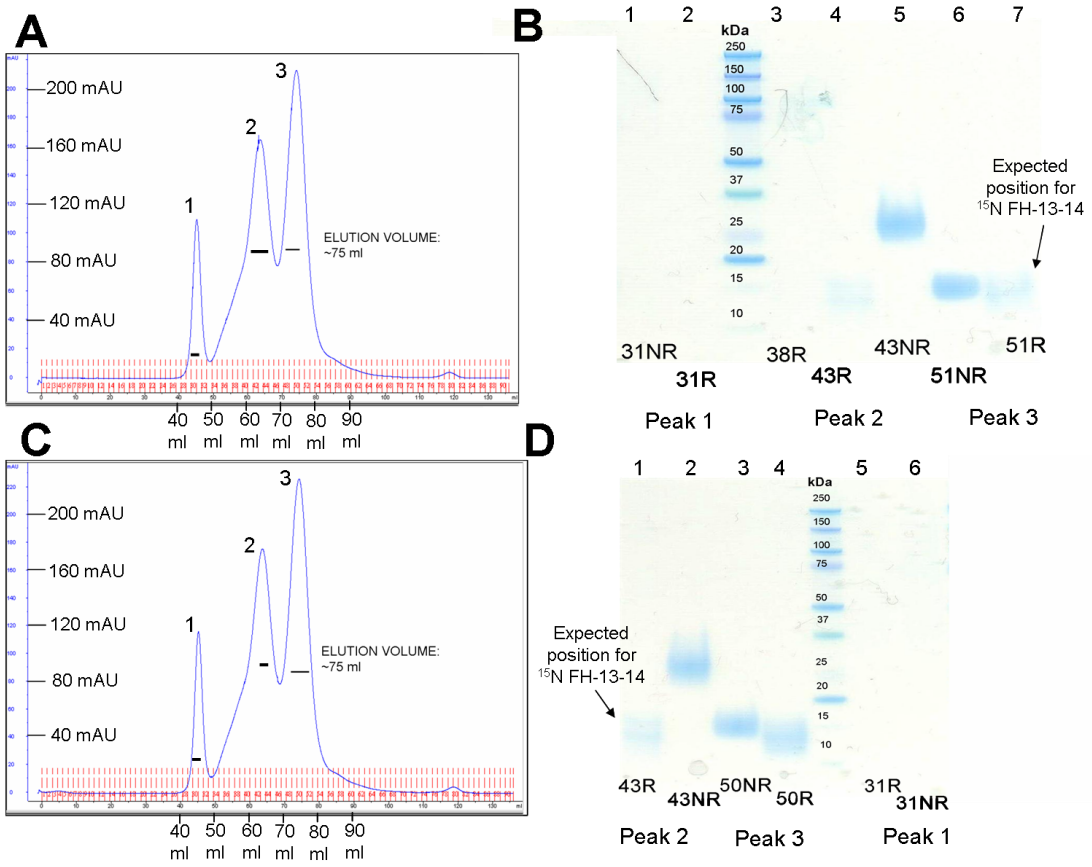


Figure 3.23 Size-exclusion chromatography of ^{15}N-FH-13-14 (batches 1 and 2), using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram for first batch. (B) Gradient SDS-PAGE (for first batch). Lanes 1-7: fractions as indicated below gel. (C) Chromatogram for second batch. (D) Gradient SDS-PAGE for second batch Lanes 1-6: fractions as indicated below the gel.

For both batches, three main peaks were eluted from the size-exclusion chromatography column, with the latest-eluting and highest peak in each case corresponding to material of the expected size for FH-13-14. The SDS-PAGE analysis revealed that individual peaks corresponded to what appeared to be single bands (under both non-reducing and reducing conditions) indicating that oligomers had been resolved from the monomer by this method. Unfortunately, closer inspection of the actual gels indicated that extra bands were present (not easily discernible in the figure above) under reducing conditions corresponding to degradation products. This is consistent with cleavage having occurred within the FH-13-14 polypeptide – at some stage in protein production, harvesting or purification - that is only revealed when the disulfides are reduced. Nonetheless, fractions 48-52 of the two runs were combined separately. The sample originating from the first size-exclusion chromatography was ~150 µM in 300 µl volume. It was

subsequently buffer-exchanged into 20 mM potassium phosphate buffer, pH 6.6, and used to record the ^{1}H, ^{15}N-HSQC spectrum shown in figure 3.24. The pooled fractions from gel-filtration chromatography of the second batch were stored at -80 ^{o}C.

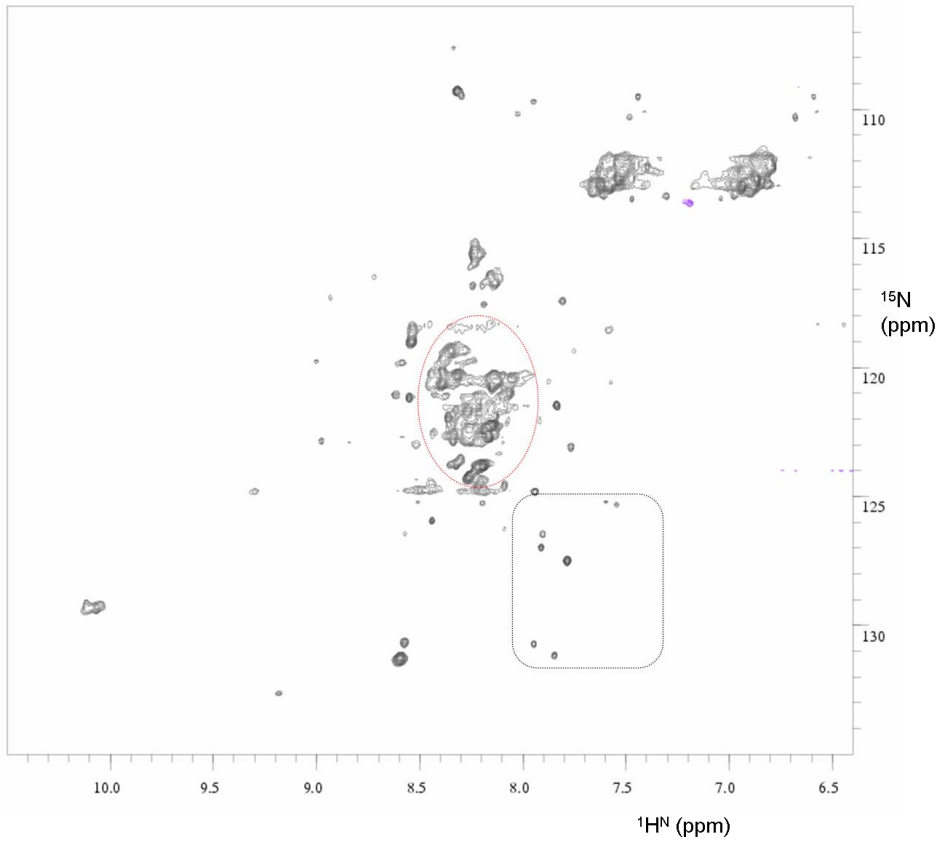


Figure 3.24 ^{1}H, ^{15}N-HSQC spectrum of ^{15}N-FH-13-14. The sample was 150 µM in 20 mM potassium phosphate buffer, pH 6.6. A black box highlights unidentified cross-peaks that have previously been observed in degraded FH fragment preparation, while a red circle highlights one of the clusters of overlapped peaks potentially due to the presence of unfolded material.

The presence of aggregated and unfolded material is evident in the ^{1}H, ^{15}N-HSQC spectrum from the non-uniformity of peak intensities and line-widths and the heavily overlapped clusters of peaks with random-coil shifts (for example in the middle region of the spectrum, also indicated with a circle). In addition there are weak but well-dispersed peaks that might correspond to compactly folded protein. It is difficult to judge the relative proportions of folded and unfolded/aggregated material. Close inspection of the overlay of the acquired ^{1}H, ^{15}N-HSQC with the ^{1}H, ^{15}N-HSQC for FH-13 alone (which shows that the protein is properly folded), (Fig.

3.25) revealed that a small proportion of the well-dispersed peaks observed in the former could possibly arise from FH-13 (highlighted with circles in figure 3.25). Although both samples were in the same buffer and conditions it is likely that there could be a deviation in the actual pH between the two (e.g. due to experimental error when adjusting pH or calibration of the pH meter) which could in turn affect the chemical shift positions of the residues. All in all it seems probable that FH-14 as well as potentially the linker between the two modules are responsible for the properties of the 13-14 pair.

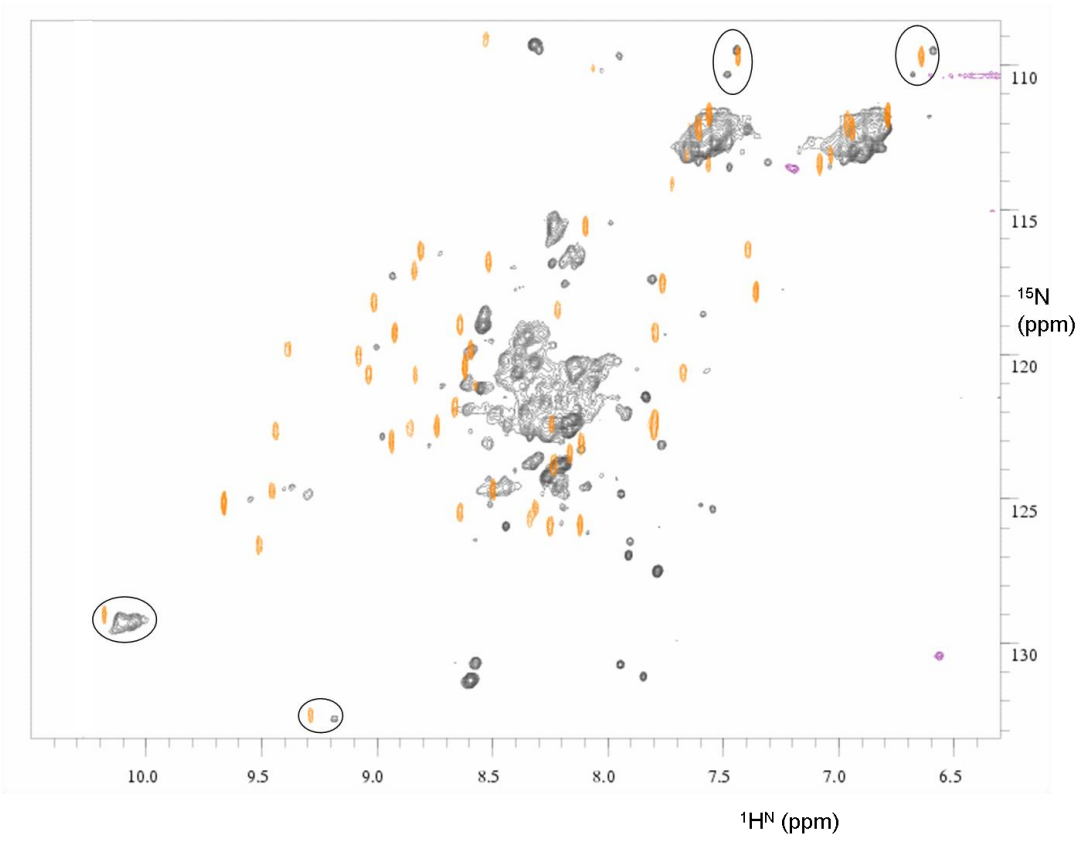


Figure 3.25 Overlay of ^{1}H, ^{15}N-HSQC spectrum of ^{15}N-FH-13-14 (shown in dark grey) with ^{1}H, ^{15}N-HSQC spectrum of ^{15}N-FH-13 (shown in orange). Circles highlight cross-peaks that may possibly belong to FH-13. Both samples were in 20 mM potassium phosphate buffer pH 6.6.

In conclusion, FH-13-14 does not appear to be a well folded protein. It is possible that changes in the conditions of the protein sample, such as pH and salt concentration could have an impact to its properties but this investigation was not pursued further mainly due to the focus of this work on the remaining aims.

However, it can be said that this construct is most likely not suitable for NMR-based structure determination.

3.1.5 Construct FH-14-15

3.1.5.1 Mini-scale expression trials

This construct was produced for assessment of its suitability for NMR-based structural analysis. Five colonies were selected for mini-scale production trials (Fig. 3.26). Upon concentration and EndoH$_f$ treatment, a dominant band appears in SDS-PAGE at ~17 kDa under reducing conditions (which might, in fact, be two bands). This corresponds approximately to the right size for FH-14-15 (~15 kDa). However, under non-reducing conditions there is clearly a second band (Fig. 3.26) at below 17 kDa, indicating some degradation. As was observed also for FH-13-14 there appear to be a ladder of oligomers under non-reducing conditions and a proteolytic degradation product emerges when samples are reduced. Despite this unpromising mini-scale production trial, a shaker-flask production was undertaken with the prospect of recording a ^{1}H-NMR experiment in mind.

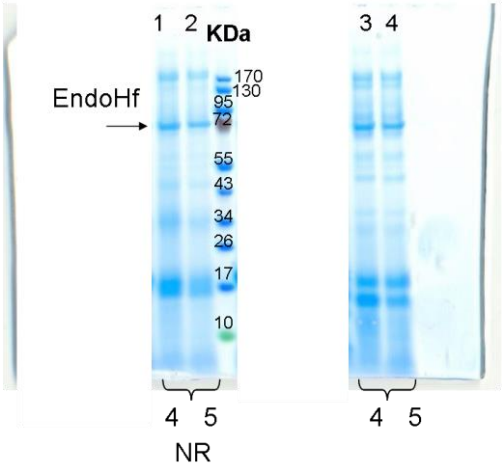


Figure 3.26 Gradient SDS-PAGE of the resultant spin-concentrated samples under reducing and non-reducing conditions. The lanes prior to lane 1, and the two lanes between the molecular weight markers and lane 3 have been masked with a white box, as these correspond to samples not relevant to this section.

3.1.5.2 Production trials of FH-14-15 in shaker flasks

The supernatant of the culture (~125 ml) was harvested, diluted, and subjected to cation-exchange (SP-Sepharose) chromatography at pH 4.0 (see section 2.2.6). Fractions from the elution of the column were subjected to $EndoH_f$ treatment and analyzed by SDS-PAGE (Fig. 3.27 A.). The gel showed bands where expected however it is evident that severe degradation has occurred as judged by the overloaded bands below 10 kDa. In addition, protein was still present in the flow-through, although in small quantities. Hence, the flow-through was diluted four-fold with sodium acetate buffer, pH 4.0 and re-applied to the same column (Fig. 3.27 B).

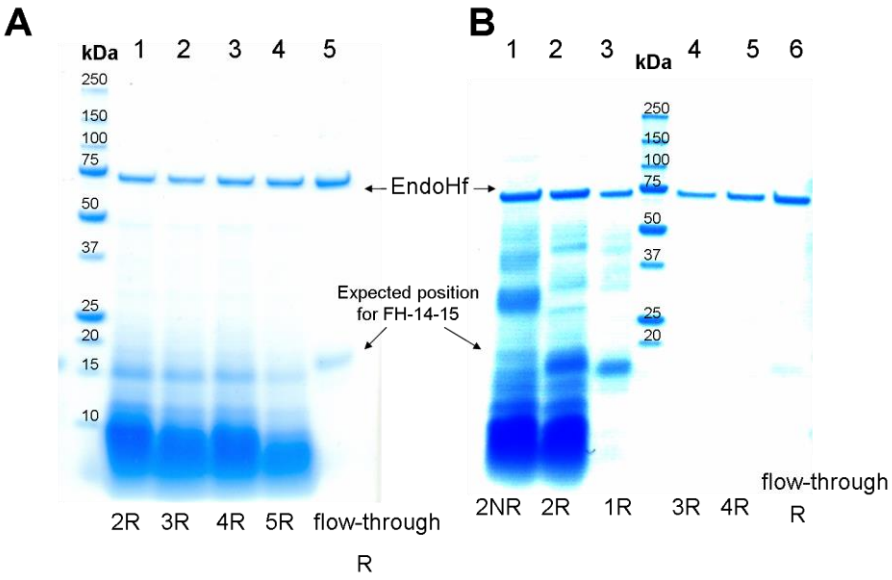


Figure 3.27 Gradient SDS-PAGE of the fractions from the elution of the SP-Sepharose column. (A) Lanes 2-4 correspond to fractions 2-5 that were eluted from the column. Lane 5 corresponds to sample from the flow-through (B) Lanes 1-3 and 4-6 correspond to fractions from the elution of the column after the flow-through was re-applied.

SDS-PAGE of the fractions from the SP-Sepharose elution revealed a rather weak band at the expected size for FH-14-15 (15 kDa) however the gel is dominated by very strong (overloaded) bands running ahead of the 10-kDa molecular weight marker (lanes 2-5). Fractions 2-5 from (A) and 1-3 from (B) (in Fig. 3.27) were combined, concentrated and applied to the size-exclusion chromatography column (see section 2.2.6) (Fig. 3.28). Resultant fractions were analyzed by SDS-PAGE (figure not shown) but no bands at an appropriate size were detectable.

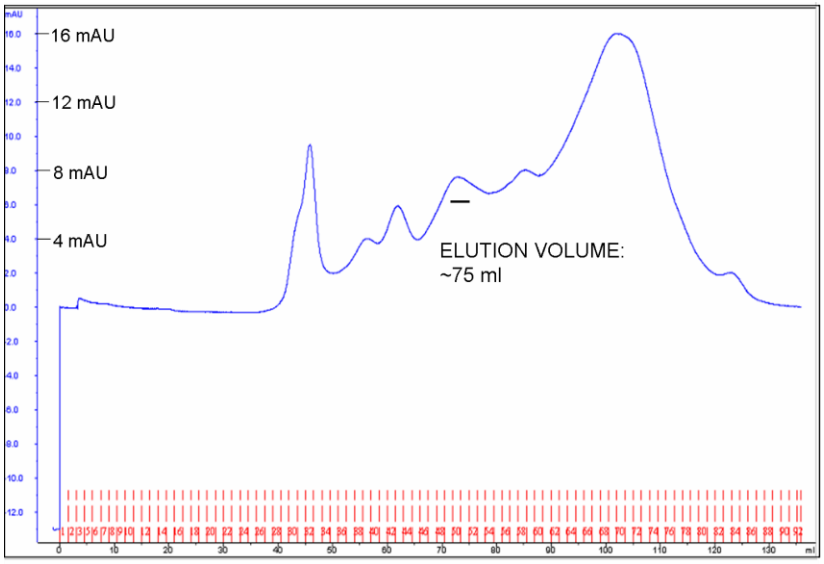


Figure 3.28 Size-exclusion chromatography of FH-14-15, using a HiLoad 16/60 Superdex 75 column.

The size-exclusion chromatogram contained multiple peaks indicative of a wide range of oligomers and breakdown products. A minor peak occurred at 75 ml, corresponding to the expected elution volume of a monomeric, double-CCP module. Analysis of fractions by SDS-PAGE, however, revealed that neither this nor any of the other peaks corresponded to an intact polypeptide chain of the expected size. It was concluded that recombinant *P. pastoris*-derived FH-14-15 was susceptible to misfolding (which could either be a cause of, or an outcome of, inappropriate disulfide bond formation), to proteolysis and/or to aggregation. This was not investigated further after considering the time-limitations and main aims of this project. It is interesting, however, to note that FH-15 folds compactly, with no degradation issues, in the FH-15-16 context (44).

3.1.6 Construct FH-10-15

Both unlabeled and ^{15}N-enriched batches of FH-10-15 were required as they were intended for binding studies with the PspC constructs, and for recording ^{1}H, ^{15}N-HSQC and TROSY NMR experiments, respectively.

3.1.6.1 Purification of unlabeled FH-10-15

Crude EndoH_f-treated and concentrated cell-depleted (by centrifugation and filtering) supernatant was retrieved from storage at -80 ^{o}C, and samples of 0.5-1.0 ml were loaded onto the size-exclusion chromatography column (as described in section 2.2.6). Fractions corresponding to peaks that eluted where expected were analyzed by SDS-PAGE, then subjected to a second size-exclusion chromatography step (on the same column). Data can be found in the supplementary CD section 1.5 while figure 3.29 shows the final sample produced as analyzed by SDS-PAGE.

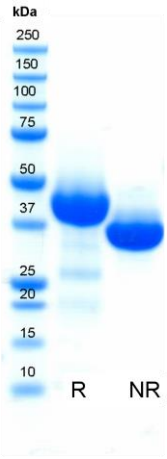


Figure 3.29 Gradient SDS-PAGE of the final sample for unlabeled FH-10-15. Under these reducing conditions, some minor degradation is visible.

The final sample, 140 μM in 1 ml, was used for subsequent SPR studies. Mass spectrometry analysis was not performed in the present work; however it has been investigated previously (148).

3.1.6.2 ^{15}N-enriched expression in a fermentor

P. pastoris glycerol stocks containing the DNA for FH-10-15 were re-streaked on a YPD plate containing 100 μg/ml zeocin. Single colonies were used for the starter culture in BMG. This was subsequently used to inoculate the fermentor. The fermentation procedure was carried out as described in Materials and Methods. The culture (~900 ml) was harvested and was then diluted ten-fold. FH-10-15 is known (on the basis of prior experience in the lab) to precipitate at low pH, therefore cation-exchange chromatography was ruled out as a first capture step. Anion-exchange chromatography was used, instead, on a Q-Sepharose column equilibrated at 20 mM Tris-HCl, pH 8.5. Due to incomplete solubilisation (as judged by the blockage of the column) it was subsequently decided to split the tenfold-diluted supernatant (10 L) into five 2-L batches, further dilute these threefold and apply one batch at a time to the column with the aim to elute one of each of the five 6-L batches daily. The elution was achieved using 20 mM Tris-HCl over a linear gradient from 0 to 1 M NaCl. Each batch-elution was subsequently analyzed by SDS-PAGE then fractions containing FH-10-15 were pooled, EndoH_f-treated and prepared for two consecutive rounds of gel-filtration chromatography. All fractions were analyzed by SDS-PAGE. The results for one representative purification are shown (batch 1)[$] while data for batches 2-5 may be found in the supplementary CD sections 1.6-1.9.

[$]This work was performed in conjunction with Vanessa Vongvarotai, an undergraduate student working under the supervision of the author.

3.1.6.2.1 Representative results for batch 1

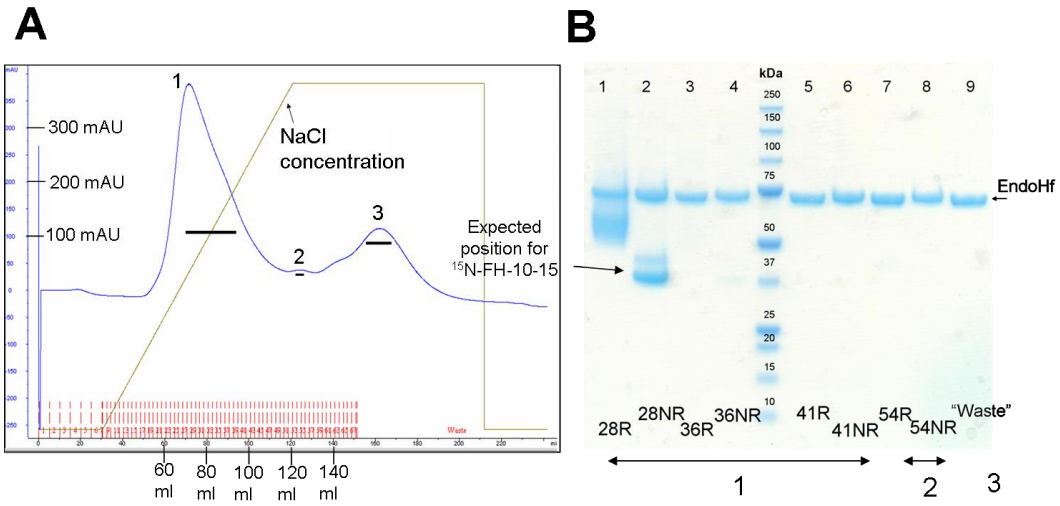


Figure 3.30 Anion-exchange chromatography of batch 1 of ^{15}N-labeled FH-10-15. (A) Chromatogram of the elution of the Q-Sepharose column. (B) Gradient SDS-PAGE of fractions from the elution as indicated, as well as from the "waste".

The chromatogram contains one main peak (of three) that, according to SDS-PAGE of the relevant fractions (28-36), contains protein of the expected size (42 kDa) (lanes 1-4 in figure 3.30). Fractions 24-34 were combined, EndoH$_f$- treated, concentrated, buffer exchanged into 20 mM potassium phosphate buffer, pH 7.4, containing 0.5 M NaCl and split into two samples. Each was subjected separately to size-exclusion chromatography, (Fig. 3.31).

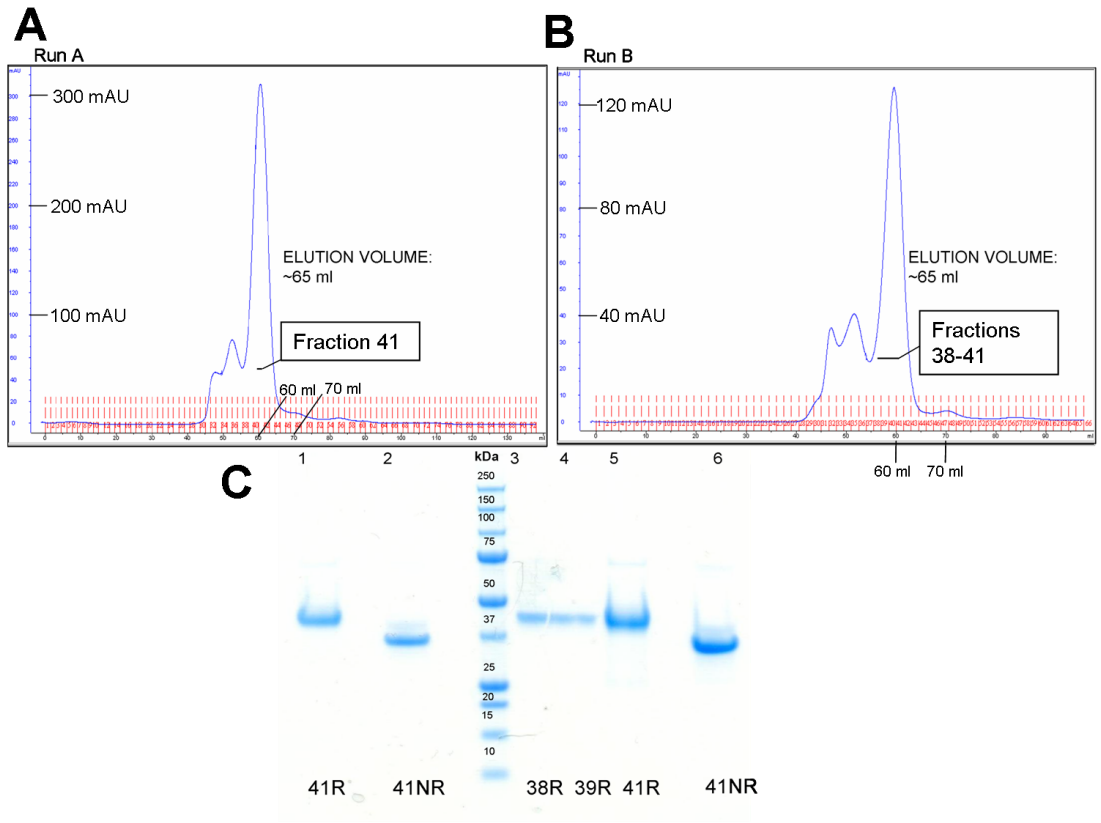


Figure 3.31 Size-exclusion chromatography for batch 1 of ^{15}N-FH-10-15, using a HiLoad 16/60 Superdex 75 column. (A), (B) Chromatograms (C) Gradient SDS-PAGE of the fractions as indicated below the gel from the two gel-filtration runs (from the peaks indicated with a black bar). Lanes 1-2: fractions from the middle of the main peak shown in (A). Lanes 3-6: fractions from the main peak shown in (B).

Size-exclusion chromatography revealed that near-homogeneous samples of ^{15}N-FH-10-15 had been prepared (as judged from the SDS-PAGE gel). Fractions from the main peak of both runs were combined and subjected to a second round of size-exclusion chromatography (Fig. 3.32).

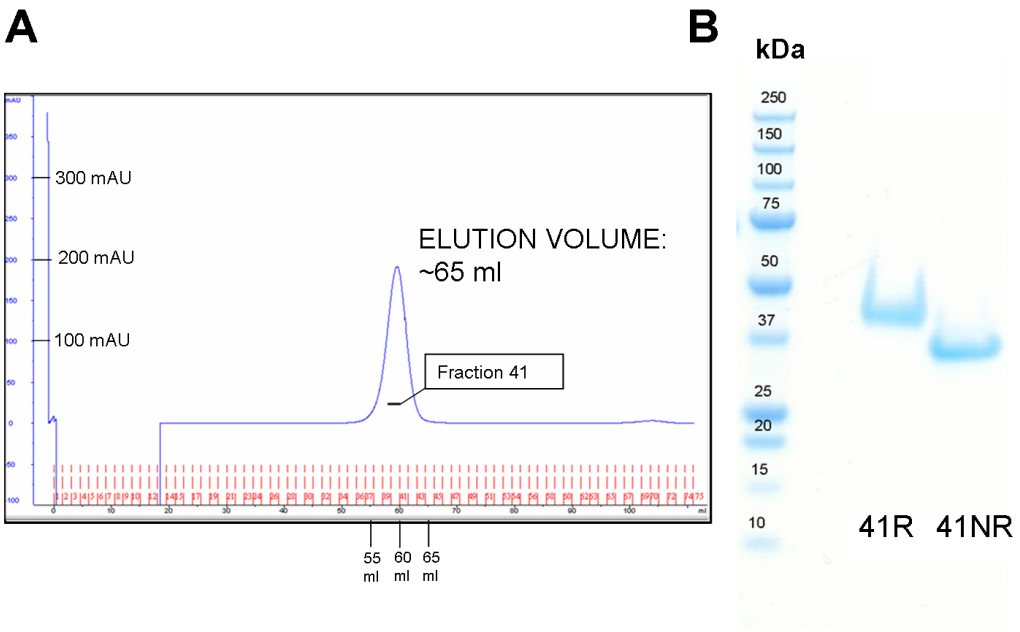


Figure 3.32 Final purification for batch 1 of ^{15}N-FH-10-15, using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram. (B) Gradient SDS-PAGE of fraction 41.

The size-exclusion chromatography profile featured just one peak that eluted at the expected volume for ^{15}N-FH-10-15 whilst the SDS-PAGE gel of the fraction from the middle of the peak confirmed its purity.

The NMR sample used to record 1D and 2D experiments was ~100 μM in 0.5 ml and resulted from the combination of some of the purest fractions shown in this section along with fractions for the remaining batches shown in the supplementary CD sections 1.6-1.8. An overloaded sample of that was run on a SDS-PAGE gel to ensure optimal purity and is shown in figure 3.33.

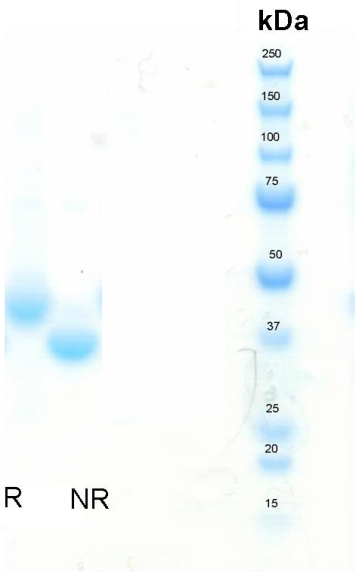


Figure 3.33 Gradient SDS-PAGE of the final sample used for the NMR experiments for ^{1}H, ^{15}N-FH-10-15. Strong bands at the expected positions appear under reducing and non reducing conditions consistent with purified protein.

In addition to the NMR sample, combination of some of the purest fractions resulted in the following 0.5 ml samples that were subsequently stored for future usage: a 15 μM sample, a 30 μM sample and a 20 μM sample.

3.1.6.3 NMR experiments on ^{15}N-labeled FH-10-15

A purified NMR sample (550 μl) of ~100 μM ^{15}N-labeled FH-10-15, which ran as a single band on SDS-PAGE, was prepared in 20 mM potassium phosphate pH 6.8. Figures 3.34-3.36 show the 1-D and 2-D NMR experiments performed for ^{15}N-FH-10-15.

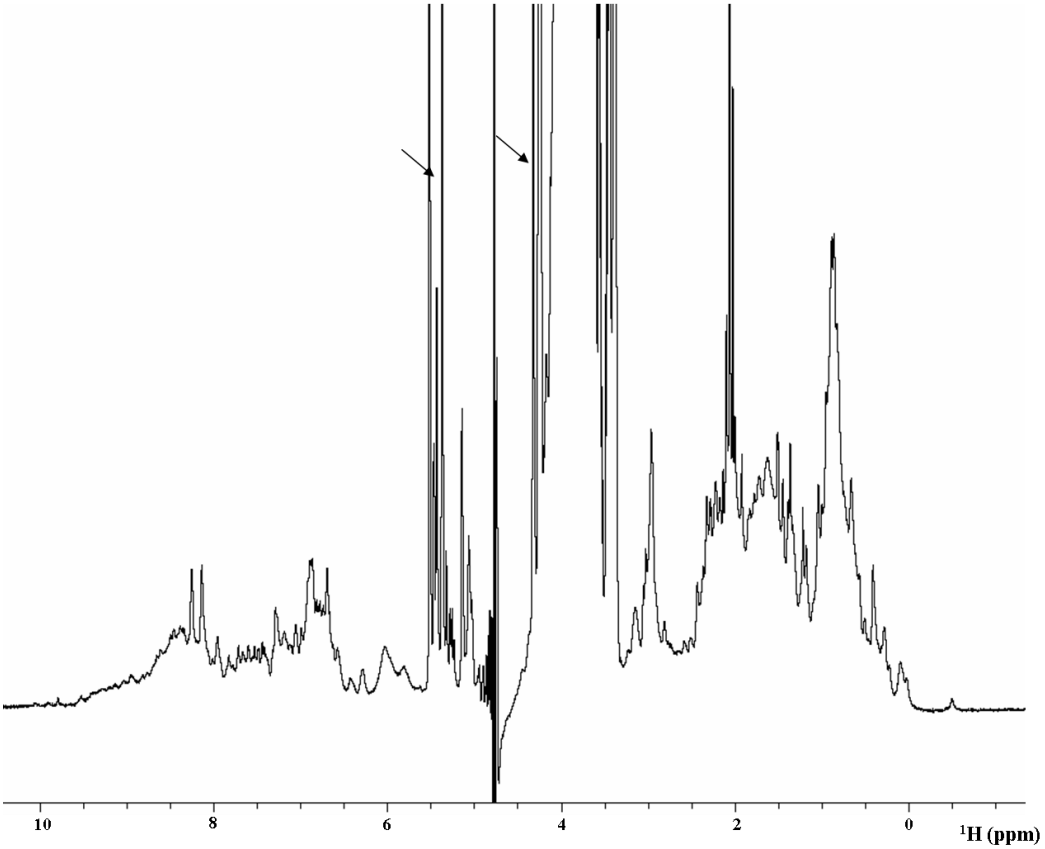


Figure 3.34 1D ^{1}H-NMR spectrum of ^{15}N-FH-10-15. Arrows indicate peaks thought to derive from contaminating carbohydrates (147).

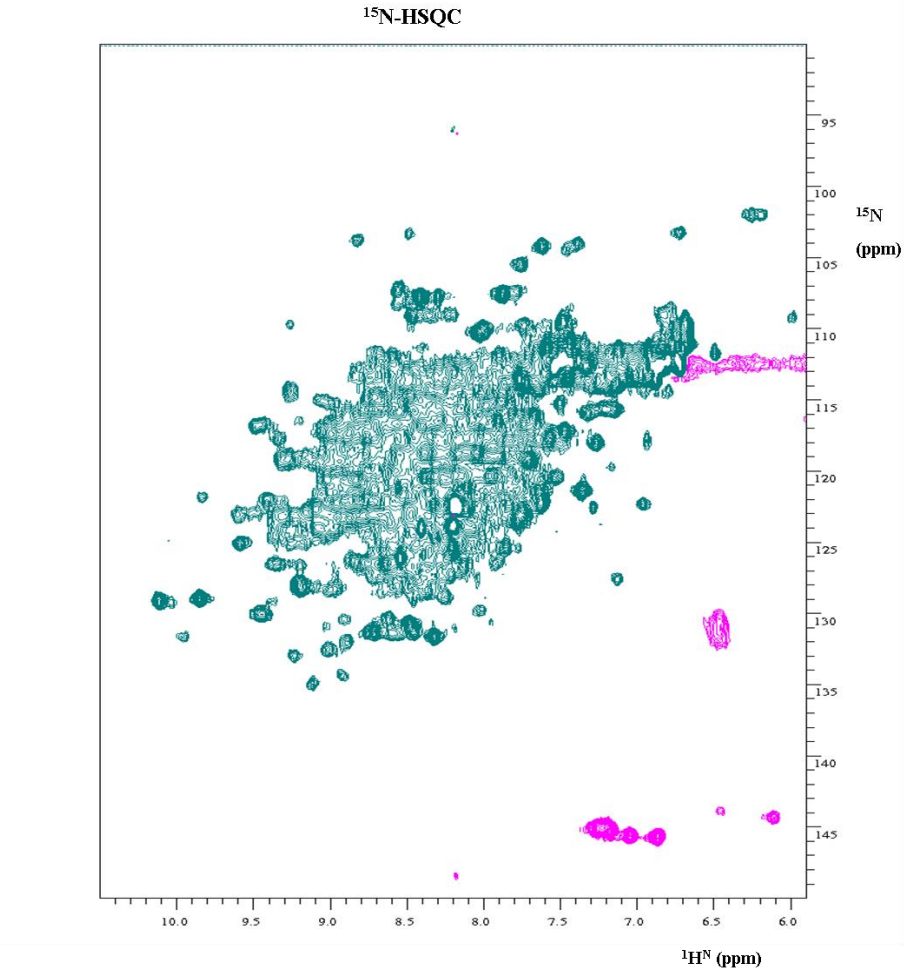


Figure 3.35 ^{1}H, ^{15}N-HSQC spectrum for ^{15}N-FH-10-15. Peaks in magenta are negative (folded).

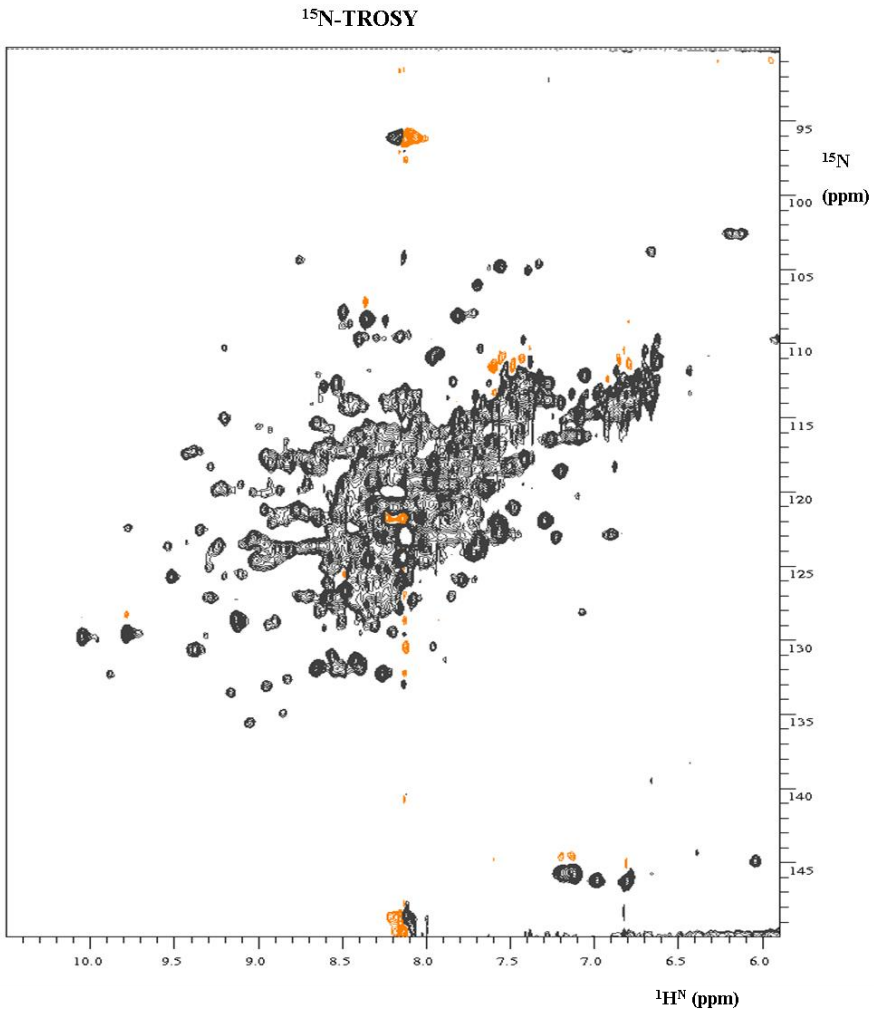


Figure 3.36 ^{1}H, ^{15}N-TROSY spectrum for ^{15}N-FH-10-15. Peaks in orange are negative (folded).

The ^{1}H-NMR spectrum shows a relatively good dispersion with peaks that are well resolved considering the relatively high molecular weight of this protein, along with some very broad peaks that could reflect self-association or multiple conformations. In addition, there appears to be contamination with carbohydrates. This had also been observed in the case of FH-13-14 and could potentially be due to co-purification with *P. pastoris*-derived oligosaccharides (147). The ^{1}H, ^{15}N-HSQC spectrum showed significant overlap as well as some well-dispersed and resolved cross peaks. Some of the overlap might be ascribed to a combination of line-broadening and the presence of many cross peaks both of which are a consequence of the relatively large size of the protein. On the other hand there are a wide range of peakwidths and a large number of cross-peaks at random coil shifts (around 8.3 ppm) that could reflect conformational mobility on the intermediate timescales.

In general, line widths depend on the tumbling of a molecule in solution and hence on its size. A protein that exhibits a ^{1}H, ^{15}N-HSQC spectrum with a wide range of cross-peak widths could imply self-association and formation of a large complex with some flexible loops giving rise to sharp peaks. Furthermore, in a ^{1}H, ^{15}N-HSQC spectrum of an unfolded protein signals appear as clusters of not well-defined peaks around 8.3 ppm (random-coil shifts) with little dispersion in either dimension.

Many signals appear to be sharper in the ^{1}H, ^{15}N-TROSY spectrum as expected since this pulse sequence promotes mutual cancellation of T_2 relaxation pathways. Nonetheless there remains a cluster of cross peaks near the centre of the spectrum consistent with the presence of disorder and line broadening due to interchanging (intermediate timescale) conformations. This sample (stored for future use at -80 °C) could be applied to a ConA column to remove the carbohydrate contaminants and further work could then be done to optimise conditions such as pH, temperature and salt concentration. In the future, FH-10-15 could be expressed under ^{15}N, ^{2}D-labelling conditions to improve the quality of the TROSY spectrum and possibly allow some assignments based on the assignments of the component modules.

3.1.7 Construct FH-8-15

This construct was prepared for binding studies with PspCN/PspCNR1-proteins from *S. pneumoniae*. Cell-depleted growth media was retrieved from storage at -80 ^{o}C and deglycosylated with $EndoH_f$. In a trial, six samples of 0.5-1 ml each were flowed, separately, down a size-exclusion chromatography column, then appropriate fractions were pooled and re-loaded (in two batches) onto the same column for a second size-exclusion chromatography step. Data for this purification can be found in the supplementary CD section 1.10. The final resultant sample (~50 μM sample in 0.5 ml) was run on SDS-PAGE to confirm its purity (Fig. 3.37).

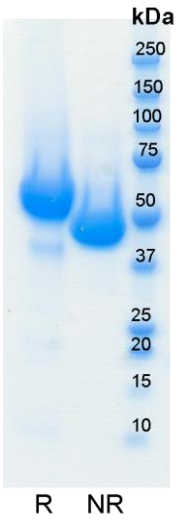


Figure 3.37 Gradient SDS-PAGE of the final sample for FH-8-15. Strong bands appear at the expected molecular weight; under reducing conditions some mild degradation is visible while under non-reducing conditions a single band appears.

The final sample for FH-8-15 was homogeneous as judged by SDS-PAGE and was used for the SPR experiments.

3.2 Production, purification and characterization of the PspC constructs

The first PspCN construct to be produced was His-tag PspCN-GSGC. The expression construct for PspCN with a non-native C-terminus GSGC sequence was designed for its potential future use as a platform for labeling or thiol chemistry and was already available in the lab. The protein was used in initial SPR studies of its interaction with FH. The second segment produced was His-tagged PspCN. In this case the non-native C-terminal amino acid addition was absent. This protein was used for follow-up SPR analysis of PspCN-FH interactions. Finally, the constructs SUMO-PspCN and SUMO-PspCNR1 were produced, utilizing the pE-SUMOPro-Kan vector. This final approach was adopted because of the increased yields and solubility that can be achieved. In addition, cleavage of SUMO (that is fused to the protein of interest) results in the native form of the proteins (no N- or C- terminal tags and/or non-native residues). These constructs were subjected to the final and most extensive SPR experiments with FH and C3.

The table below (Table 3.4) summarizes the PspC constructs produced in this study, their position in the native PspC protein and the non-native C- and N- terminal tags/sequences.

Construct	*Position in PspC*	*Non-native sequence (N-terminal)*	*Non-native sequence (C-terminal)*
His-tag PspCN-GSGC	Residues 37-140	MGSSHHHHHHSSGLVPRGSHM	GSGC
His-tag PspCN	Residues 37-140	MGSSHHHHHHSSGLVPRGSHM	-
SUMO-PspCN	Residues 37-140	MGHHHHHHGSLQDSEVNQEAK PEVKPEVKPETHINLKVSDGSSE IFFKIKKTTPLRRLMEAFAKRQG KEMDSLRFLYDGIRIQADQAPED LDMEDNDIIEAHREQIGG	-
PspCN	Residues 37-140	-	-
SUMO-PspCNR1	Residues 37-292	MGHHHHHHGSLQDSEVNQEAK PEVKPEVKPETHINLKVSDGSSE IFFKIKKTTPLRRLMEAFAKRQG KEMDSLRFLYDGIRIQADQAPED LDMEDNDIIEAHREQIGG	-
PspCNR1	Residues 37-292	-	-

Table 3.4 Summary of the PspC constructs used in this study.

The cloning, expression, purification and characterization performed for each of the PspC constructs employed in this study is summarized in Tables 3.5 (parts 1 and 2) while table 3.6 shows the expected and experimentally determined molecular weights.

PspC construct	His-tag PspCN-GSGC 1st batch	His-tag PspCN-GSGC 2nd batch	His-tag PspCN 1st batch	His-tag PspCN 2nd batch
Expression system	*E.coli*	*E. coli*	*E. coli*	*E. coli*
Starting material	PspCN DNA in pET-15b vector	PspCN DNA in pET-15b vector	PspCN DNA in pET-15b vector	PspCN DNA in pET-15b vector
Purification step 1	Cobalt affinity	Cobalt affinity	Cobalt affinity	Cobalt affinity
Purification step 2	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion
Purification step 3	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion
Validation of primary sequence		MS-Tryptic digest		Accurate mass determination by MS
Assessment of folding	-	-	-	-
Comments	-	Used for SPR	-	-

Table 3.5 Overview of the PspC constructs employed in this study (Part 1).

PspC construct	His-tag PspCN 3rd batch	SUMO-PspCN	PspCN	SUMO-PspCNR1	PspCNR1
Expression system	*E. coli*	*E. coli*	*E .coli*	*E. coli*	*E. coli*
Starting material	PspCN DNA in pET-15b vector	PspCN DNA in pMA-T vector (Geneart)	Cleaved from SUMO-PspCN	PspCNR1 DNA in pMA-T vector (Geneart)	Cleaved from SUMO-PspCNR1
Purification step 1	Cobalt affinity	Cobalt affinity	Re-pass, Cobalt affinity	Cobalt affinity	Re-pass, Cobalt affinity
Purification step 2	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion	Size-exclusion
Validation of primary sequence	Accurate mass determination by MS	Accurate mass determination by MS	Accurate mass determination by MS	Accurate mass determination by MS	Accurate mass determination by MS
Assessment of folding	-	-	^{1}H, ^{15}N-HSQC	-	^{1}H, ^{15}N-HSQC
Comments	Used for SPR	Used for SPR	Used for SPR	Used for SPR	Used for SPR

Table 3.5 Overview of the PspC constructs employed in this study (Part 2).

PspC construct	Expected molecular weight (kDa)	Measured molecular weight (kDa)
His-tag PspCN-GSGC	14,356	–
His-tag PspCN	14,052	14,050
PspCN	11,887	11,886
SUMO-PspCN	24,147	24,146
PspCNR1	28,913	28,913
SUMO-PspCNR1	41,173	41,174
SUMO	12,279	12,279

Table 3.6 Summary of the expected and experimentally determined molecular weights for all the PspC constructs prepared in this study.

3.2.1 Construct His-tag PspCN-GSGC

Figure 3.38 shows the position of the His-tag PspCN-GSGC construct within the native PspC protein as well as the amino acid sequence including the non-native N- terminal tag and the C- terminal GSGC amino acid addition.

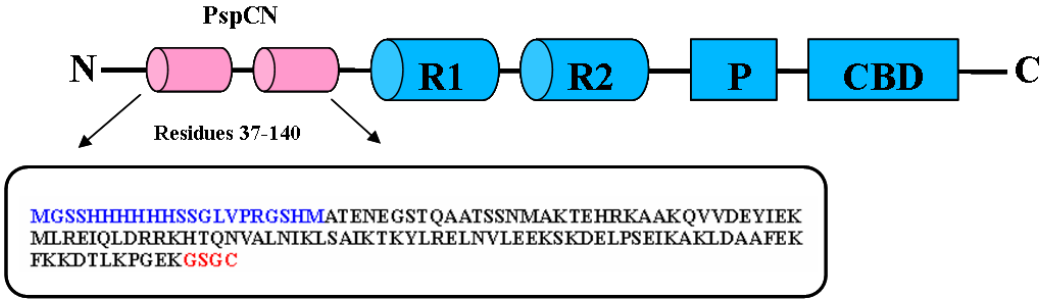


Figure 3.38 Schematic representation of the PspC protein. Residues 37-140 correspond to the N-terminal PspCN construct. The native sequence for this construct is shown in black, the N-terminal non-native tag in blue and the C-terminal non-native GSGC amino acid sequence in red.

In an initial purification step, Co^{2+}-affinity chromatography was employed. The buffer used was 20 mM potassium phosphate pH 7.0, 500 mM NaCl, 20 mM imidazole (buffer A) while the His-tagged protein was eluted using the same buffer but with 500 mM imidazole instead of 20 mM (buffer B). Two batches of His-tag PspCN-GSGC were produced and the data are presented for just one of these (data for the remaining batch can be found in the supplementary CD section 1.11). In both cases, the soluble fraction obtained after cell lysis (see Materials and Methods section 2.1.16.3 for details) was diluted with an equal volume of buffer A before application to the pre-equilibrated IMAC column. Protein was then eluted over a linear gradient from buffer A (20 mM imidazole) to buffer B (500 mM imidazole) [$].

3.2.1.1 Representative results for batch 1

Figure 3.39, shows the results of SDS-PAGE performed to assess the presence of His-tag PspCN-GSGC before and after induction with IPTG.

[$]This work was performed in conjunction with Vanessa Vongvarotai, an undergraduate student working under the supervision of the author.

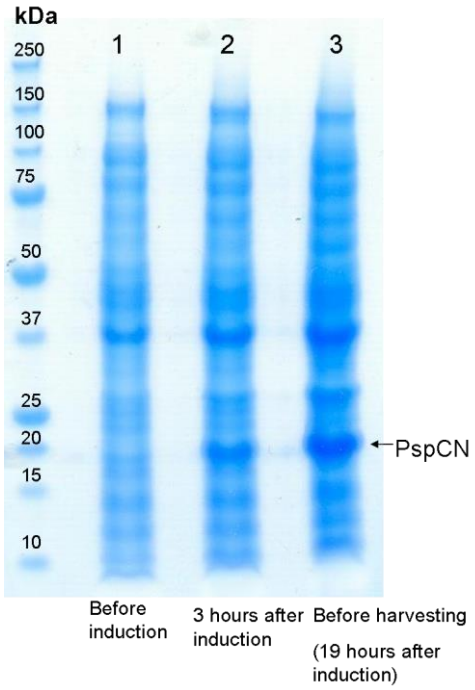


Figure 3.39 Gradient SDS-PAGE of samples from a culture of *E. coli* cells engineered to overproduce His-tag PspCN-GSGC, taken before and after induction with IPTG.

According to SDS-PAGE gel, a strong band appears just below the 20 kDa size marker after 3 h of induction which becomes more pronounced 19 h after induction. The equivalent band is not observed in the absence of IPTG. PspCN has a molecular weight of only ~15 kDa (including the His-tag, thrombin site and cloning artifacts), however it is strongly positively charged (pI = 9.46) consistent with slower migration on SDS-PAGE than the standards and an apparently higher molecular weight.

Eight individual colonies were selected and each grown in 500 ml LB media, as described in Materials and Methods, section 2.1.16. Following harvesting, cells were weighed and frozen. Four out of the eight were pooled (as an initial batch, while the rest were kept for subsequent purification), giving a total wet cell weight of 7.4 g. Cells were lysed and the lysate was applied on the IMAC column as described in Materials and Methods section 2.2.6, (Fig. 3.40 A). Fractions from this step were analyzed by SDS-PAGE (Figs 3.40 B and C).

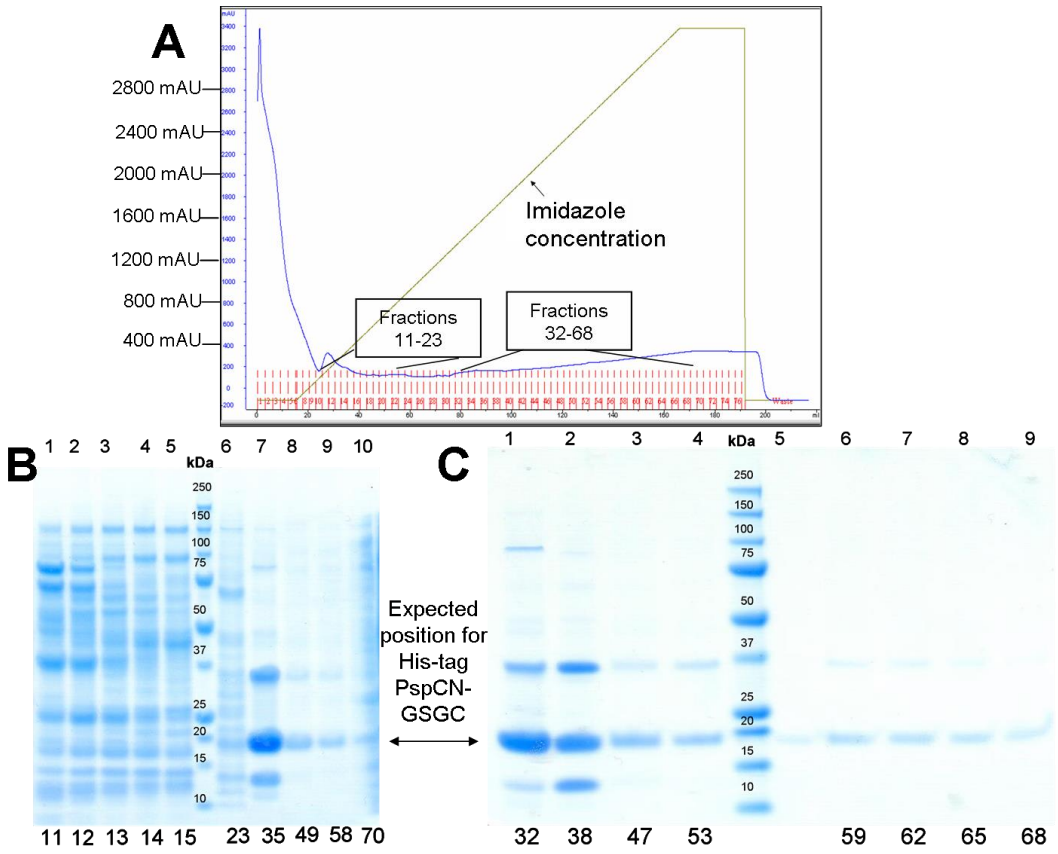


Figure 3.40 First purification step for His-tag PspCN-GSGC; metal-affinity chromatography using a HiLoad IMAC column. (A) Chromatogram. (B) and (C) Gradient SDS-PAGE of the fractions of the elution as indicated.

No distinct peaks were eluted from the IMAC column, rather a continuous increasing trace in the A_{280} nm UV absorbance was obtained due to the imidazole gradient. Despite the lack of a peak of absorbance, from SDS-PAGE it appears that fractions between 32 and 70 are potentially enriched with the protein of interest. Hence, fractions 32-40 and fractions 44-70 were combined separately. The pool corresponding to fractions 44-70 was concentrated and buffer-exchanged into PBS containing 5 mM EDTA in preparation for size-exclusion chromatography (Fig. 3.41 A). Fractions from this second purification step were analyzed by SDS-PAGE (Fig. 3.41 B).

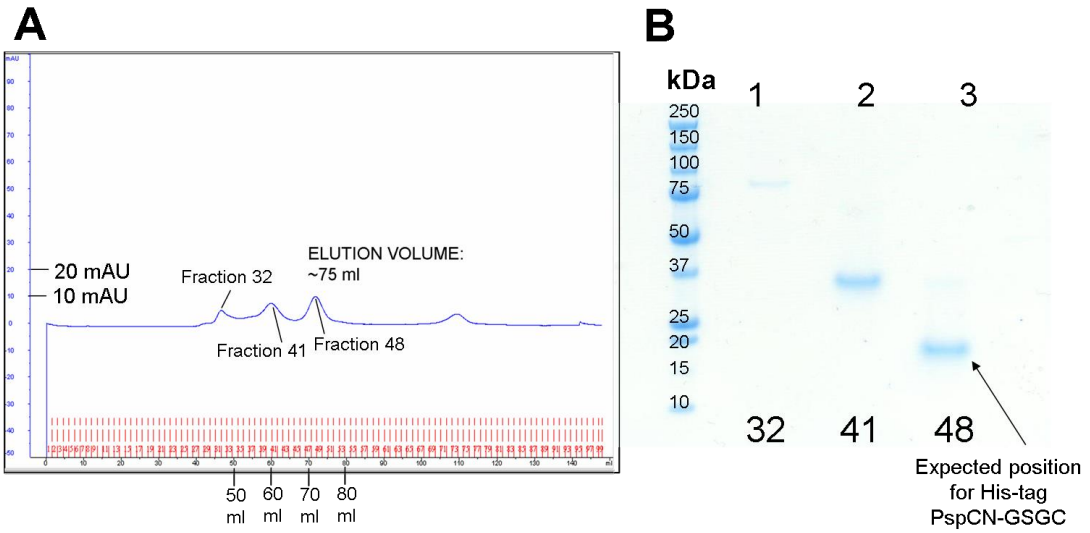


Figure 3.41 Size-exclusion chromatography for His-tag PspCN-GSGC, using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram of the size-exclusion chromatography run. (B) Gradient SDS-PAGE of the most representative fractions corresponding to the peaks visible in the chromatogram shown in (A).

The main feature of this size-exclusion chromatography run was the very low UV absorbance as seen in the chromatogram indicating that only a very small fraction of the original protein had survived the IMAC step. In the SDS-PAGE, faint band appears just above 75 kDa for fraction 32 (corresponding to the first peak that eluted from the column) which could be an *E. coli* contaminant. Fraction 41 shows a band around 37 kDa and could be potentially arising from the formation of a dimer (note the C-terminal Cys residue that could form a disulfide bond). Finally, fraction 48 exhibits a band at the expected size for monomeric His-tag PspCN-GSGC just below 20 kDa. In spite of the low UV absorbance of the eluted peaks, fractions were stored.

3.2.2 Construct His-tag PspCN

Figure 3.42 shows the amino acid sequence of the target protein.

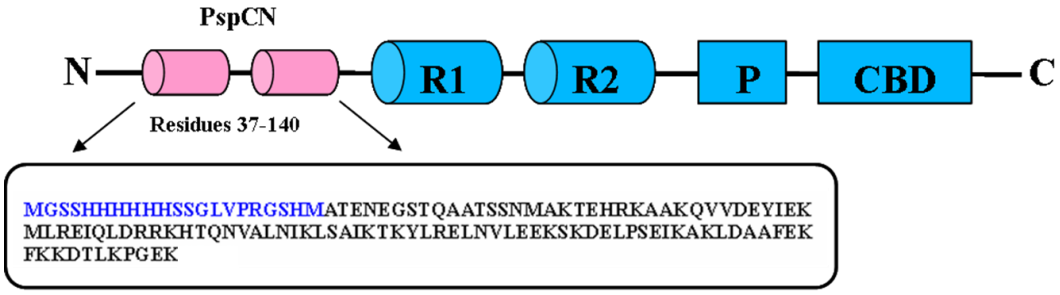


Figure 3.42 Representation of the amino acid sequence used in this part of the study. The non-native His-tag followed by the thrombin cleavage site and the cloning artifact are highlighted in blue one-letter code. The native PspC sequence of the N-terminus region of PspC, termed PspCN is shown in black.

Three batches of cells were cultured, (each corresponding to 3L LB media) harvested (wet cell weights were 8.7 g, 10 g and 6.5 g for the first, second and third batches) and lysed as described in Materials and Methods, section 2.1.16.3. Protein was eluted from the metal-affinity chromatography capture step using an imidazole gradient from 20 to 500 mM imidazole. Size-exclusion chromatography was used as a second purification step. Representative data for batch 2 will be presented below (results for batches 1 and 3 can be found in the Supplementary CD sections 1.12 and 1.13).

3.2.2.1 Representative results for batch 2

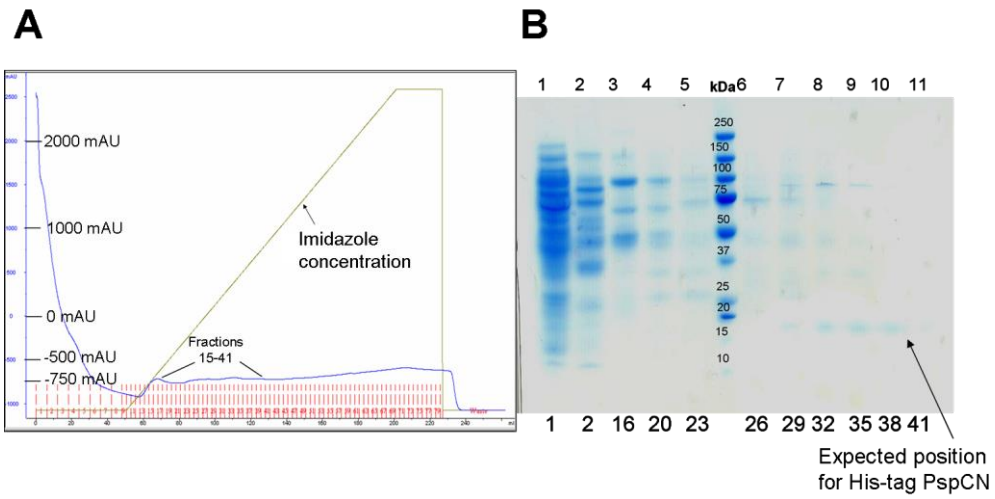


Figure 3.43 First purification/capture step for His-tag PspCN (Batch 2) using immobilized metal-affinity chromatography. (A) Chromatogram of the elution of the IMAC column. (B) Gradient SDS-PAGE of fractions from the elution as indicated below the gel. Fractions 44-80 were also run on a separate gel however no bands were visible (data not shown).

The lack of distinct peak/peaks following the launch of the gradient has been observed and commented on previously (section 3.2.1.1). Fractions 32-41 (lanes 8-11) showed bands in the SDS-PAGE gel (Fig. 3.49 B) at the expected size for His-tag PspCN from the elution of the IMAC column (migrating to just below 20 kDa) (Fig. 3.43 A). Fractions 44-80 which were included in a separate gel did not show any bands (data not shown). Nonetheless fractions 35-75 were pooled, concentrated and buffer exchanged in PBS containing 5 mM EDTA in preparation for size-exclusion chromatography. The resultant sample was split in two and two runs were performed for each of the two samples (Fig. 3.44 A and B). Fractions from the main peaks of those two runs were combined, further concentrated and subjected to a second round of size-exclusion chromatography (Fig. 3.44 C). Fractions from this final step were monitored by SDS-PAGE (Fig. 3.44 D).

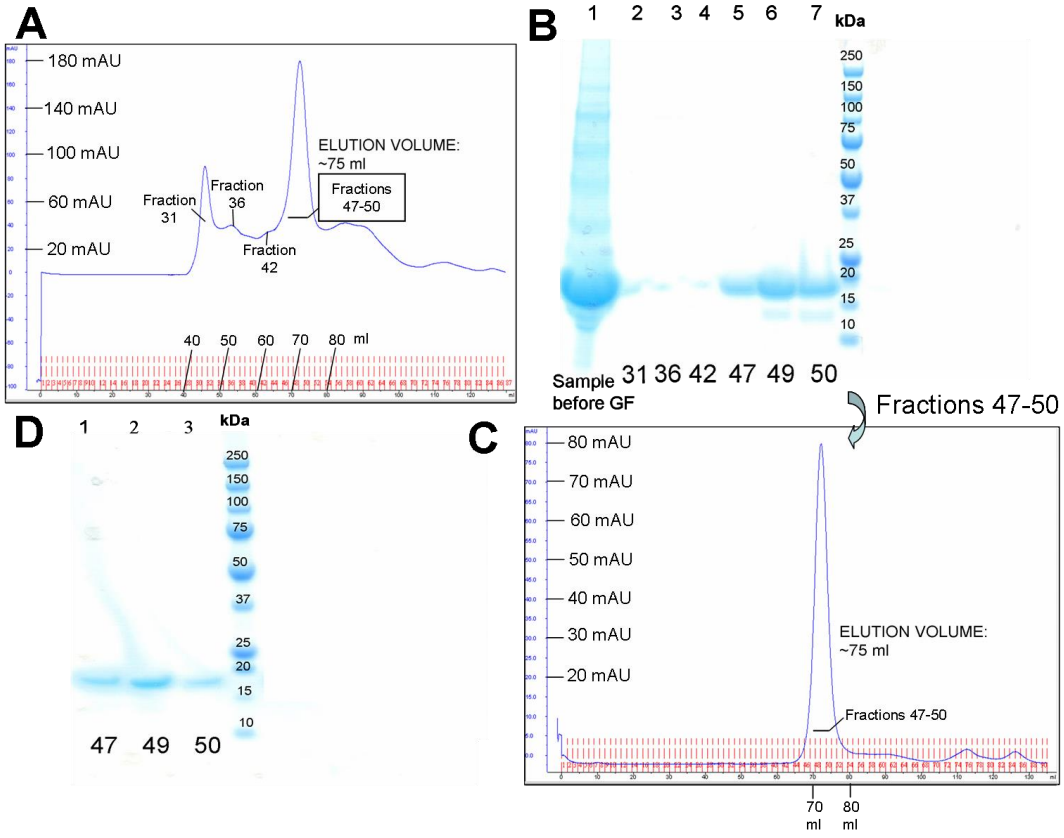


Figure 3.44 Second and final purification for His-tag PspCN, using a HiLoad 16/60 Superdex 75 column. (A), (B) Chromatograms of the two size-exclusion chromatography runs. The main peaks, indicated with a black bar, correspond to PspCN. (C) Chromatogram of the size-exclusion chromatography run performed on fractions 47-50 of the runs shown in (A) and (B). One main peak eluted which corresponds to PspCN. (D) Gradient SDS-PAGE of fractions from the main peak indicated with a black bar in (C).

Two consecutive rounds of size-exclusion chromatography resulted in protein eluting as a single peak. It is evident from the final SDS-PAGE gel performed that the majority of the protein was purified while some degradation might have occurred as reflected by the bands below the expected size for PspCN (~20 kDa). Fractions 47, 48, 49 and 50 were frozen and kept separately. From those, samples for fractions 49; 130 μM in 1 ml, and 50; 95 μM in 1 ml, were submitted for accurate mass analysis (Figs. 3.45, 3.46).

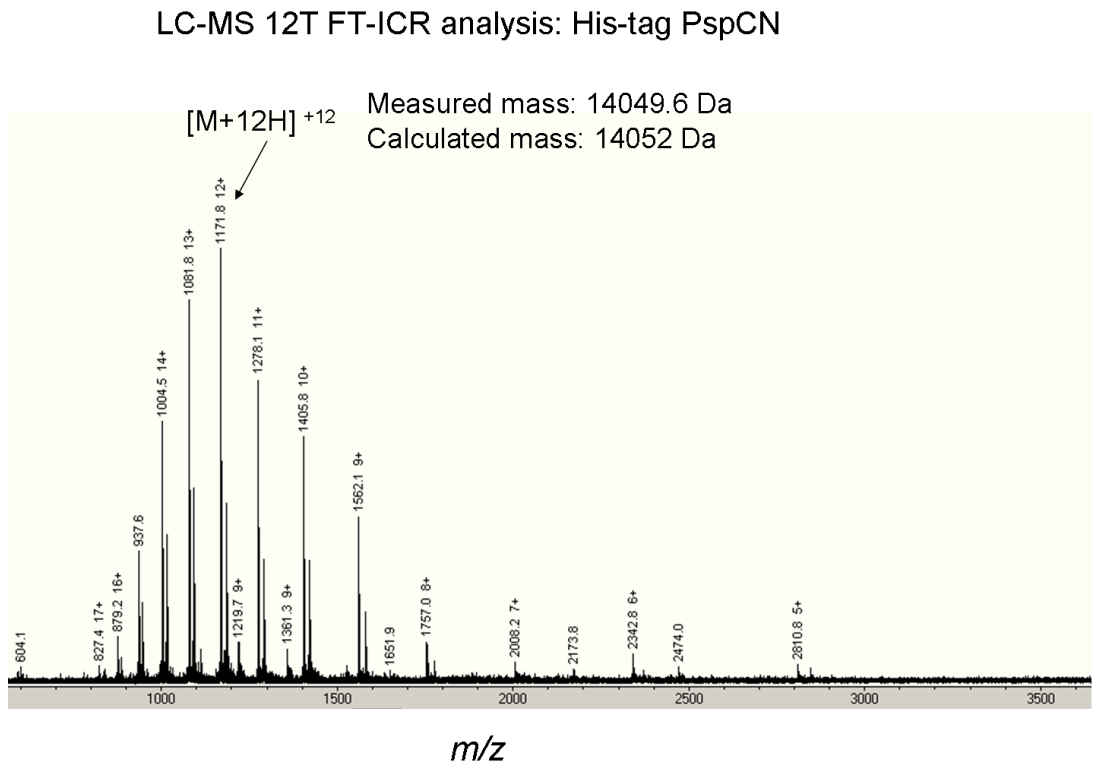


Figure 3.45 LC-MS spectrum for fraction 49 of purified His-tag PspCN. The highest peak (charged state 12+) on the spectrum corresponds to intact protein.

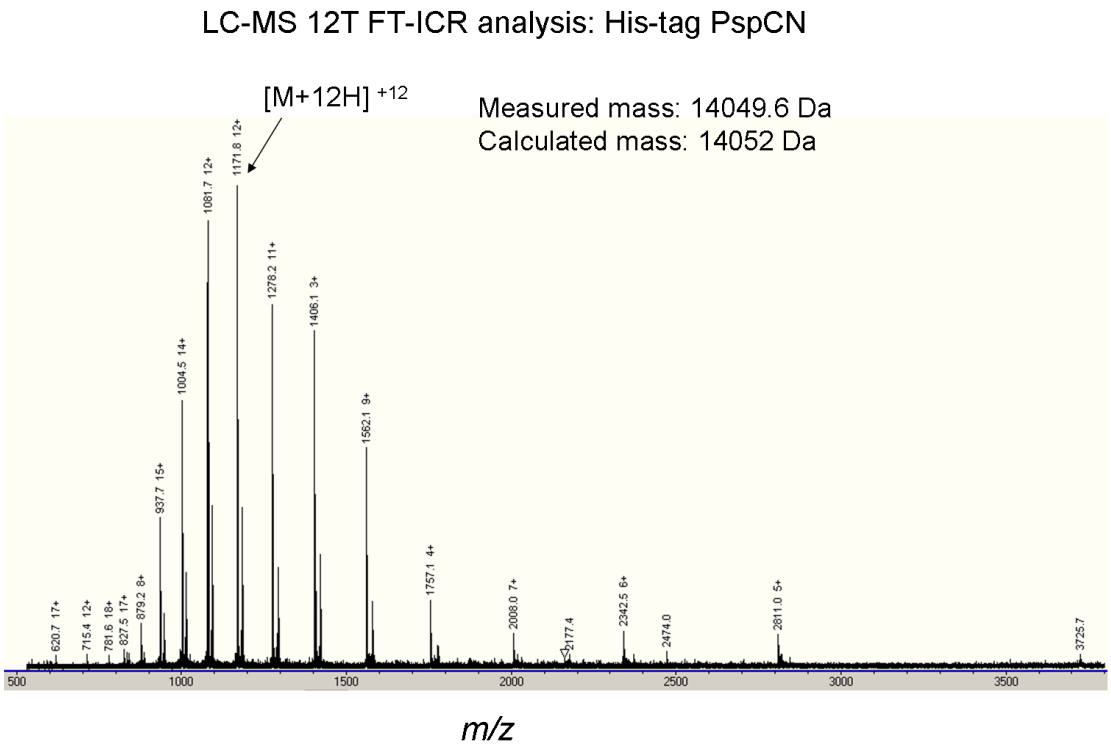


Figure 3.46 LC-MS spectrum for fraction 50 of purified His-tag PspCN. The measured mass was identical to what was found for fraction 49 shown in figure 3.98.

From the LC-MS analysis it can be concluded that although other species were present in the spectrum, the highest peaks for both fractions 49 and 50 correspond to intact protein.

3.3 Production of PspC constructs in pE-SUMOPro Kan vector

3.3.1 Construct SUMO-PspCN

Following transformation and DNA manipulation (Materials and Methods sections 2.1.11-2.1.13), purified plasmid DNA was used to transform two *E. coli* strains in order to decide which was better for expression: BL21 (DE3) and BL21 (DE3) - pLysS. Colonies were used to inoculate 5 ml of LB media containing 30 μg/ml kanamycin and both strains were tested at different temperatures and concentrations of IPTG (Fig. 3.48).

The figure below (Fig. 3.47) shows the amino acid sequence used consisting of the SUMO sequence shown in magenta and the native PspCN sequence shown in black.

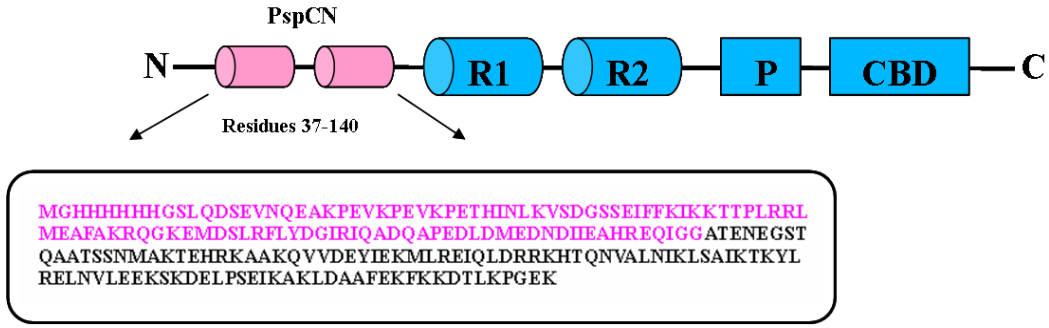


Figure 3.47 Representation of PspC in which the sequence corresponding to the N-terminus used in this study, denoted PspCN fused with SUMO is indicated.

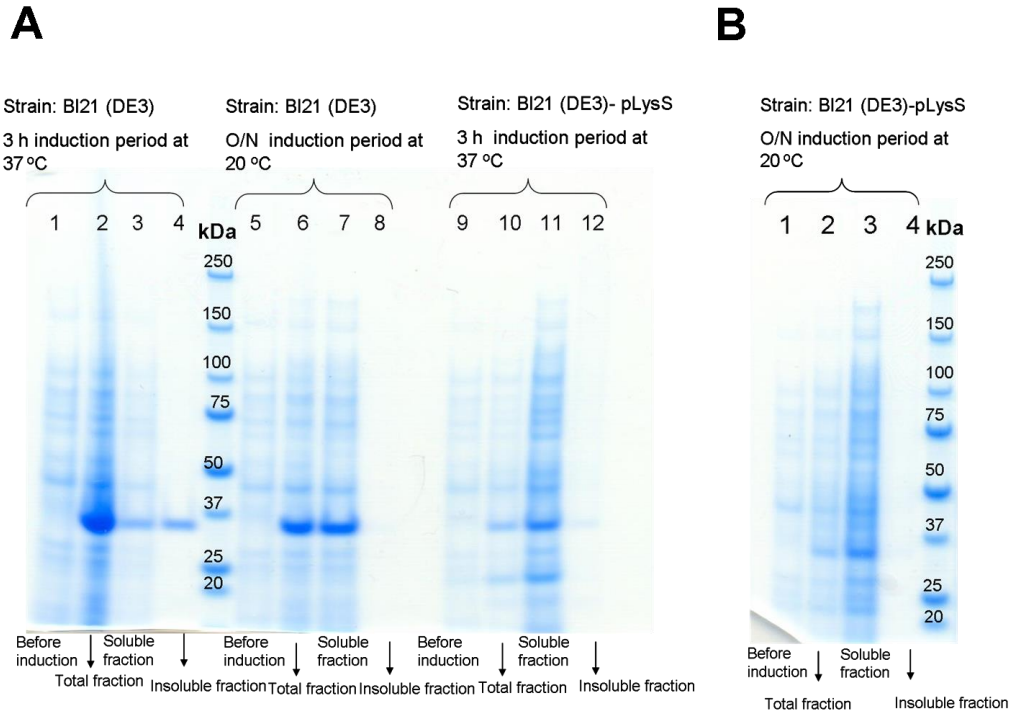


Figure 3.48 Gradient SDS-PAGE of the trial expressions for SUMO-PspCN. (A) Lanes 1-4 show the expression obtained by using the *E. coli* strain BL21 (DE3), as indicated. The band appearing just below 37 kDa probably corresponds to SUMO-PspCN and is visible in lanes 2-4. Lanes 5-8 show the expression as indicted on the gel. Bands for the expected protein appear just below 37 kDa in lanes 6-8. Lanes 9-12 show the expression for the strain BL21 (DE3)-pLysS, as indicated. SUMO-PspCN once again appears just below 37 kDa in lanes 10-12. (B) Lanes 1-4 show the expression for the strain BL21 (DE3)-pLysS. Bands for SUMO-PspCN are visible just below 37 kDa in lanes 2-4.

It is evident from SDS-PAGE (Fig. 3.48) that use of the *E. coli* strain BL21 (DE3) and an overnight induction period at 20 ^{o}C yielded both the highest amount of

protein (as reflected by the stronger bands), and the highest ratio of soluble-to-insoluble protein (note that a very faint band appeared in the actual gel in the lane for the insoluble fraction, however this is not visible in the figure). This strain and these conditions were used for a larger-scale growth.

The wet weight of cells following harvesting (of the 2-L LB cultures), after combining all four flasks was 7 g. The soluble lysate was applied onto the IMAC column and eluted as already described (the buffers were identical to those described in the sections 3.3.1 and 3.3.2) (Fig. 3.49 A and B). Fractions from the elution were analyzed by SDS-PAGE and are shown in figure 3.50 B & C.

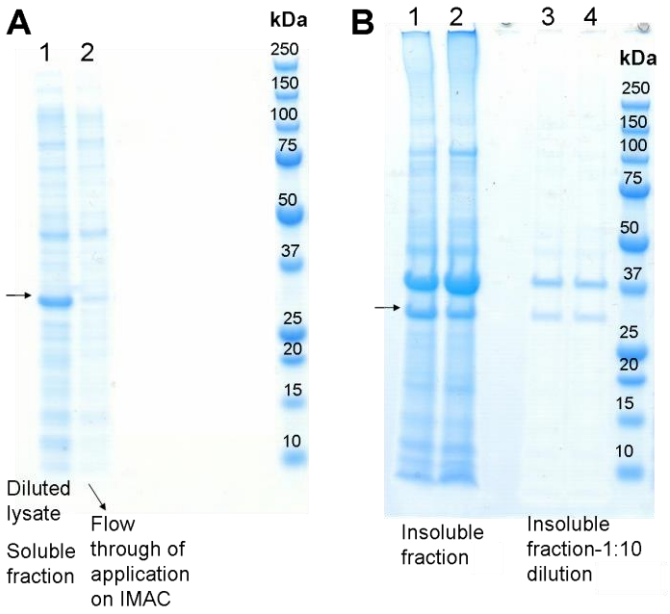


Figure 3.49 Gradient SDS-PAGE gel of SUMO-PspCN, prelude to purification attempt on a HiTrap chromatography column: (A) Lanes 1 and 2: samples as indicated. (B) Lanes 1-4: samples as indicated. All lanes contain bands at the expected molecular weight for SUMO-PspCN (indicated with an arrow). The lanes between lane 2 and the molecular weight markers in (A) and between the molecular weight markers and lane 1 in (B) have been masked with white boxes, as these contained samples not relevant to this section.

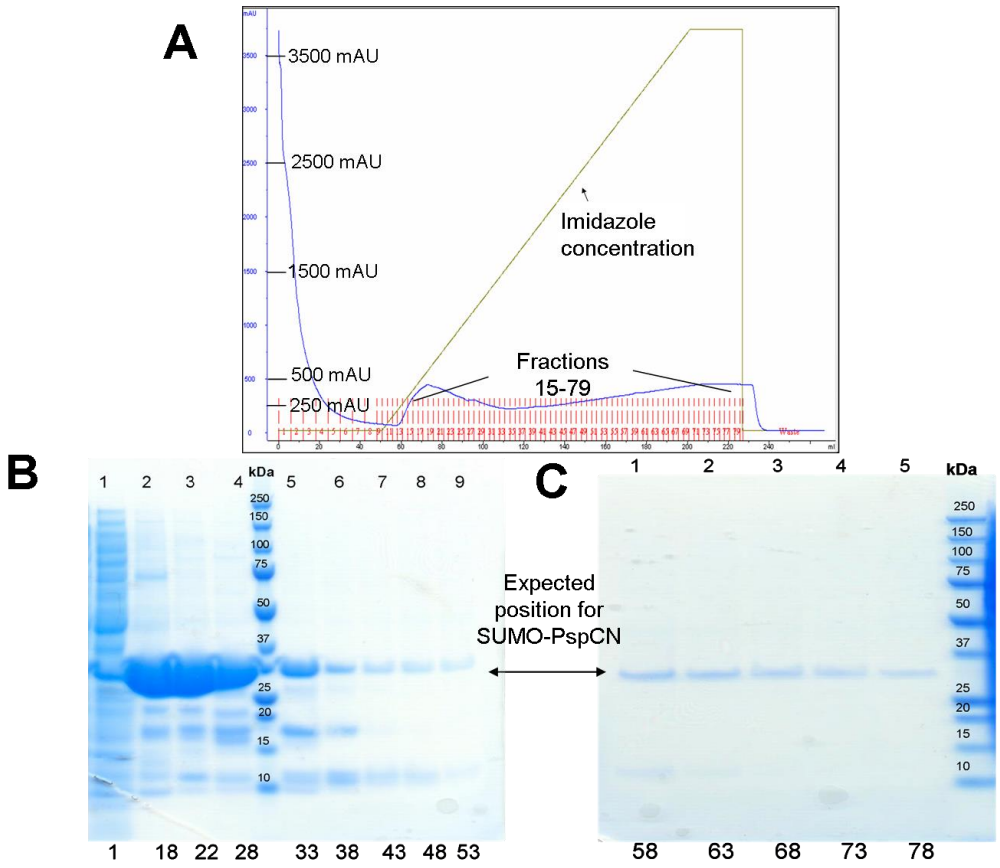


Figure 3.50 First purification step of SUMO-PspCN. (A) Chromatogram of the elution of the IMAC column. (B) Gradient SDS-PAGE of fractions from the elution shown in (A). Lanes 2-9 contain fractions-every 5th fraction as shown. (C) Gradient SDS-PAGE gel for the remaining fractions shown in (A).

The profile of the IMAC elution, with no distinct peaks, was similar to that obtained previously. Fractions were analyzed by SDS-PAGE (Fig. 3.50 B and C). (B) shows there is no detectable protein in the flow-through. All fractions from the elution show bands that presumably correspond to SUMO-PspCN that migrates to just below 37 kDa (despite the protein having a molecular weight of 24 kDa). Lanes 1-5 in (C) all contain bands for SUMO-PspCN.

It appears that fractions 18-78 all contain SUMO-PspCN. Fractions 15-80 were combined and split into two samples. The first sample (I) (~165 ml) was subjected to cleavage of the SUMO protein (as described in Materials and Methods section 2.2.5) while the other sample (II) was concentrated and buffer-exchanged in PBS containing 5 mM EDTA, in preparation for size-exclusion chromatography.

Once sample I had been subjected to cleavage, it was subsequently re-passed over the IMAC column and the cleaved protein was collected (that should have lost

its N-terminal His-tag). The column was then washed and once again the flow-through was collected. Finally, 20 ml (of 20 mM potassium phosphate buffer, pH 7.0, 500 mM NaCl, 500 mM imidazole) were applied and the flow-through collected separately and the results were analyzed by SDS-PAGE (see Fig. 3.51).

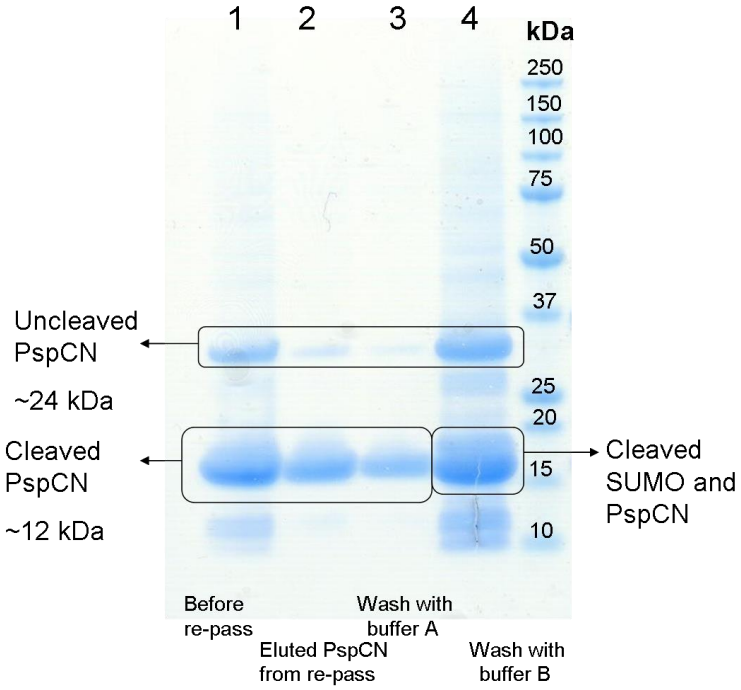


Figure 3.51 Gradient SDS-PAGE of samples collected after the re-pass through the IMAC column of SUMO-PspCN following cleavage.

From SDS-PAGE gel, it looks as if the majority of SUMO-PspCN was cleaved to afford the native PspCN, as reflected by the strong bands at the expected positions in lanes 1, 2 and 3 (Fig. 3.51). The fractions containing cleaved protein eluted from the second re-pass on the IMAC column were combined and concentrated in preparation for size-exclusion chromatography (Fig. 3.52). Samples from the size-exclusion chromatography runs were analyzed by SDS-PAGE.

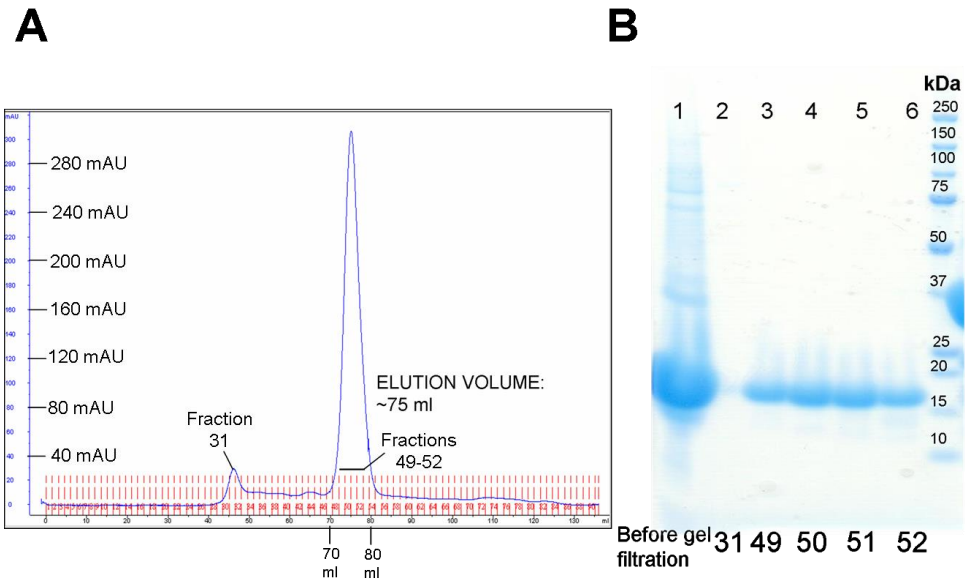


Figure 3.52 Size-exclusion chromatography for cleaved PspCN, using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram. The main peak is indicated with a black bar. (B) Gradient SDS-PAGE of samples/fractions of the main peak shown in (A) with the exception of lane 2 which contains a sample from the first lower peak in (A) and lane 1 which contains a sample prior to application on the column.

The profile of the chromatogram showed a main peak eluting at ~75 ml. When fractions from this peak were analyzed by SDS-PAGE (Fig. 3.52 B) they all showed single bands appearing just above 15 kDa thus leading to speculation that purified cleaved PspCN was contained in those fractions. Fractions 49-51 were combined and concentrated affording a 1 ml sample at a concentration of ~580 μM. From this, 400 μl were used to record NMR experiments (Figs. 3.53, 3.54), 5 μl in order to obtain accurate mass analysis (Fig. 3.55) while the rest (~600 μl) of the sample was frozen.

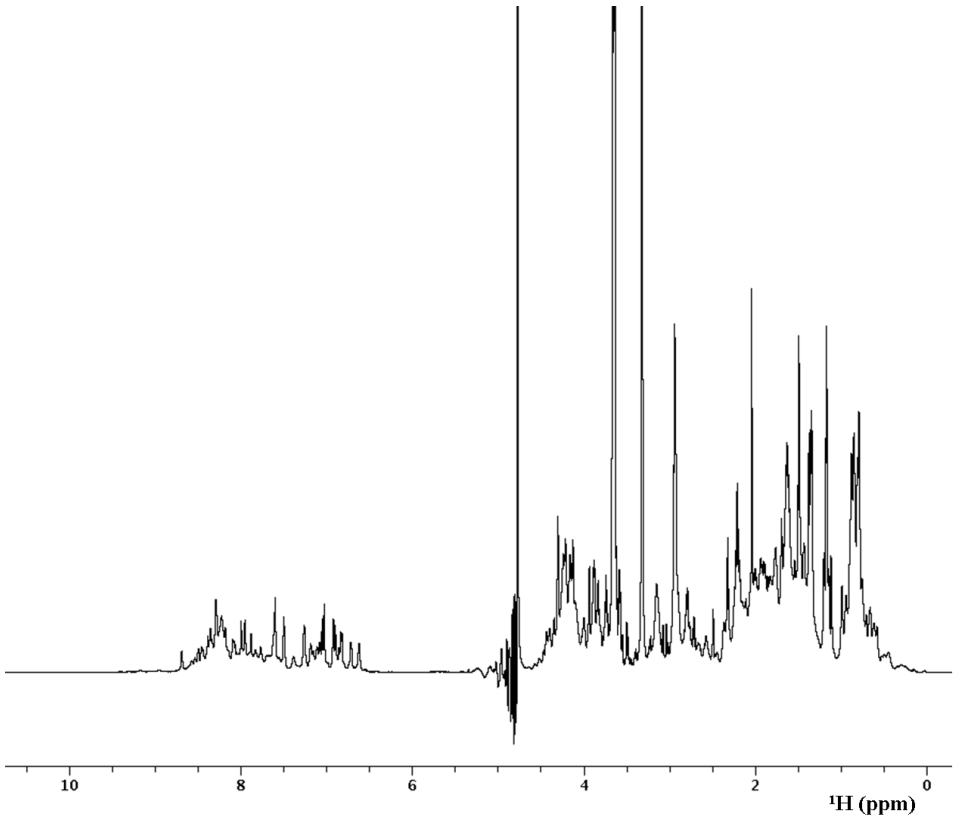


Figure 3.53 1D ^{1}H-NMR spectrum for PspCN. The sample was 580 µM in 20 mM potassium phosphate pH 6.4.

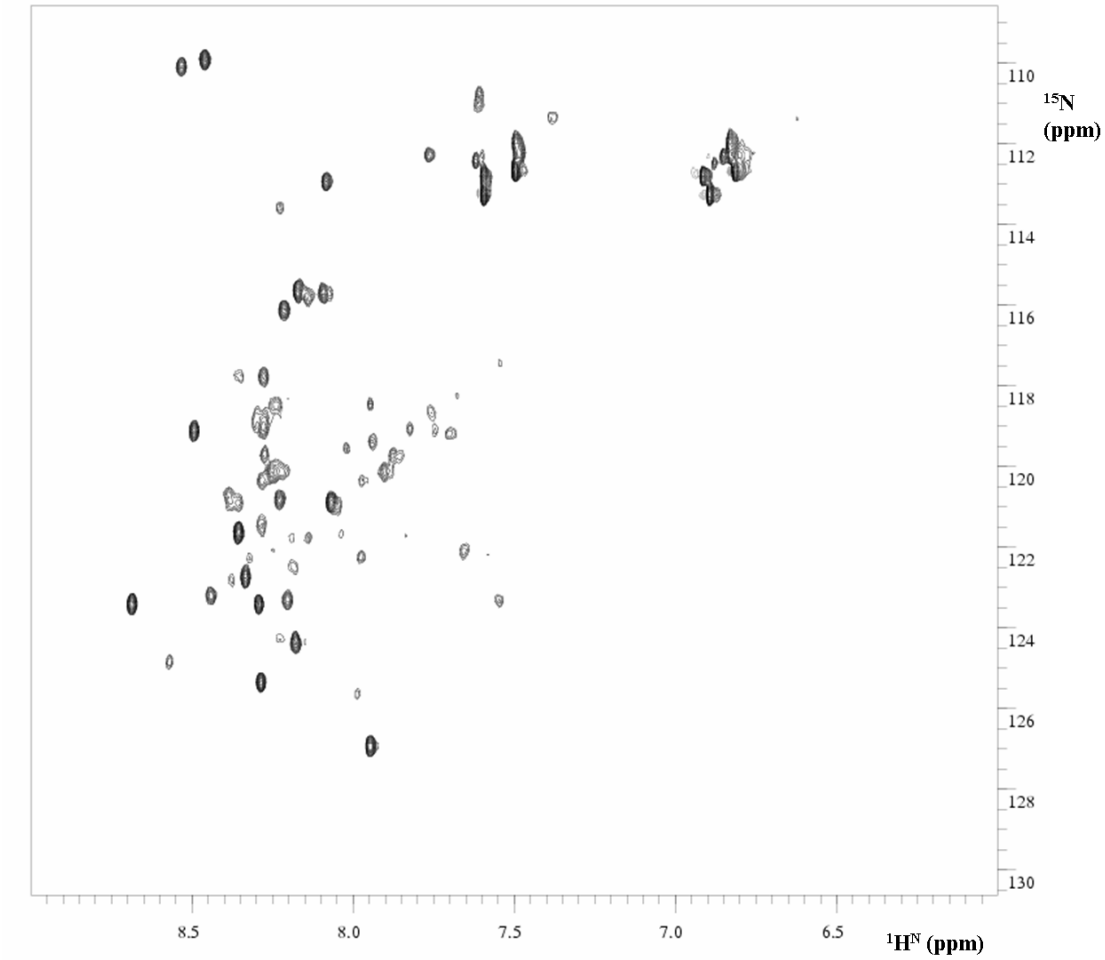


Figure 3.54 Natural abundance ^{1}H, ^{15}N-HSQC spectrum for PspCN. The sample was 580 µM in 20 mM potassium phosphate buffer pH 6.4.

The 1D ^{1}H, and 2D ^{1}H, ^{15}N-HSQC spectra indicate that there is some folded material present. This is reflected by the good dispersion and resolution exhibited by some of the peaks, however there is some potentially unfolded material as may be judged from clusters of not well-defined peaks appearing in parts of the spectrum. This can be more clearly observed in the natural abundance ^{15}N-HSQC.

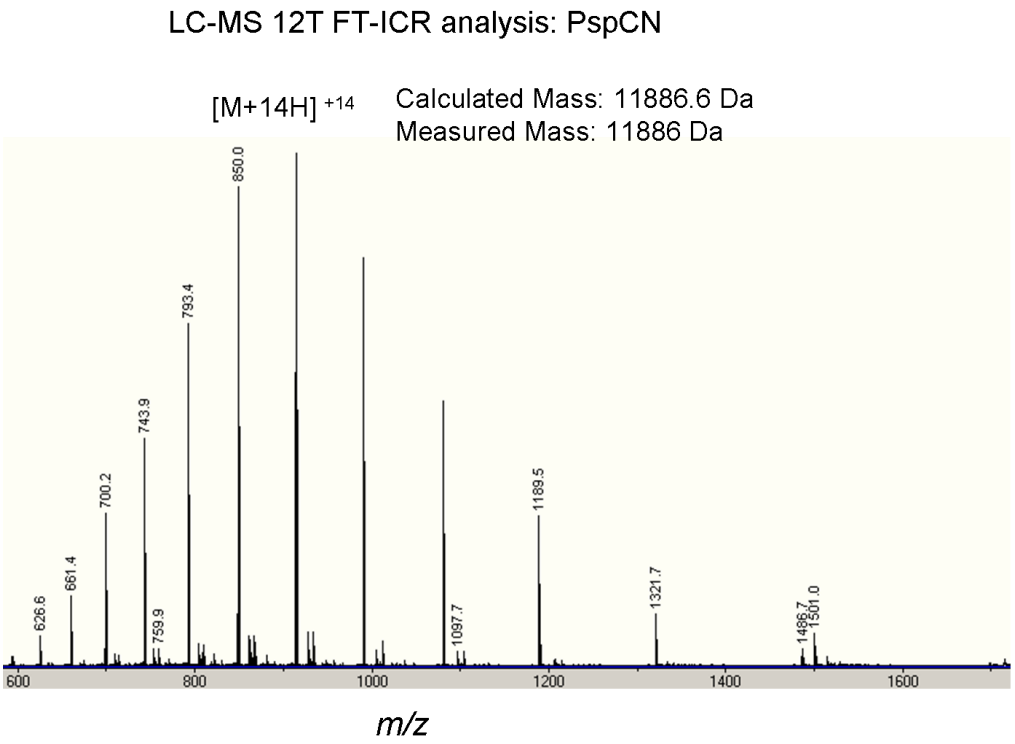


Figure 3.55 LC-MS spectrum for PspCN. All peaks correspond to charge state 14+.

According to the accurate mass analysis it can be concluded that the protein had the mass expected judging from the comparison between the calculated mass (without the N-terminal Met residue) and the measured mass, both of which are indicated on the spectrum.

Following the NMR experiments a SDS-PAGE gel was run to test the state and the stability of the protein after four days in the NMR instrument, the temperature of which was set at 25 ^{o}C (Fig. 3.56).

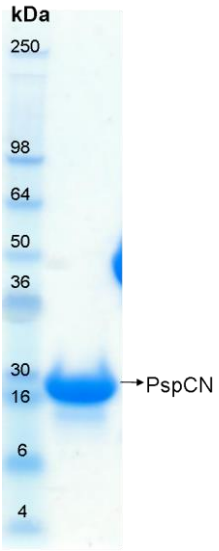


Figure 3.56 Gradient SDS-PAGE of a purified sample of PspCN after the completion of the NMR experiments.

In SDS-PAGE gel an overloaded band appears where expected, while negligible degradation is visible below the main band. This experiment suggested that the protein remained intact.

The second batch originating from the concentrated IMAC fractions, which remained uncleaved, was split in two samples and two size-exclusion chromatography runs were performed, one for each. Data for this can be found in the Supplementary CD section 1.14.

3.3.2 Construct SUMO-PspCNR1

The same procedure described for PspCN was followed and two *E. coli* strains were used in order to decide which was best for expression: BL21 (DE3) and BL21 (DE3) - pLysS. Single colonies were used to inoculate 5 ml of LB media containing 30 μg/ml kanamycin and both strains were tested at different temperatures and concentration of IPTG (Fig. 3.58).

The figure below (Fig. 3.57) shows the amino acid sequence used consisting of the SUMO sequence shown in magenta and the native PspCNR1 sequence shown in black.

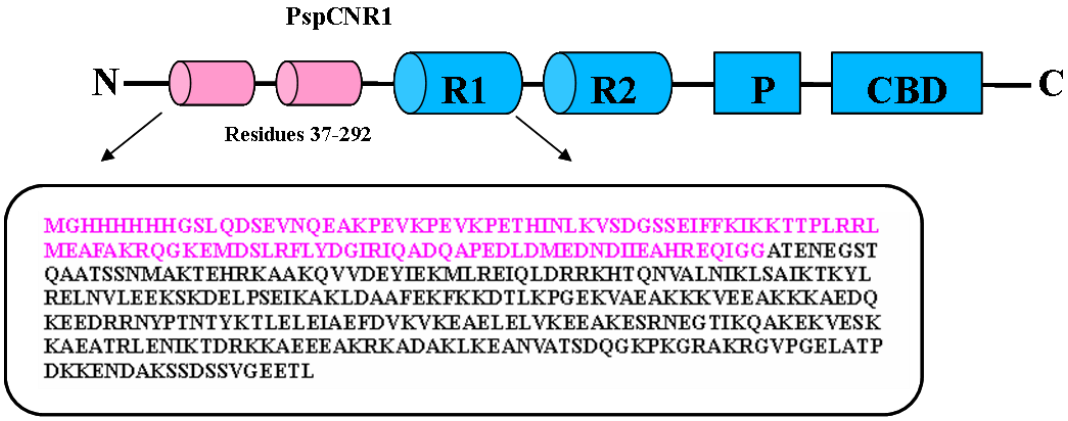


Figure 3.57 Representation of PspC where the sequence corresponding to the N-terminus-R1 domains fused with SUMO used in this study is indicated.

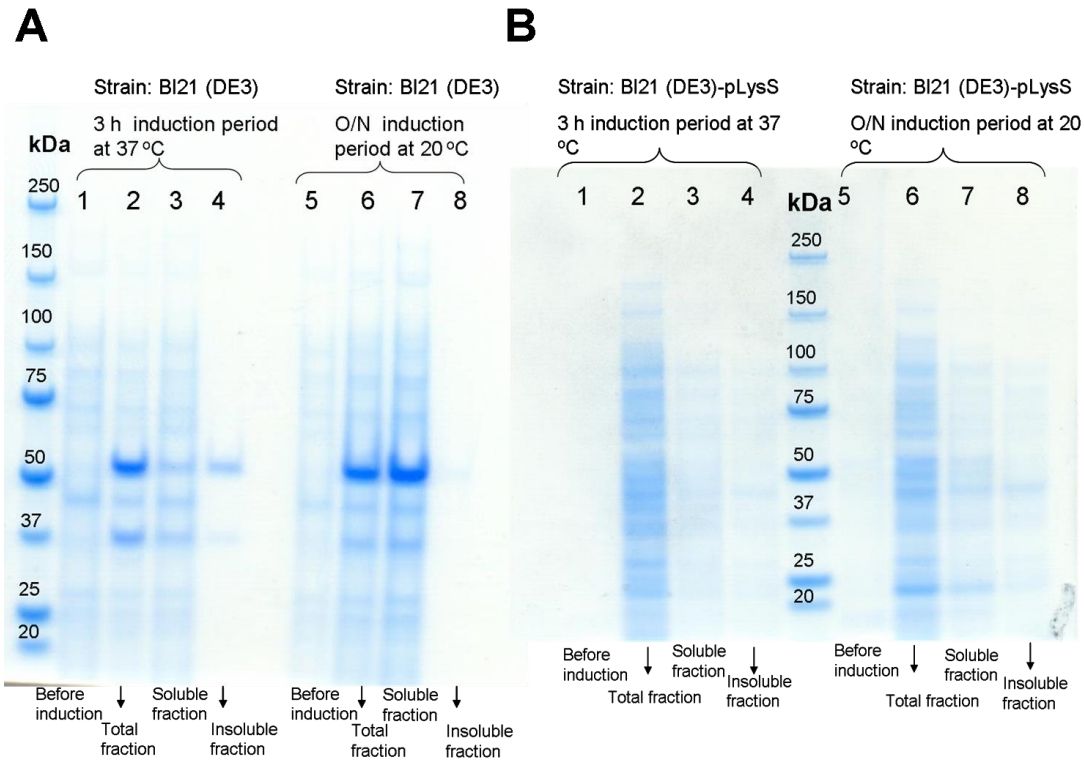


Figure 3.58 Gradient SDS-PAGE following the trial expressions of SUMO-PspCNR1. (A) Lanes 1-4 show the expression obtained by using the *E. coli* strain BL21 (DE3), as indicated on the gel. The band appearing just above 50 kDa corresponds to SUMO-PspCNR1 and is visible in lanes 2-3. Lanes 5-8 show the expression for the strain BL21 (DE3) as indicated. Bands for the expected protein appear just above 50 kDa in lanes 5-8. (B) Lanes 1-4 show the expression for the strain BL21 (DE3)-pLysS. SUMO-PspCNR1 does not appear where expected in any of the lanes. Lanes 5-8 show the expression for the strain BL21 (DE3)-pLysS. Bands for SUMO-PspCNR1 cannot be seen in any of the lanes.

Analysis by SDS-PAGE indicates that the strain BL21 (DE3) –pLysS was not at all suitable for the expression of SUMO-PspCNR1. The strain that yielded the most protein was BL21 (DE3), following overnight induction at 20 ^{o}C. These conditions also yielded the most favorable ratio of soluble to insoluble protein.

For large-scale expression, the procedure followed was identical to that already described for SUMO-PspCN in Section 3.4.1. The wet cell weight was 6.5 g. The outcome of the first purification step (IMAC) is shown in Figures 3.59 and 3.60.

.

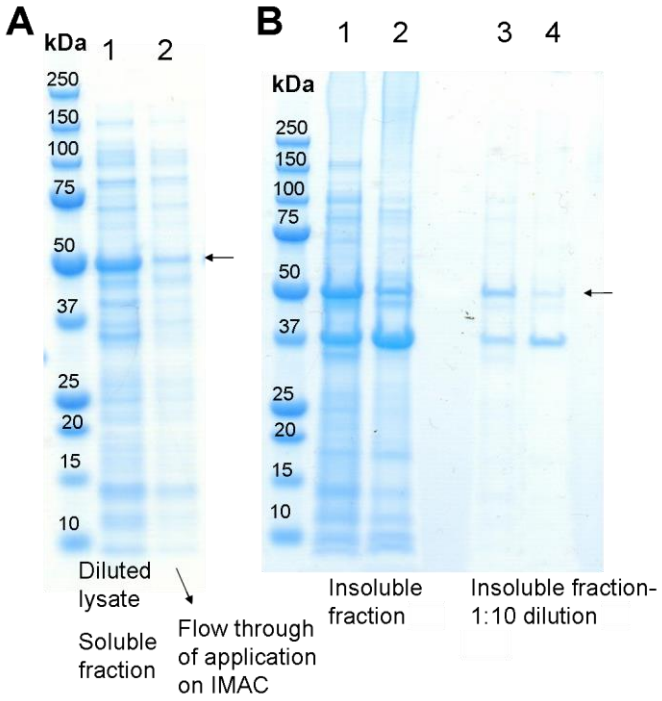


Figure 3.59 Gradient SDS-PAGE of the following samples: (A) Lanes 1 and 2: samples as indicated. (B) Lanes 1-4: samples as indicated. The lane between lane 2 and the molecular weight marker in (A) has been masked with a white box as it contained a sample not relevant to this section. All lanes contain bands at the expected molecular weight of around 50 kDa (indicated with an arrow).

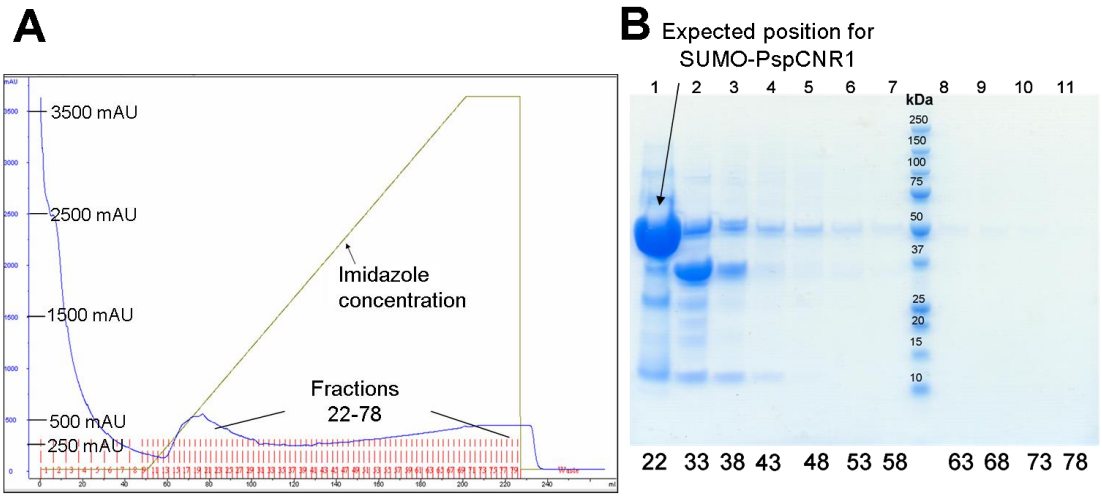


Figure 3.60 First purification step for SUMO-PspCNR1. (A) Chromatogram of the elution of the IMAC column. (B) Gradient SDS-PAGE of fractions from the elution shown in (A) as indicated below the gel.

The profile of the elution following launch of the gradient is similar to what has been previously observed. In SDS-PAGE analysis, lanes 1-11 in (B) correspond to fractions spanning 22-78 and all contain bands for SUMO-PspCNR1, around 50 kDa, although in lanes 8-11 they are very faint.

The expected molecular weight for SUMO-PspCNR1 is 41.173 Da. However, it runs more slowly than standard proteins of the same size so the band appearing to have a MWt of ~50 kDa could be the target protein. Fractions 15-80 were pooled and spilt into batches I and II which were subsequently treated as described for SUMO-PspCN in section 3.4.1.

After overnight dialysis/cleavage of Batch I, a sample was kept aside for SDS-PAGE and the rest re-applied to the IMAC column that had been pre-equilibrated with the same buffer. This was eluted and analyzed by SDS-PAGE as in the case of SUMO-PspCN (Fig. 3.61).

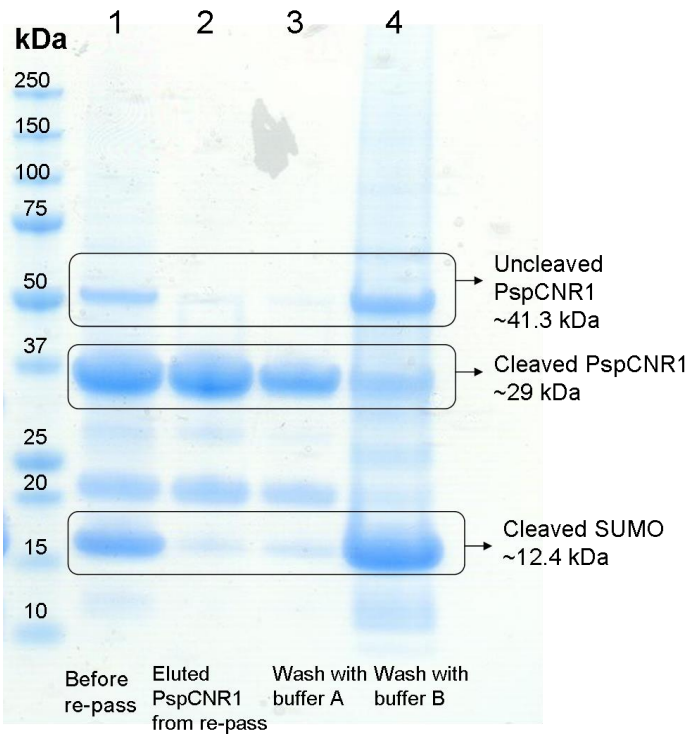


Figure 3.61 Re-pass of cleaved SUMO-PspCNR1 on the IMAC column.

As can be observed from SDS-PAGE the majority of the protein batch was successfully cleaved. Fractions from the IMAC column containing the cleavage product were combined and concentrated for size-exclusion chromatography (Fig. 3.62).

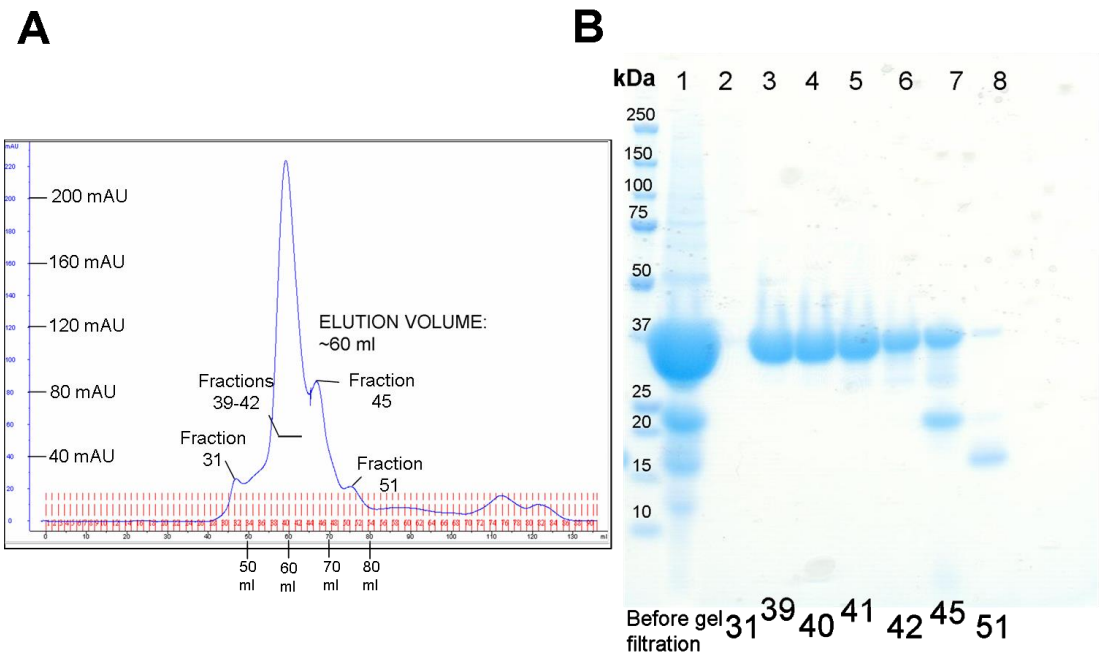


Figure 3.62 Size-exclusion chromatography for cleaved PspCNR1, using a HiLoad 16/60 Superdex 75 column. (A) Chromatogram of the size-exclusion chromatography run. (B) Gradient SDS-PAGE of a sample before the size-exclusion chromatography run, in lane 1, and fractions from the chromatogram shown in (A) (lanes 3-8).

The highest peak, indicated with a black bar in Fig. 3.62, probably corresponds to cleaved PspCNR1. In the SDS-PAGE that followed, lanes 3-5 contain fractions from the main peak indicated in the chromatogram and all show single bands just below 37 kDa corresponding to pure protein. PspCNR1 is 28.9 kDa however it migrates in an SDS-PAGE gel to just below 37 kDa due to its high positive charge. More pronounced degradation is visible in lane 6 containing fraction 42 while in lane 7 which contains a fraction from the following shoulder of the main peak degraded protein amounts for half the sample. Lane 8 contains mostly degraded material.

SDS-PAGE revealed that pure cleaved PspCNR1 were eluted in fractions 39 and 40. In fractions 41 and 42 some slight degradation is visible. Therefore fractions 39-40 were combined, concentrated and afforded a sample of 1 ml at a concentration of 120 μM. Half of that sample was used to record NMR experiments (Fig. 3.63) and mass spectrometry data (Fig. 3.64) while the rest was stored in the -80 ^{o}C freezer.

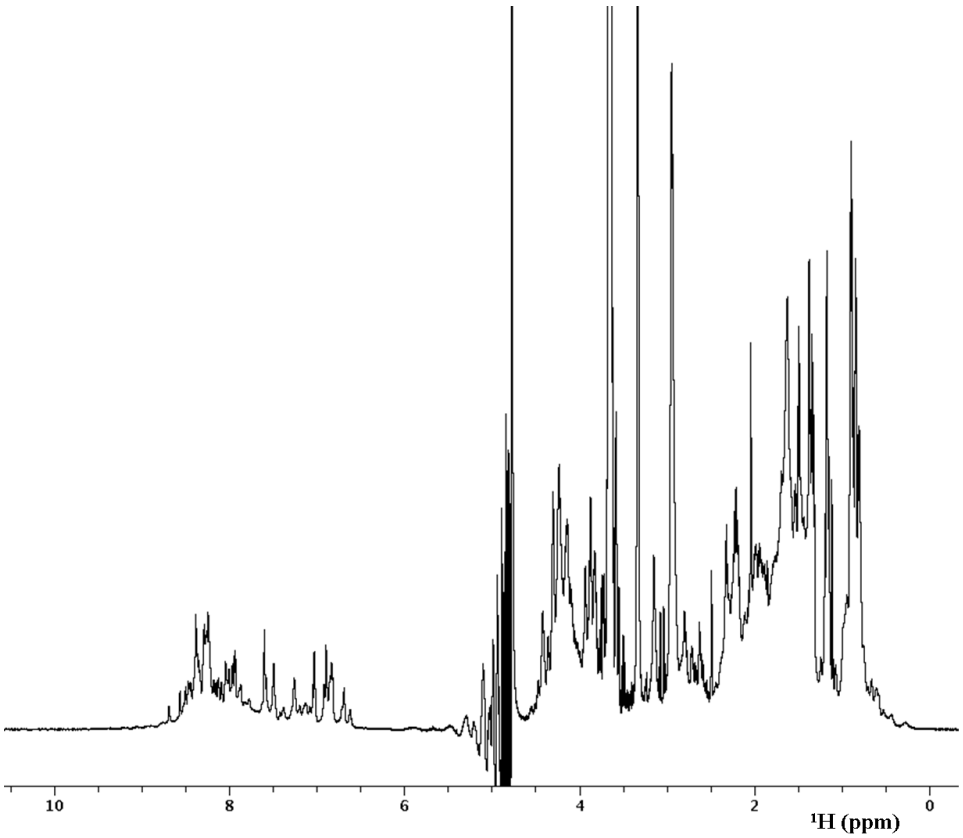


Figure 3.63 1D ^{1}H-NMR spectrum for PspCNR1. The sample was 120 µM in 20 mM potassium phosphate buffer, pH 6.4.

SDS-PAGE (Fig. 3.65) confirmed that the sample was stable after several days in the spectrometer the temperature of which was set to 25 ^{o}C.

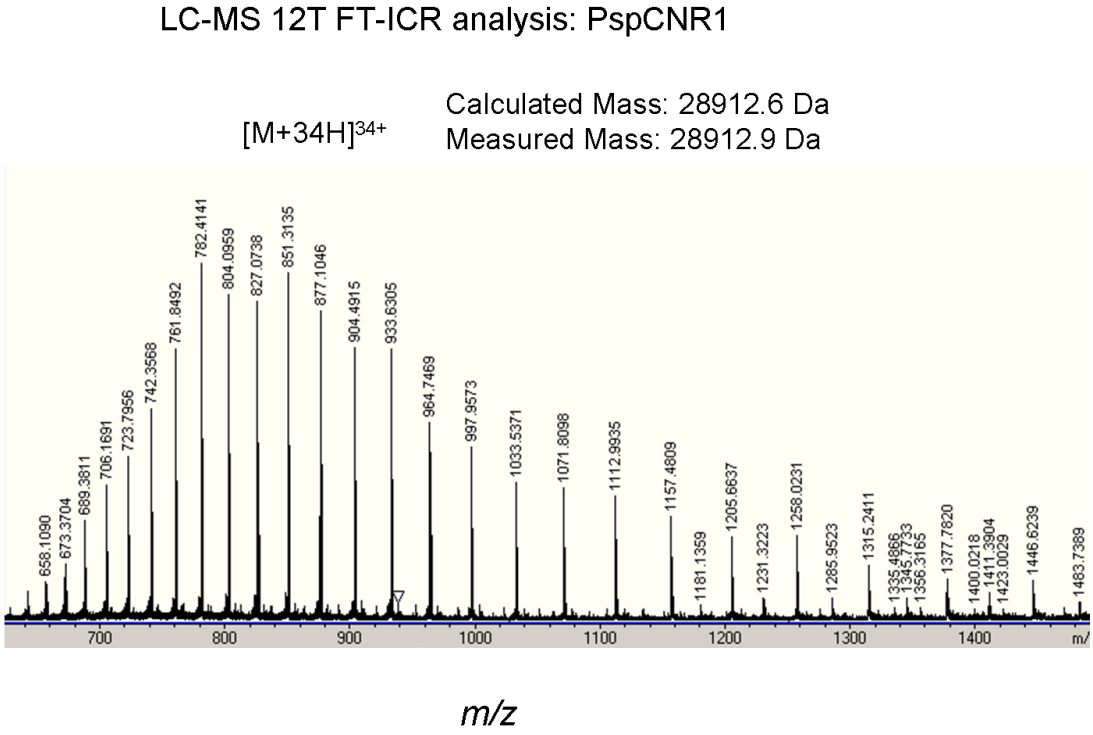


Figure 3.64 LC-MS spectrum for PspCNR1. All peaks correspond to charge state 34+.

Accurate mass analysis verified the presence of intact protein as the measured and expected mass are in very good agreement.

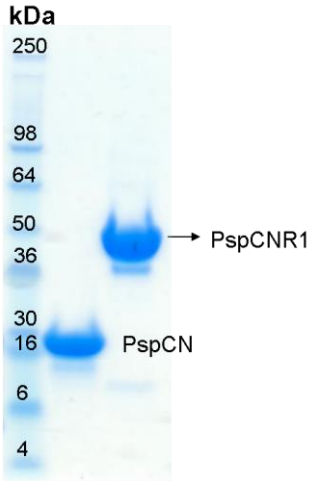


Figure 3.65 Gradient SDS-PAGE on a sample after the NMR experiments for PspCN and PspCNR1.The first lane contains the same sample as shown in figure 3.63. PspCNR1 appears as a strong in the second lane at the expected molecular weight. Lower faint bands correspond to some mild degradation.

The SDS-PAGE gel shows an overloaded band at the expected weight consistent with intact protein. Despite some very mild degradation compared to the intact portion that the protein has undergone, it was considered suitable for further experiments at room temperature.

A similar approach was followed for the purification of SUMO and uncleaved SUMO-PspCNR1. The results for both can be found in the Supplementary CD sections 1.15 and 1.16.

3.4 Conclusions- Protein production, purification and characterization

In this study three segments of the complement regulator, FH, and two segments of PspC, a bacterial protein and important virulence factor expressed by *S. pneumoniae*, were overproduced as recombinantly expressed gene products. The remaining FH constructs presented in this chapter were already available either as DNA incorporated in the *P. pastoris* genome, or as expressed and unpurified material and were further processed accordingly by the author for use in functional studies.

One of the FH segments, FH-10-11, was chosen for further expression in ^{15}N and ^{15}N, ^{13}C-enriched media for the purpose of three-dimensional structural studies. Two ^{15}N-labeled and three ^{15}N, ^{13}C-labeled were produced in total in order to cover all experiments required. The quality of the obtained material was excellent as reflected by purity confirmed by SDS-PAGE in conjunction with accurate mass analysis and finally by how well the proteins were folded as judged by the NMR experiments. The majority of the remaining FH proteins prepared in the current work proved to be pure as judged by SDS-PAGE, folded as judged by NMR experiments, and produced in adequate yields for the studies in plan, while in some cases accurate mass analysis confirmed the correct primary sequence. The exceptions were FH-4-6, FH-13-14 and FH-14-15. In the case of FH-4-6 a relatively pure sample was obtained however at a low yield. In the case of FH-13-14 adequate yields were achieved, however the protein became degraded almost immediately. In addition, as judged by the ^{1}H, ^{15}N-HSQC spectrum it appears that some part of the construct is folded while the remaining is aggregated and/or unfolded. This could potentially arise from FH-14 as a ^{1}H, ^{15}N-HSQC spectrum of FH-13 alone showed

properly folded material (30). One possibility is that in the context of a larger construct FH-13-14 is stabilized by its neighboring CCPs. If that were the case, then the site(s) in FH-13-14 susceptible to cleavage (when the construct is produced as a double CCP module) may no longer be accessible to proteases when in the context of FH-10-15 or FH-8-15. In the case of FH-14-15 due to the limited investigation performed it can be said that, similarly to FH-13-14, the protein becomes degraded immediately, however at which stage takes place is not known. It is interesting though that this construct also involves FH-14. The fact that neighboring CCP modules of 13, 14 and 15 may be required for the observed shift in the behavior of the proteins (when going from double CCP constructs to bigger segments), may be evidenced by the fact that larger FH regions, *i. e.* FH-10-15 and FH-8-15 appear to be much more stable, not as susceptible to degradation and to a large extent folded. For FH-10-15 this conclusions can be drawn based on the present work while for FH-8-15 also based on previous work performed by Dr C. Q. Schmidt. Overall, it appears that within the FH-10-15 structure there are compactly folded modules along with at least one less well-ordered module (see also section 3.1.6.3 of this chapter). Taking into account the observations above for FH-10-11, FH-13-14 and FH-14-15 together with the excellent quality of NMR spectra obtained for FH-11-12 (discussed further in Chapter 4), FH-13 (30) and FH-15 (44), the best candidate for a disordered module is FH-14. The presence of disorder in FH-10-15 might explain the difficulty that was encountered in crystallizing this construct despite a serious effort (Dr C. Q. Schmidt, personal communication).

The use of *E. coli* for the expression of the PspC constructs was fast and efficient. Three different versions of PspCN were produced and one version of PspCNR1. Adequate yields were produced in all cases. In addition, SDS-PAGE confirmed their purity while accurate mass analysis confirmed the correct primary amino acid sequences. When the "SUMO" approach was followed (utilizing the pE-SUMOPro Kan vector) higher yields and increased solubility were achieved. In addition subsequent cleavage of the SUMO from the expressed fusion products (*i. e.* SUMO-PspCN and SUMO-PspCNR1) resulted in the native sequence for both proteins. A natural abundance ^{1}H, ^{15}N-HSQC performed on PspCN revealed the presence of folded material (supported by the good peak-dispersion observed in parts

of the spectrum) however there is also evidence for incomplete folding of the protein. A similar however not as conclusive comment can be made for PspCNR1 based on the ^{1}H-NMR experiment. It is likely that the PspC proteins adopt a folded conformation upon binding to a target partner. In fact a ^{1}H, ^{15}N-HSQC spectrum recorded on PspCN in complex with FH-8-9 was consistent with proper folding of PspCN (experiments performed by Dr A. Herbert). In conclusion this work successfully produced a panel of proteins the majority of which were suitable for studies based on NMR and SPR.

CHAPTER 4

NMR-based structural studies of FH-10-11 and FH-11-12[$]

[$] Parts of this chapter have been published in (149).

Overview

Recent models for the mechanism of action of FH envisage its central region as playing a major architectural role that is key to its biological function of minimising complement-mediated damage to the host (15), (43). According to this model, the C- and N- terminal regions are brought into spatial proximity, thereby allowing for cooperation in engaging their binding partners, namely C3b and polyanionic carbohydrates.

As detailed in the Introduction, a binding site for C3b is located within FH-1-4. Indeed these modules can, alone, carry out the fluid-phase CA and DAA of FH. FH-7 contains a polyanion-binding site that probably aids in self-surface recognition. At the other end of the molecule, FH-19-20 bind C3b as well as polyanionic self-surface markers. There is a plethora of high-resolution structural data for these regions. Conversely, the lack of binding sites for C3b or polyanions in FH-8-15 has resulted in less attention as only limited structural information is available for this region. Intriguingly, the data that do exist suggest these modules adopt a unique compact conformation (15, 43). Even the incomplete picture currently available is suggestive of a mechanism of how the N- and C-terminal binding sites might cooperate. There is a clear need to acquire more structural information with the aim of further elucidating the spatial organisation of the central modules of FH.

Within FH, CCPs 10-15 are connected by the longest intermodular linking sequences. These allow for bends between modules that can thereby make more contacts with one another and form compact and potentially rigid structures as indicated by recent biophysical data (43). Previous attempts to crystallise FH-10-15 (personal communication, Dr C. Q. Schmidt, University of Edinburgh) were unsuccessful. The only high-resolution structure available to date is that of CCP-module pair FH-12-13 solved by NMR (43). Interestingly, a rigid 80^{o} tilt or bend angle is formed between these two modules. This information was helpful, but inadequate, for the unambiguous interpretation of SAXS curves generated from samples of FH-10-15 and FH-8-15.

In order to obtain additional high-resolution information on the modules that constitute the central region of FH, two module pairs, FH-10-11 and FH-11-12 (Fig.

4.1) were selected for solution structural determination using NMR. These recombinant protein fragments were judged to be suitable for NMR in terms of their yield of production, size (~13 kDa), solubility, homogeneity following purification and relative stability of the final samples. The aim was to determine their structures, then combine these, and the previously solved FH-12-13 structure, to generate models of FH-10-12 and FH-10-13. These could then be compared against SAXS data collected both on the module-pairs and on longer fragments (e.g. FH-10-12, 11-14, 10-15 and 8-15) (SAXS data will be presented in Chapter 5).

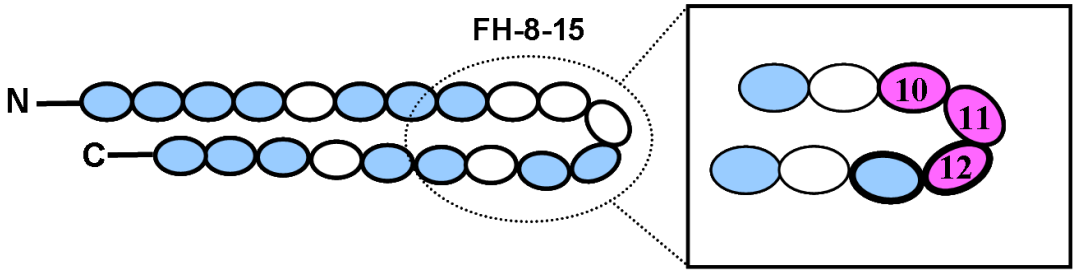


Figure 4.1 Representation of the 20 CCP modules that comprise FH. The CCP modules for which high-resolution structures are available are coloured in light blue. The FH pairs 10-11 and 11-12 of the central region that were subjected to structural studies in the current work are shown in pink. The structure of FH-12 was also solved by NMR in the context of FH-12-13 (43) (drawn using bolder outlines).

4.1 NMR structures of FH-10-11 and FH-11-12

For 3D structural studies by NMR of the two pairs, FH-10-11 and 11-12, triple-resonance techniques were required and hence double-labeled samples were produced. Pure samples were prepared as follows: 400, 150 and 500 μM FH-10-11 in 20 mM potassium phosphate buffer, pH 6.7; and 650 μM FH-11-12 in 20 mM potassium phosphate buffer, pH 6.3. These samples were subjected to a standard suite of double- and triple- resonance NMR experiments as described in Materials and Methods.

Good quality spectra obtained for both proteins enabled for the near-complete assignment of ^{1}H, ^{15}N, and ^{13}C nuclei. Specifically, for FH-10-11 94 % of amide, 97 % of backbone and 97 % of sidechain H nuclei were assigned (by the author). In the case of FH-11-12, 94 % of amide, 95 % of backbone and 96 % of sidechain H

atoms were assigned (by Ilias Matis). Figures 4.2 and 4.3 show the assigned ^{1}H, ^{15}N-HSQC spectra of FH-10-11 and FH-11-12.

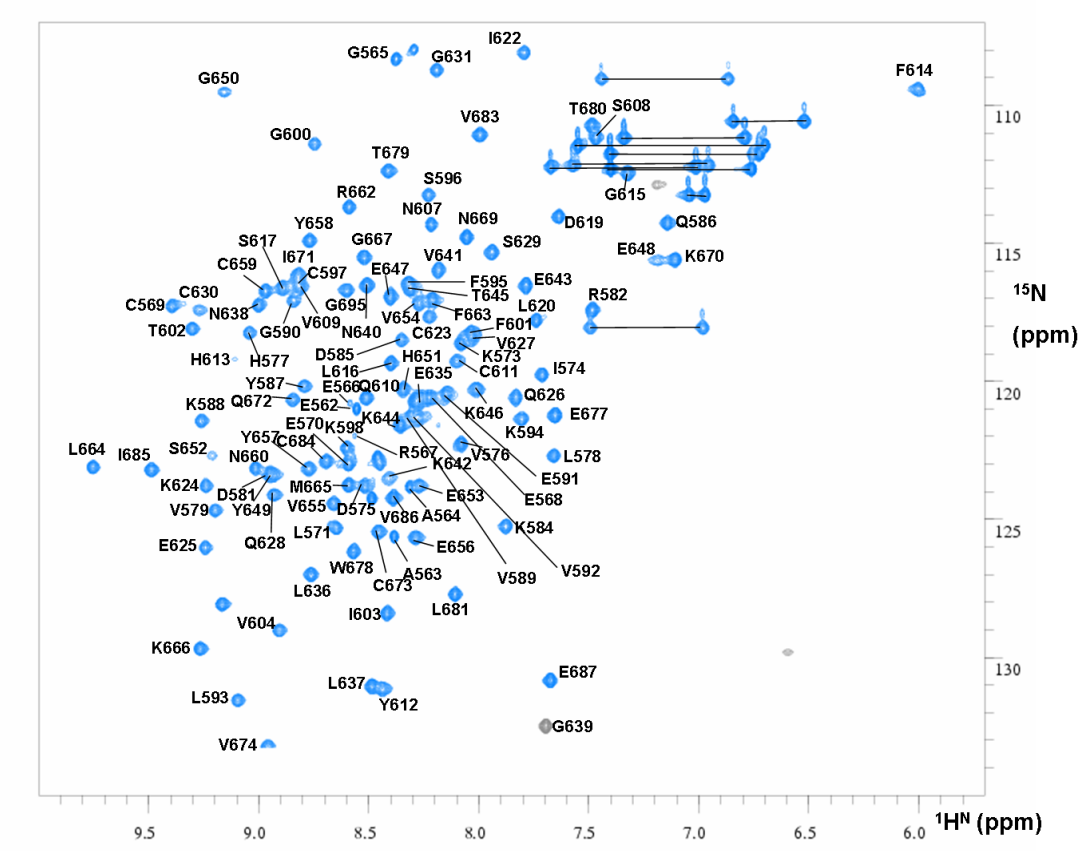


Figure 4.2 Assigned ^{1}H, ^{15}N-HSQC spectrum of FH-10-11. The sidechain NH resonances of Gln and Asn residues are connected by a horizontal line. Folded (negative or antiphase) cross peaks are shown in grey.

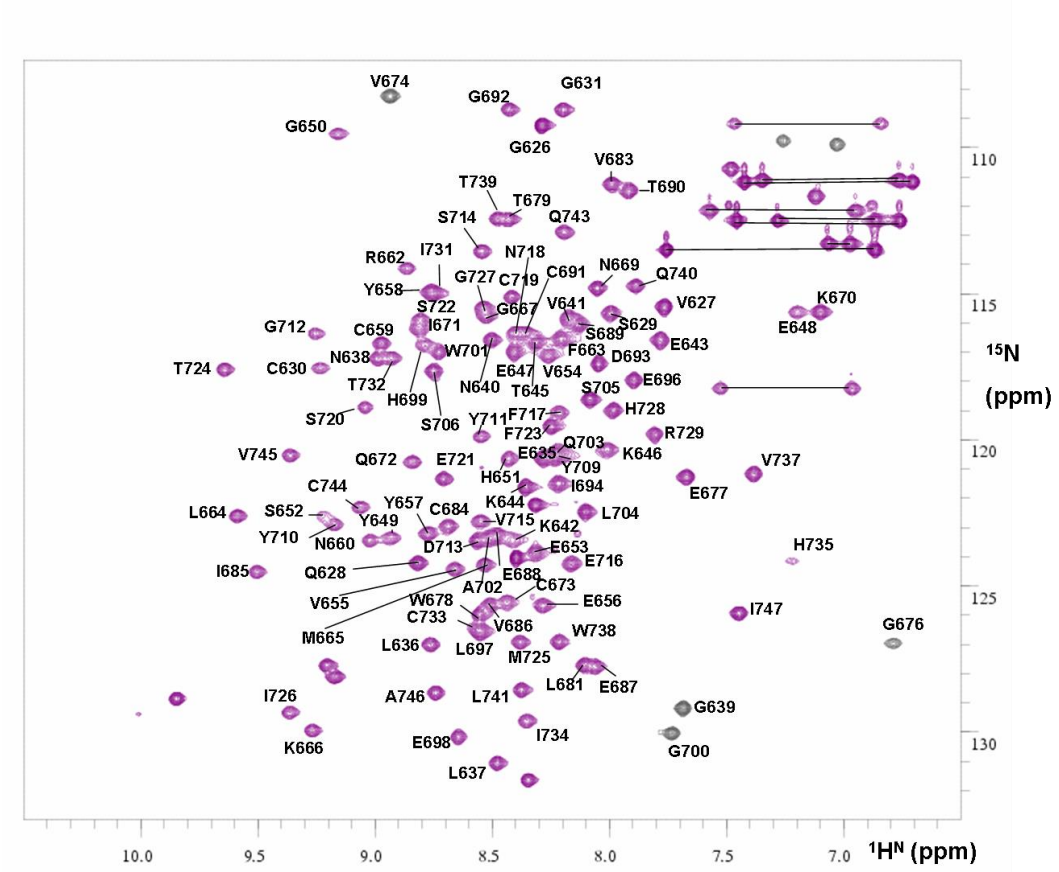


Figure 4.3 Assigned ^{1}H, ^{15}N-HSQC spectrum for FH-11-12. See legend to Figure 4.2 for explanation.

4.1.1 Structure calculations

The figure below (Fig. 4.4) summarizes the basic steps followed en route to obtaining the NMR-derived solution structures of FH-10-11 and FH-11-12.

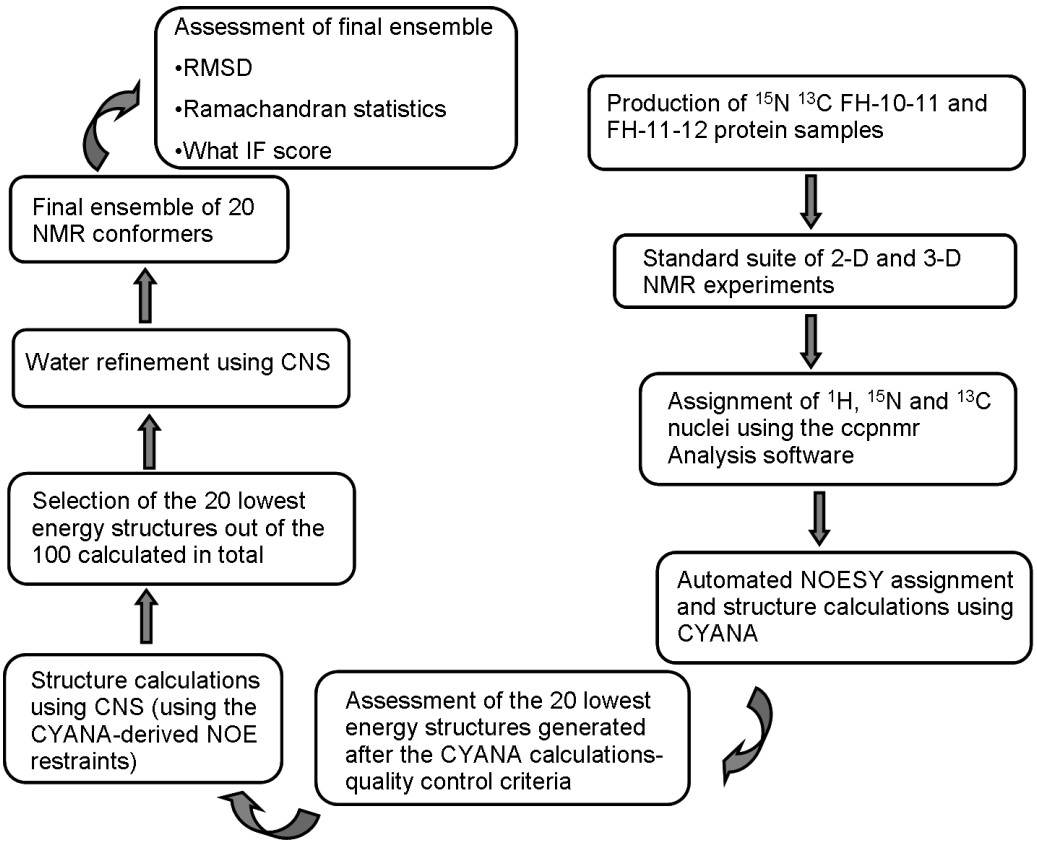


Figure 4.4 Summary of the most important steps followed in this study for the purpose of obtaining the two NMR-derived ensembles of FH-10-11 and FH-11-12.

To determine the 3D structures by NMR, distance constraints are extracted from nuclear Overhauser effect spectroscopy (NOESY) spectra. These NOEs may be equated to upper-distance bounds for one (or more, if ambiguous) pairs of protons. A large number of pairs of interacting hydrogen nuclei must be identified to generate adequate sets of distance restraints.

For the purpose of this study, the program CYANA (126) was utilized (see section 2.3.4.9.2 of Materials and Methods). Following the complete structure calculation, *i.e.* after seven CYANA cycles, the 20 lowest-energy structures were selected out of the 100 calculated in total, and their reliability was assessed according to several quality-control criteria (see Appendix section C1 for details).

After the structure calculations in CYANA a total of 100 structures were calculated within CNS (130) for each of FH-10-11 and FH-11-12, and these were ranked according to overall and NOE energies (see Figs. 4.5 and 4.6).

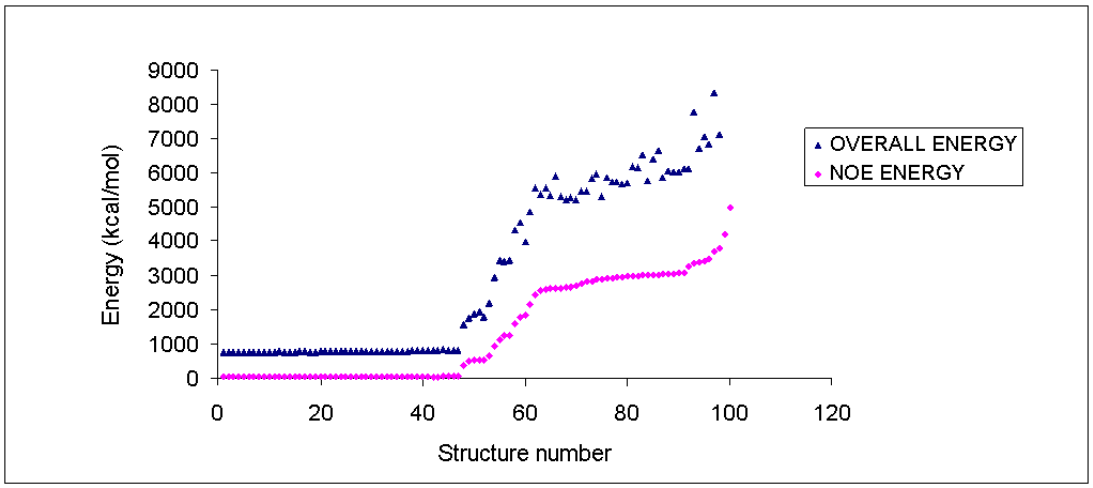


Figure 4.5 Energy plots for the 100 structures calculated in the final CNS run for FH-10-11. Overall and NOE energy are shown for each structure.

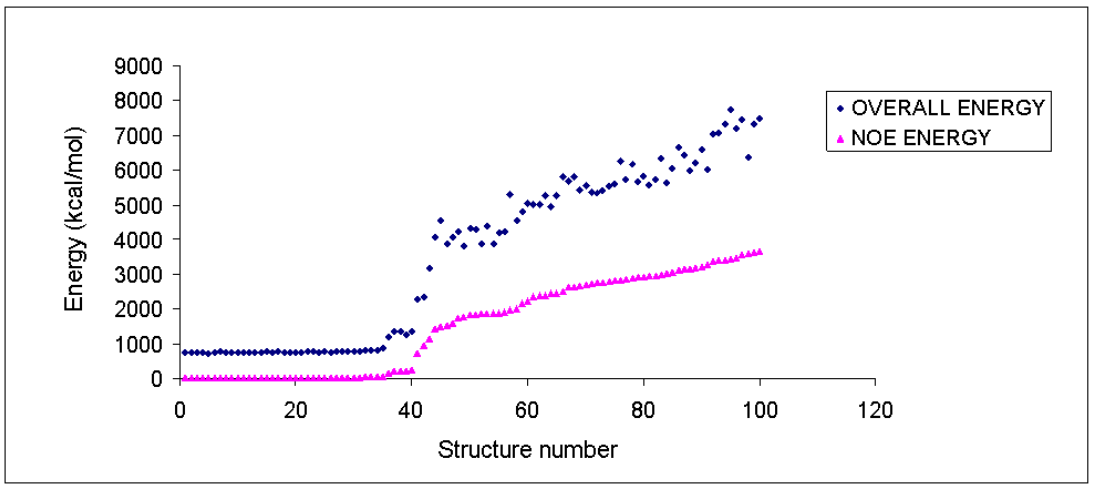


Figure 4.6 Energy plots for the 100 structures calculated in the final CNS run for FH-11-12. Overall and NOE energy are shown for each structure.

In these plots, some 30-40 calculated structures lie to the left of the first inflection point, suggesting convergence. It was decided, however, that the 20 lowest-energy structures in each case represented an adequate ensemble and all of these structures also met the criterion of having no NOE violations in excess of 0.5 Å. These ensembles were then subjected to water refinement using the appropriate scripts in CNS. The resultant ensembles are presented in figures 4.7 and 4.8.

As may be been judged from RMSD values for the two ensembles (when superimposed on the backbone atoms), excellent structural convergence was

obtained at the level of individual modules. Moreover, the RMSD values of the module pairs are also relatively low, consistent with small variations in inter-modular orientation within each ensemble.

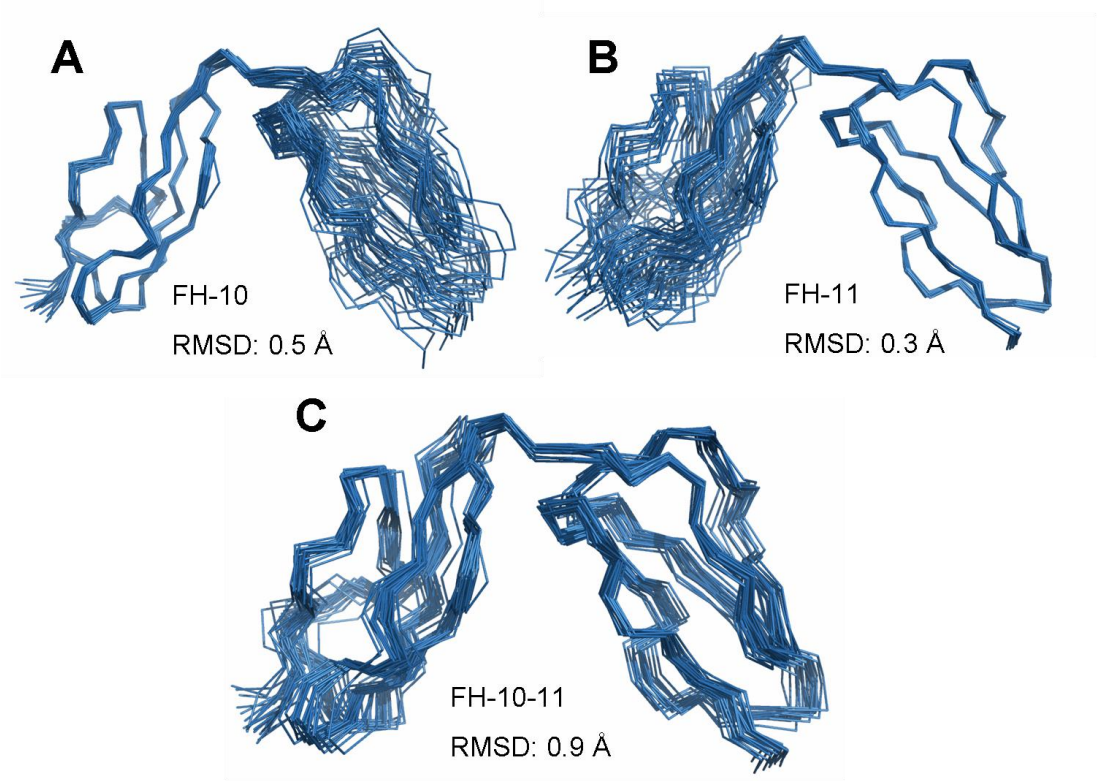


Figure 4.7 Ensemble of NMR-derived structures for FH-10-11 after water refinement in CNS. Backbone overlays (and RMSDs referring to backbone atoms) of the 20 lowest-energy structures. A. Overlaid on module 10. B. Overlaid on module 11. C. Overlaid on both modules. The 0.9 Å RMSD indicates a good convergence and a small deviation in intermodular orientation.

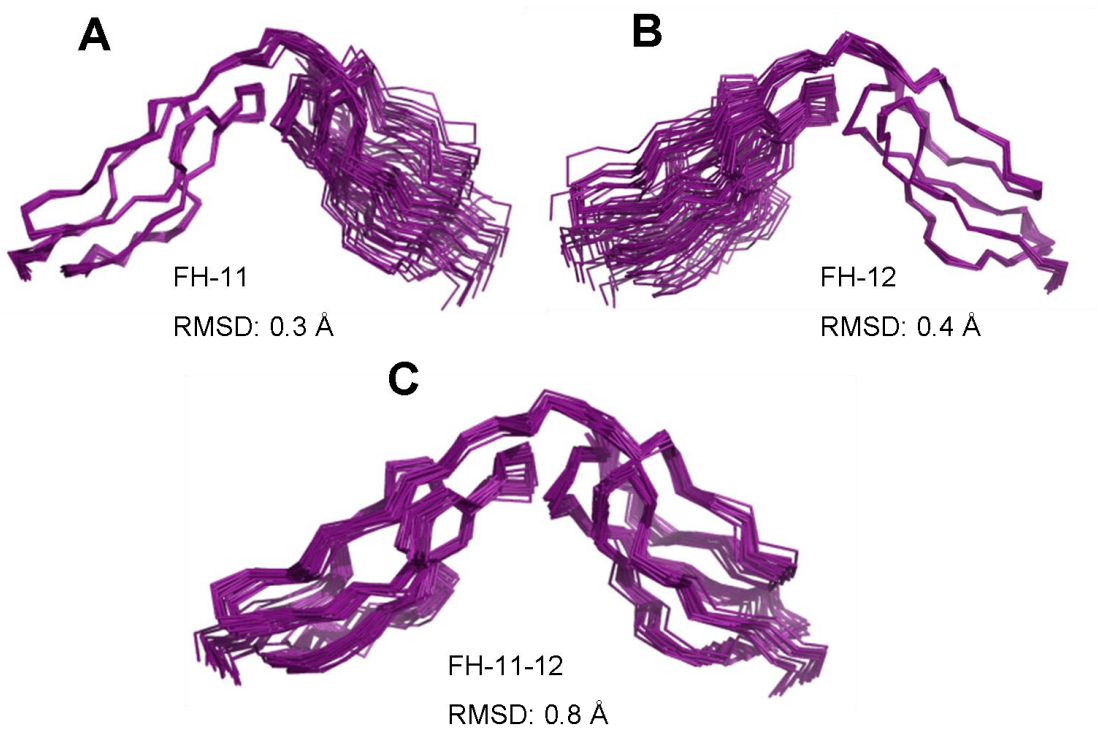


Figure 4.8 Ensemble of NMR-derived structures for FH-11-12 after water refinement in CNS. Backbone overlays (and RMSDs referring to backbone atoms). A. Overlaid on module 11. B. Overlaid on module 12. C. Overlaid on both modules. The 0.8 Å RMSD indicates a good convergence and a small deviation in intermodular orientation.

4.1.2 Description and characteristics of the structures

The modules 10, 11 and 12, have structures that are typical of CCP modules. Each module contains two disulfide bonds that are formed in a Cys (I)-Cys (III), Cys (II)-Cys (IV) manner. In FH-11 and 12, the consensus Trp residue located between the third and the fourth Cys residue is buried in a small hydrophobic core. However in FH-10 this Trp is absent and a Leu residue lies in its place (Fig. 4.9). Like all CCPs, modules 10, 11 and 12 of FH have elongated, ovoid structures in which anti-parallel β-strands are the predominant type of secondary structure (Figs. 4.10 and 4.11). In addition, in both pairs there is a *cis*-Pro residue (^{618}Pro in FH-10 and ^{708}Pro in FH-12) (see also Materials and Methods section 2.3.4.7).

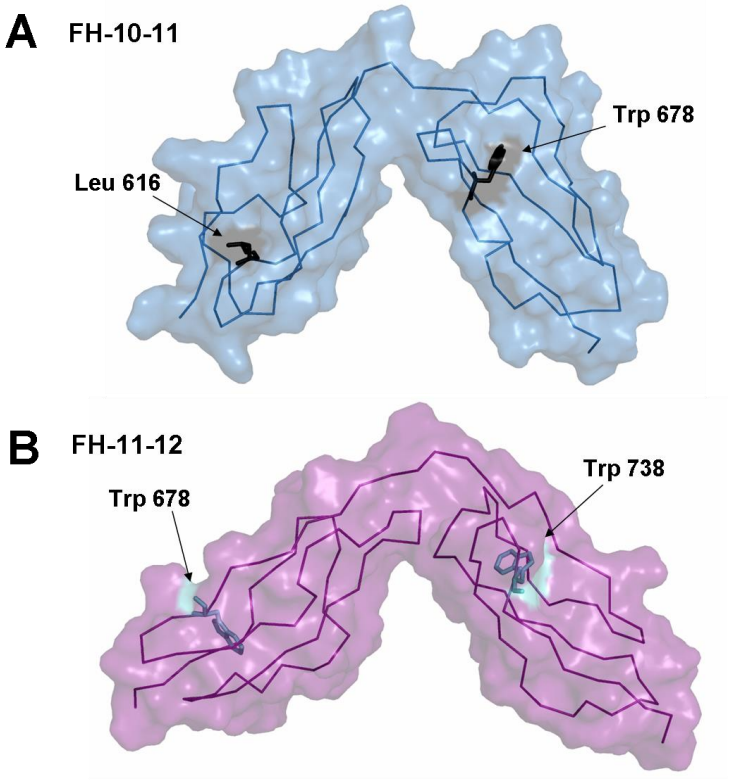


Figure 4.9 Semi-transparent surface representation of the two closest-to-mean structures in which the buried Trp and Leu residues are highlighted. (A) FH-10-11 in which ^{616}Leu and the consensus Trp are indicated. (B) FH-11-12 in which the two consensus Trp residues are indicated.

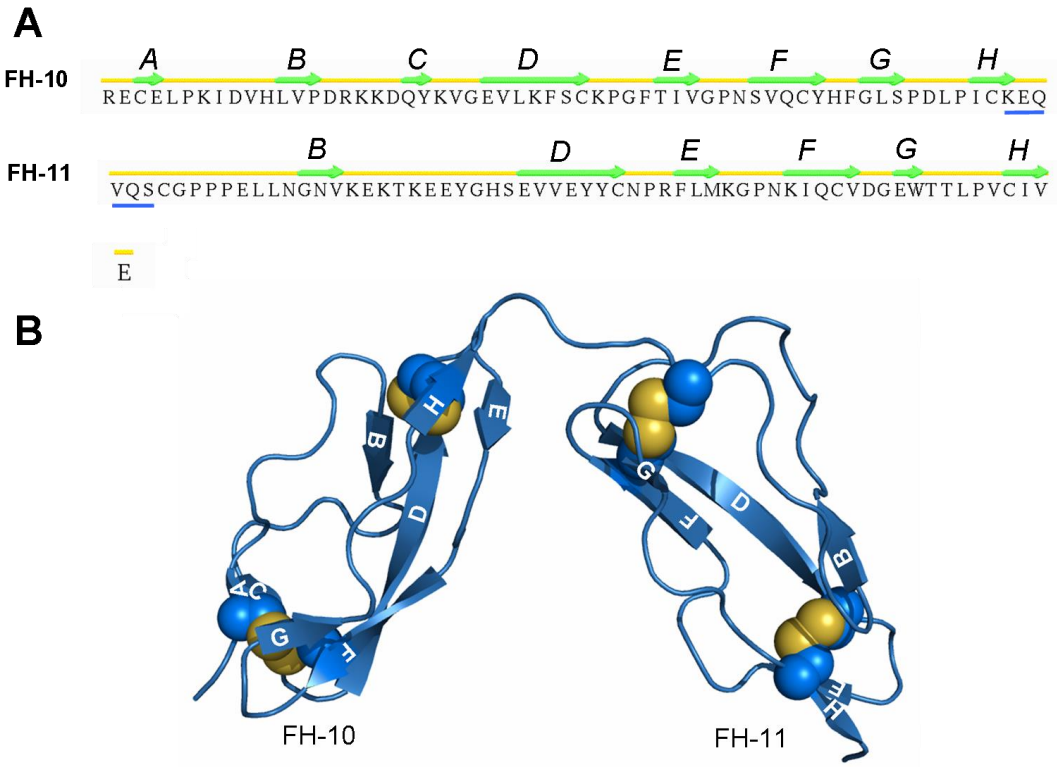


Figure 4.10 Representations of the secondary structural elements within the closest-to-mean member of the ensemble of NMR-derived FH-10-11 structures. (A) Visual illustration of the STRIDE (138) output in which the β-strands are shown labeled. The linker residues are underlined by a blue line. (B) Cartoon of the closest-to-mean structure in which the β-strands calculated using STRIDE are indicated. The strands are labeled according to the convention based on the occurrence of a maximum of eight strands (A-H) (144) in any given CCP. Disulfide bonds are highlighted by sphere representations of the β-carbon and sulfur atoms.

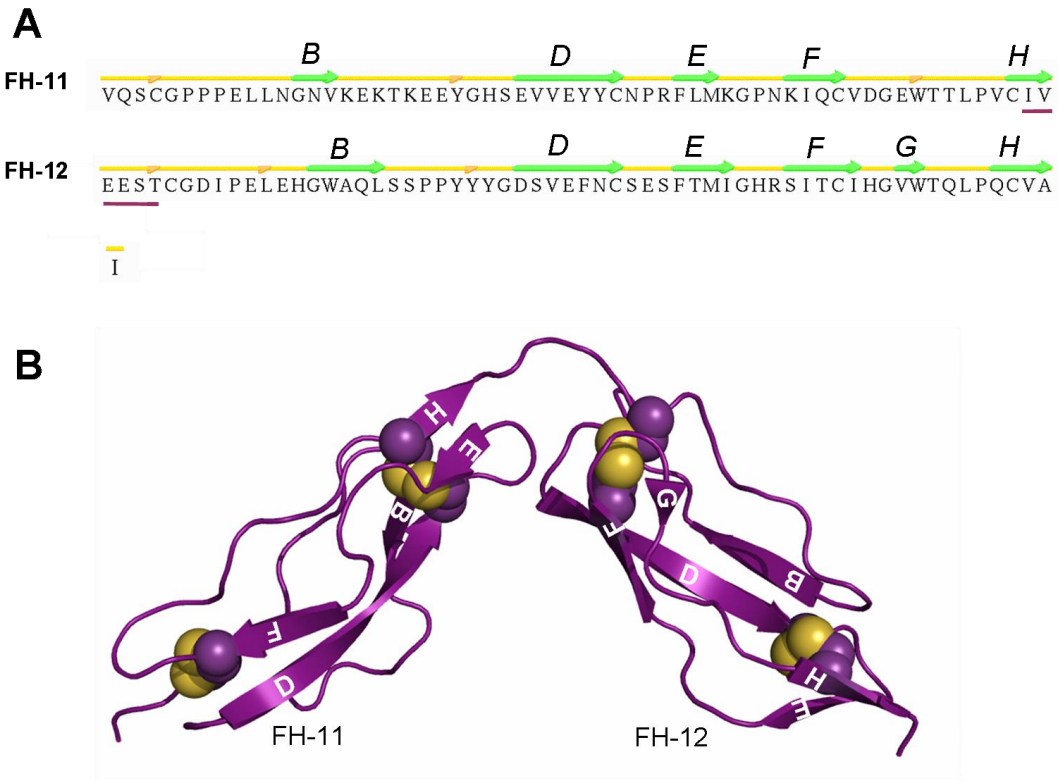


Figure 4.11 Representations of the secondary structural elements within the closest-to-mean member of the ensemble of NMR-derived FH-11-12 structures. See legend to Figure 4.10

4.1.3 Identification of slowly exchanging amide protons

Samples of both FH-10-11 and FH-11-12 were lyophilized, resuspended in 99.96% D_2O and subjected to a time-series of ^{1}H, ^{15}N-HSQC experiments. Spectra were recorded in 15-min intervals.

Exchange of N-H for N-D atoms results in a loss of cross-peaks in the HSQC spectrum, and occurs rapidly for solvent-exposed protons. On the other hand, N-H protons that are buried within the protein – and are therefore normally participants in hydrogen-bonds - exchange only slowly with solvent. Figures 4.12 and 4.13 show ^{1}H, ^{15}N-HSQC spectra for FH-10-11 and FH-11-12 after a 15-min exposure to D_2O.

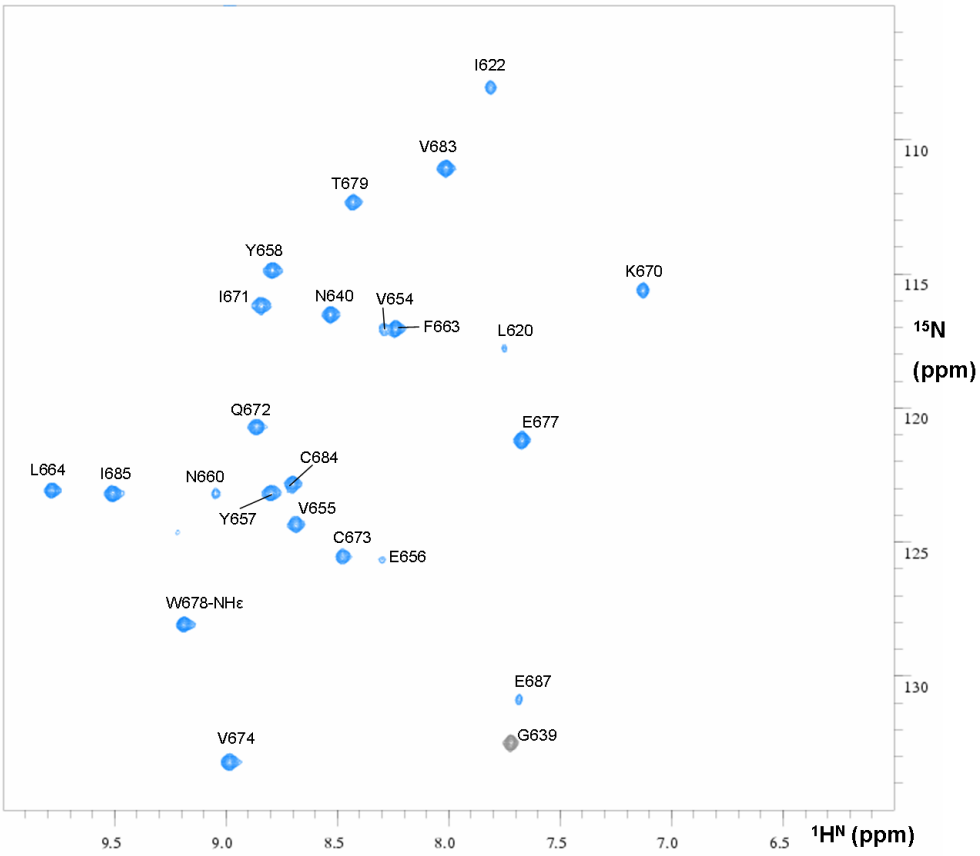


Figure 4.12 ^{1}H, ^{15}N-HSQC spectrum for FH-10-11, after 15-min exposure to D_2O.

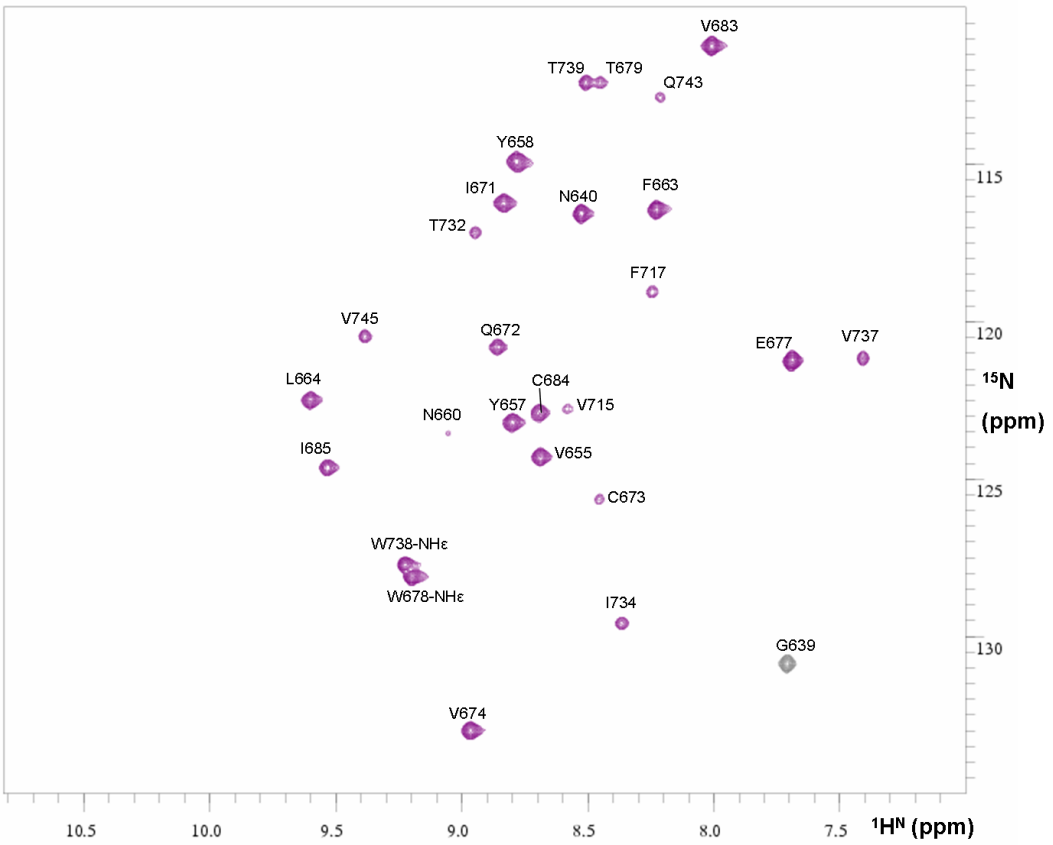


Figure 4.13 ^{1}H, ^{15}N-HSQC spectrum for FH-11-12, after 15-min exposure to D_2O.

In the case of FH-10-11, 24 cross-peaks persist in the HSQC spectrum, two from FH-10, none from the 10-11 linker region, and 22 from FH-11. The relevant slowly exchanging protons are highlighted on the structure of FH-10-11 in figure 4.14. Of these 24 protons, 21 participate in the formation of hydrogen bonds, 18 of which are within β-sheets (as recognized by STRIDE) while ^{660}Asn H^N (which lies immediately after FH-11 β-strand D) hydrogen bonds with the carbonyl of residue ^{638}Asn (which lies just before FH-11 β-strand B), ^{679}Thr H^N is hydrogen-bonded to the carbonyl of β-strand residue ^{672}Gln, and ^{687}Glu H^N (which lies just after FH-11 β-strand H) hydrogen-bonds with the carbonyl of ^{662}Arg (which lies immediately prior to FH-11 β-strand E). Of the non-hydrogen bonded protons, the Hε of the consensus Trp residue in FH-11, ^{678}Trp is buried, while ^{620}Leu H^N exhibits a weak signal that disappears in spectra recorded at later times. The only anomaly is ^{683}Val H^N in FH-11 that does not appear to be within hydrogen-bonding distance of any acceptor atom yet does not exchange even after 60 min. Its amide in all 20 conformers of the

ensemble is not exposed, and points towards the interior of the structure, thus being rather inaccessible.

It is interesting that only two residues retain their amide proton in FH-10 whereas the remaining 22 slowly exchanging amides lie in FH-11. FH-10 appears to be more dynamic. This is despite the fact that this module has the most extensive formation of β-sheets (as detected by STRIDE). It must be concluded that modules 10 and 11 have quite distinct dynamic properties even though they form a well-defined inter-modular interface. The absence of the consensus Trp may contribute to the potential lack of rigidity in FH-10. Note that the presence of FH-9 may enhance the stability of FH-10 in the context of full-length FH.

In the case of FH-11-12, a total of 27 cross-peaks remain detectable after 15 min of exposure to D_2O. The relevant protons are highlighted on the structure of FH-11-12 in figure 4.15. Of these, 17 belong to FH-11, one to the linker region and 9 to FH-12. Two of these correspond to the H^{ε} of the consensus Trp residues (^{678}Trp in FH-11 and ^{738}Trp in FH-12) and 24 correspond to H-bonded backbone amide protons. Of these 24, all lie in β-sheets except: FH-11's ^{660}Asn H^N and ^{679}Thr H^N (as was observed in FH-10-11); ^{739}Thr H^N of FH-12 lying just after β-strand G which forms a hydrogen bond with the carbonyl of ^{732}Thr (part of CCP 12 β-strand F); ^{674}Val H^N (lying immediately after CCP 11 β-strand F), which hydrogen bonds with the carbonyl of ^{677}Glu; and ^{677}Glu H^N, which presumably hydrogen bonds with the carbonyl of ^{674}Val although in the closest-to-mean structure it appears to not be positioned at an optimum geometry for this to take place. However, according to STRIDE, in 11 conformers out of the 20 of the ensemble ^{674}Val and ^{677}Glu both belong to CCP 11 β-strands F and what would be β-strand G. As in FH-10-11, slowly exchanging ^{683}Val H^N is not within hydrogen bonding distance of any acceptor however it exhibits similar characteristics as described in the case of FH-10-11.

The slowly exchanging protons of FH-11 in the two contexts (*i. e.* in FH-10-11 and FH-11-12) largely coincide. A total of 17 amides - out of 22 slowly-exchanging FH-11 amides in FH-10-11 and 18 slowly-exchanging FH-11 amides in FH-11-12- agree. This conservation of dynamics (on the time-scale probed by the observation of slowly-exchanging amides) observed between the two constructs is consistent with the similarity of structures (see below).

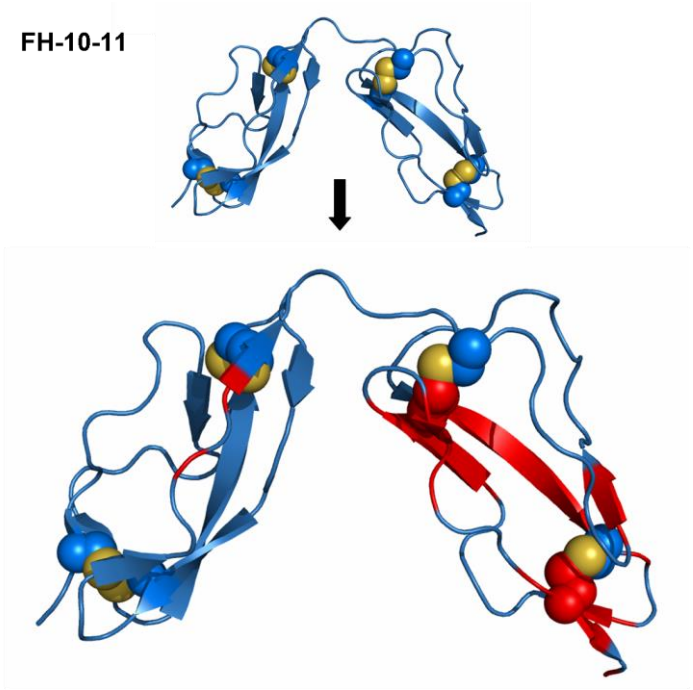


Figure 4.14 Representation of the closest-to-mean structure of FH-10-11 in which beta strands are shown, and residues that retained their amide H^N atom in the hydrogen-deuterium exchange experiment are highlighted in red.

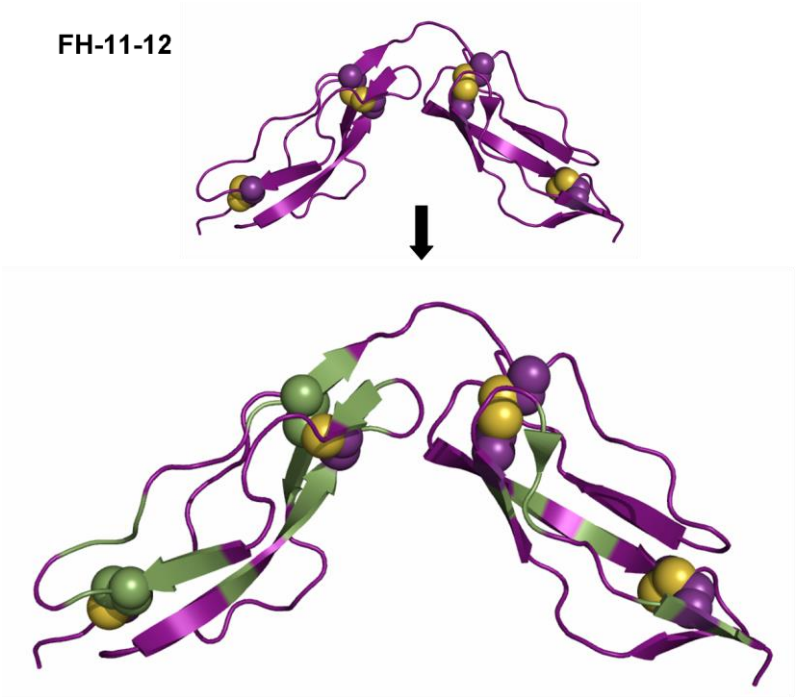


Figure 4.15 Representation of the closest-to-mean structure of FH-11-12 in which beta strands are shown, and residues that retained their amide H^N atom in the hydrogen-deuterium experiment are highlighted in green.

4.1.4 Assessment of the structures of FH-10-11 and FH-11-12

PROCHECK assesses the stereochemical quality of a protein structure. Amongst other things, it allows evaluation of the extent to which the *phi* and *psi* angles occupy energetically favored regions in the Ramachandran plot (145). After refinement using explicit water, PROCHECK was utilized to analyse the experimentally derived structures of FH-10-11 and FH-11-12 (Figs 4.16 and 4.17).

For FH-10-11, 80.5 % of residues are located in the most favored regions and 16.4 % in the additionally allowed regions. For FH-11-12, 74.6 % lie in the most favored region while 22.4 % in the additionally allowed regions. Very few residues lie in the disallowed regions (1.8 % for FH-10-11 and 1.3 % for FH-11-12) and overall these values are comparable with those obtained for other published NMR-derived structures of CCP modules (27), (39), (43).

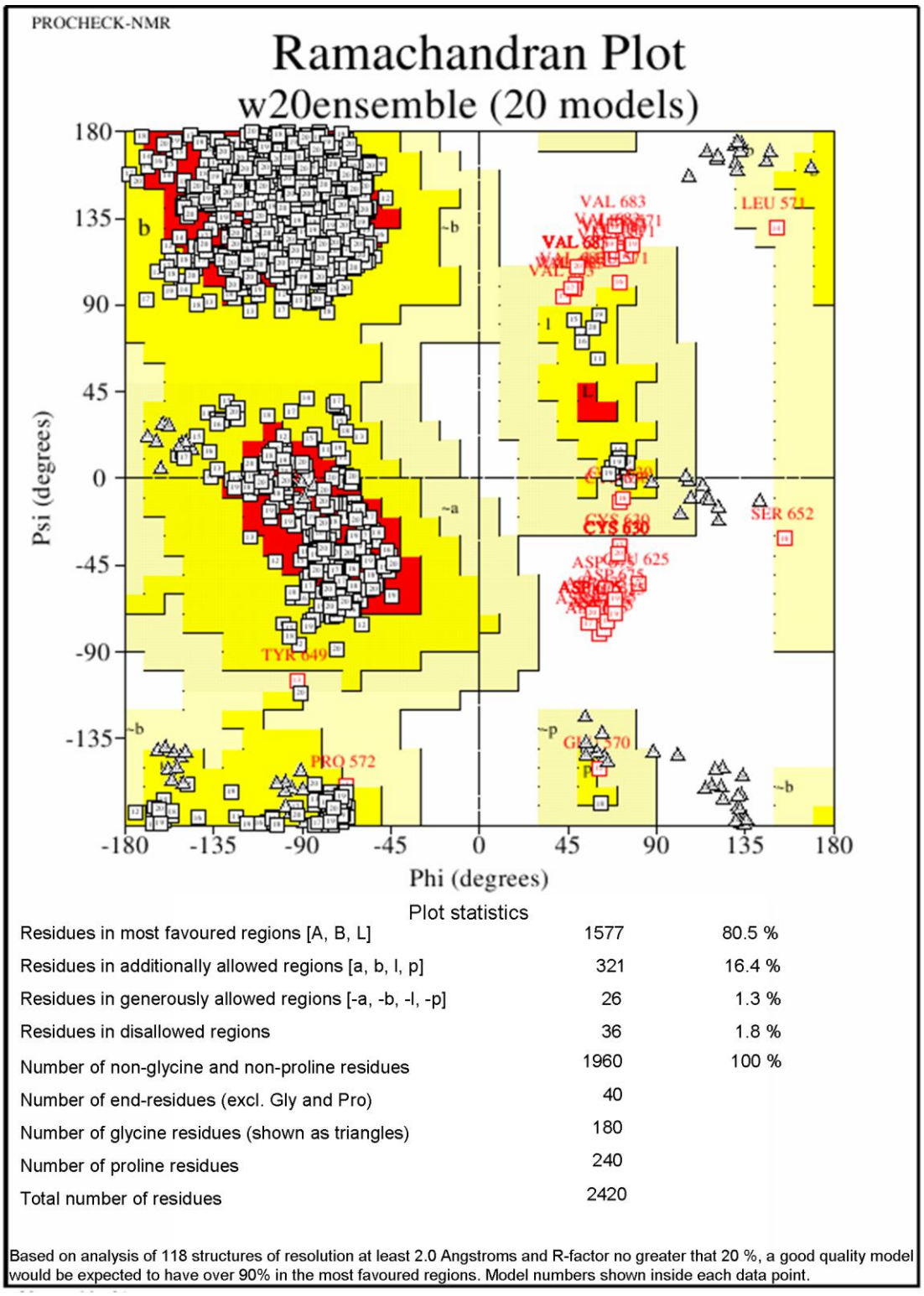


Figure 4.16 Ramachandran statistics for the 20 lowest-energy water-refined structures of FH-10-11.

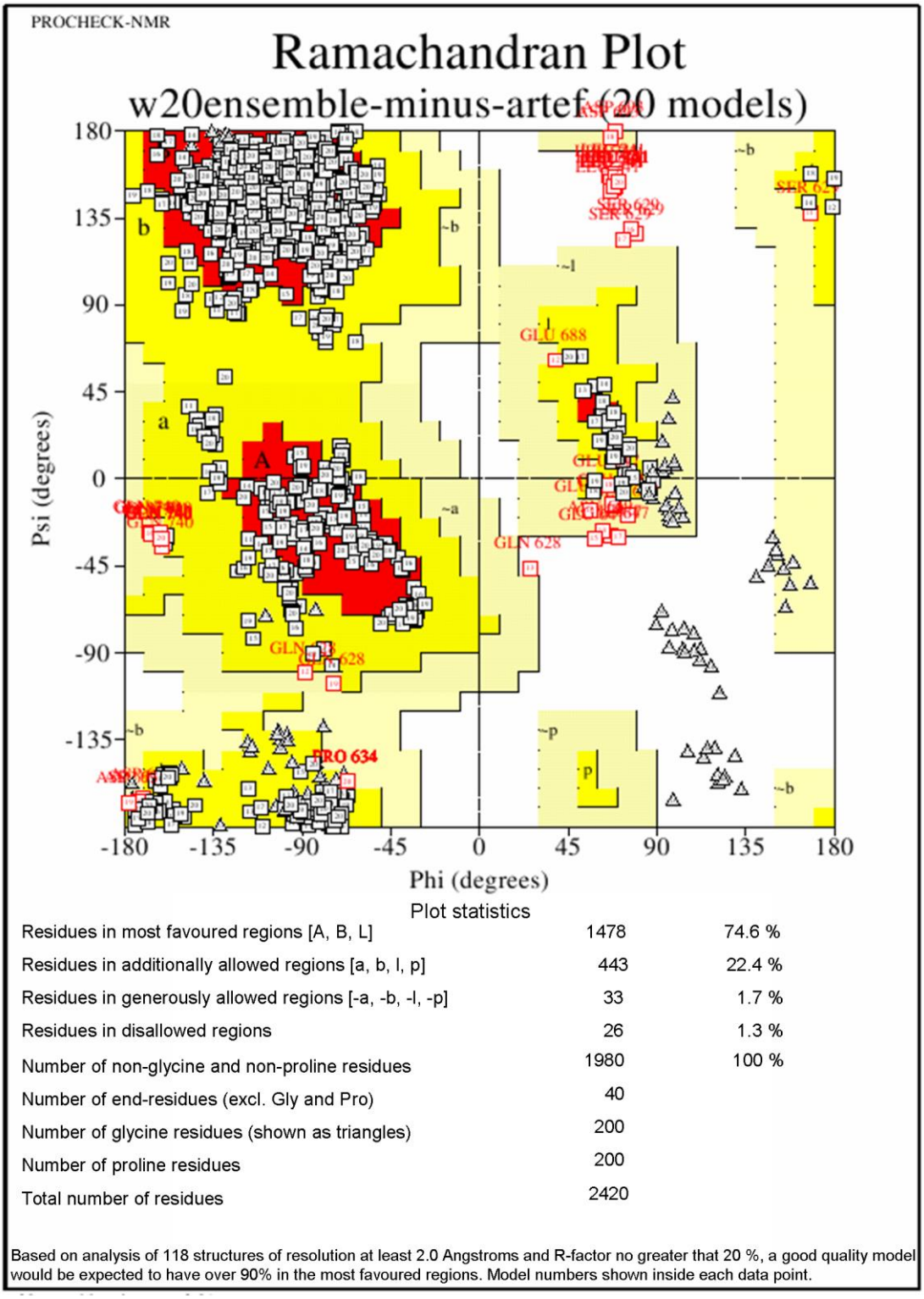


Figure 4.17 Ramachandran statistics for the 20 lowest-energy water-refined structures of FH-11-12.

4.1.5 FH-10-11 and FH-11-12 form bent structures

In FH-10-11, FH-10 is connected to FH-11 through a six-residue linker, where a linker is defined as the sequence of residues lying between Cys IV of the N-terminal CCP and Cys I of the C-terminal CCP (in fact one of the linker residues lies within CCP 10 β-strand H and therefore could be considered part of FH-10). A ~118^{o} tilt is present between the modules (according to the convention in which zero degrees corresponds to a fully extended, linear structure), a ~57^{o} twist and a ~89^{o} skew. The linker between (the cysteines of) FH-11 and 12 is also six-residues long - with two residues actually forming part of FH-11 strand H - and a ~90^{o} tilt angle occurs between these modules. The twist angle is ~72^{o} while the skew angle is ~57^{o} (see also figure 4.18 and Appendix section C2).

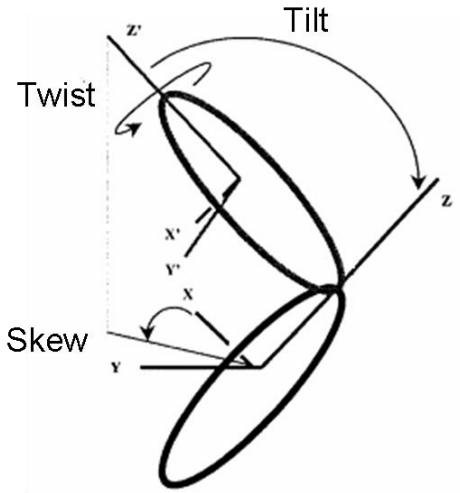


Figure 4.18 Schematic explanation of the intermodular angles tilt, twist and skew between CCP modules (Figure adapted from (150)). Tilt, twist and skew angles in degrees were determined using (for a reference *x*-axis) a vector between the principal axis of the inertia tensor (the *z*-axis) and the α-carbons of Leu616 in FH-10 and the conserved Trp678 in FH-11 or Trp738 in FH-12, respectively. Module boundaries were defined as CysI-CysIV (Cys569-Cys684 in the case of FH-10-11 and Cys630-Cys744 in the case of FH-11-12).

For calculations of surface areas buried between modules, module boundaries were defined as follows: one residue before Cys ^{I} and three residues after Cys ^{IV} for the first CCPs of the bimodules (*i.e.* ^{568}Glu-^{626}Gln for FH-10 and ^{629}Ser-^{687}Glu for FH-11) and three amino acids before Cys ^{I} and one amino acid after Cys ^{IV} for the following CCPs (*i.e.* ^{627}Val-^{685}Ile for FH-11 and ^{688}Glu-^{745}Val for FH-12). This

produced values of ~360 Å^2 for FH-10-11 and ~ ~640 Å^2 for FH-11-12. Thus (see Fig. 4.19) even though FH-10 and FH-11 have a greater tilt angle, significantly less surface area is buried compared to the FH-11-FH-12 interface. In the latter case an extended loop (between strands D and E) from FH-11 contacts the linker as well as residues from FH-12.

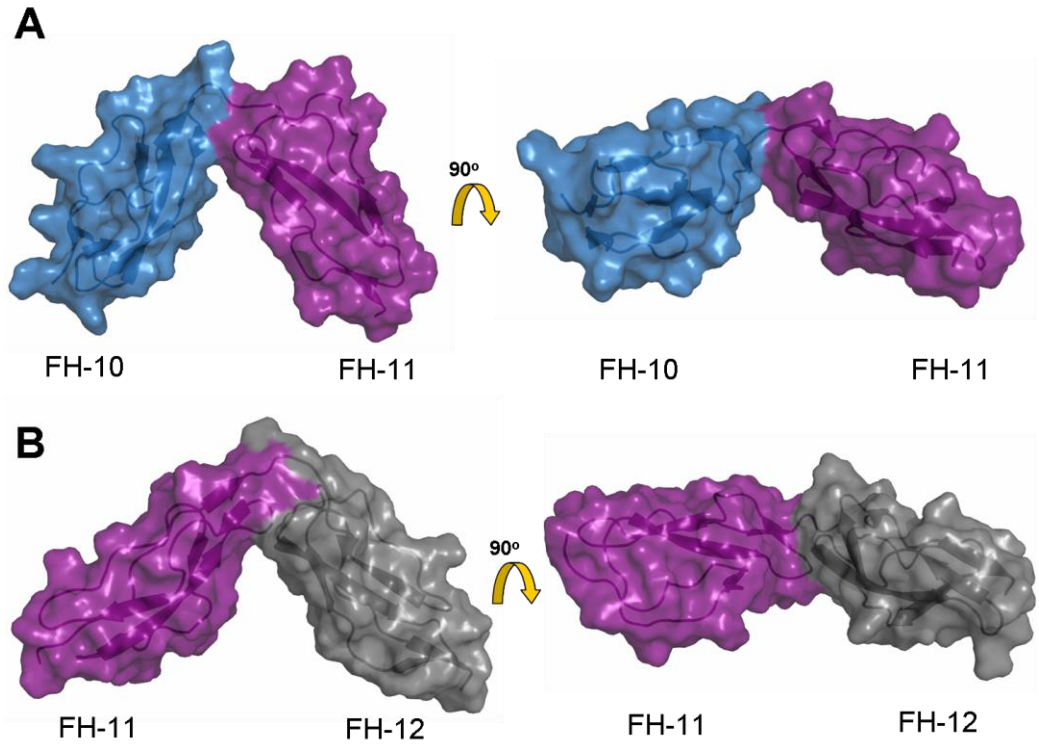


Figure 4.19 Surface representations of the structures of FH-10-11 and FH-11-12. (A) Surface representation of FH-10-11 in two views rotated by 90° about the x-axis. (B) Surface representation of FH-11-12 in two views rotated by 90° about the x -axis.

The table below (Table 4.1) summarizes the statistics for the 20 lowest-energy structures of FH-10-11 and FH-11-12.

	FH-10-11	FH-11-12
Number of structures in ensemble	20	20
Numbers of upper-limit distance constraints		
Intraresidue, |*i-j*| <=1	1405	1482
Medium range, 1< |*i-j*| < 5	411	301
Long range, |*i-j*|>=5	1391	1322
Total	3207	3105
Intermodular	8	38
Module i to linker	62	89
Module j to linker	65	39
RMSD[a]		
Bimodule (i+j)		
All atoms	1.17	1.05
Backbone atoms	0.91	0.84
Module i		
All atoms	0.79	0.73
Backbone atoms	0.50	0.33
Module j		
All atoms	0.74	0.72
Backbone atoms	0.33	0.43
Intermodular angles {minimum-maximum [mean (S.D.)]}		
Skew (^{o})	75-107 [89(7)]	149-170 [157(4)]
Twist (^{o})	40-78 [57(10)]	59-87 [72(7)]
Tilt (^{o})	111-128 [118(5)]	76-99 [88(6)]
Ramachandran assessment		
Most favoured (%)	80.5	74.6
Additionally allowed (%)	16.4	22.4
Generously allowed (%)	1.3	1.7
Disallowed (%)	1.8	1.3
Surface area buried between modules (Å^2)[b]	358	642

[a] Residues from Cys^{I} to Cys^{IV} of each CCP module
[b] Closest-to-mean structures/ Calculated by breaking the bimodules at the midpoint of the linker

Table 4.1 Statistics and characteristics for the lowest-energy structures of FH-10-11 and FH-11-12.

4.1.6 Description of the intermodular interfaces

As already mentioned, the linkers in the central region of FH (six residues between both FH-10 and 11, and between FH-11 and 12) are longer than average. Before the current study, various hypotheses had been suggested for the structural ramifications of these longer linkers. Intuitively it seemed likely, for example, that a long linker would allow flexibility or, in contrast, that a long linker would allow flanking CCPs to fold back against each other.

The resulting structures interestingly revealed that both FH-10-11 and FH-11-12 show defined, strongly bent conformations, as supported by a multitude of NOE cross-peaks allowing upper-distance restraints to be applied between modules and linkers or directly from one CCP to its neighbour. In the case of FH-10-11 there are eight intermodular NOEs, 62 NOEs from FH-10 to the linker and 65 NOEs from the linker to FH-11. For FH-11-12 there are 38 intermodular NOEs, 89 NOEs between FH-11 and the linker and 39 NOEs between the linker and FH-12. Thus far from introducing flexibility the longer linkers appear to rigidly associate the two CCPs that they connect. (Inter-modular flexibility will be discussed further in Chapter 5).

The linking sequence in FH-10-11 consists of the residues ^{624}KEQVQS^{629}. Residues ^{602}Thr and ^{604}Val from FH-10 (from β-strand E of CCP 10), and residues ^{651}His (from the FH-11 B-D loop) and ^{675}Asp (from the FH-11 F-G loop) are predominantly responsible for interacting with the linking sequence and also interact with each other. For example, NOEs were observed between atoms Hg2* of the side chain of ^{602}Thr of FH-10 and Hb atoms of ^{651}His of FH-11, and between ^{627}Val Hb atom of the linker and ^{651}His Hb atoms of FH-11. The overall result is a defined bend as well as a small hydrophobic core in which residues ^{602}Thr, ^{604}Val, ^{627}Val (from the linker and part of strand H), and ^{651}His bury their sidechains. The β- and ε-methylenes of ^{624}Lys (linker) also contribute to this core. None of these residues, however, exhibit any slowly exchanging amides. This further validates the interface. The sidechains of linker residues ^{625}Glu, ^{626}Gln, ^{628}Gln and ^{629}Ser, on the other hand, project on the outside of the inter-modular bend (see Figs. 4.21, 4.23) and are exposed to the solvent. A putative salt bridge exists between the first residue of the

linker (^{624}Lys, which is also the last residue of FH-10 β-strand H) and ^{675}Asp of the F-G β-turn of FH-11 (Fig. 4.20).

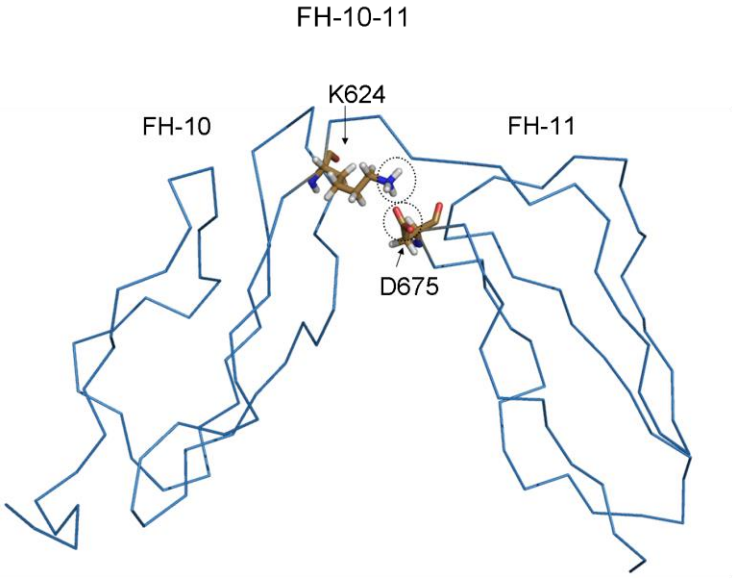


Figure 4.20 Ribbon representation of the lowest energy conformer of FH-10-11; K624 of the linker presumably forms a salt bridge with D675 of the F-G β-turn of FH-11 and these residues are highlighted.

The six-residue FH-11-12 linking sequence is ^{685}IVEEST^{690}. The amide of ^{685}Ile of β-strand H of the linker also exhibits slowly exchanging properties according to the H-D experiment. A notable feature of the FH-11-12 interface is the important role played by aromatic sidechains (Fig. 4.22). Side-chains of the D-E loop and β-strand E of FH-11 (^{662}Arg, ^{663}Phe and ^{664}Leu) make extensive van der Waals interactions with alkyl portions of linker residues (^{685}Ile, ^{686}Val, while polar portions are displayed on the convex surface of the bend) and with a segment of the B-D loop of FH-12 (^{710}Tyr, ^{711}Tyr and ^{712}Gly).

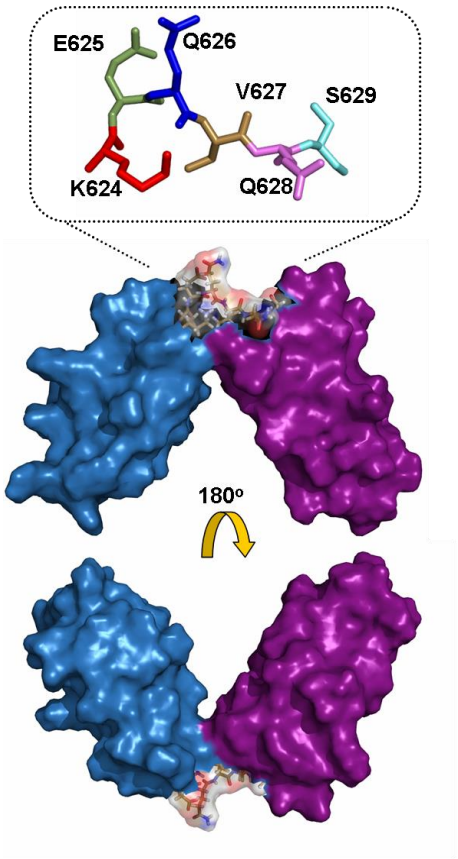


Figure 4.21 Stick and surface representation of the linking sequence for FH-10-11. Linker residues are labeled (in upper panel), and are shown by partial transparency in the context of the whole molecule in the surface display.

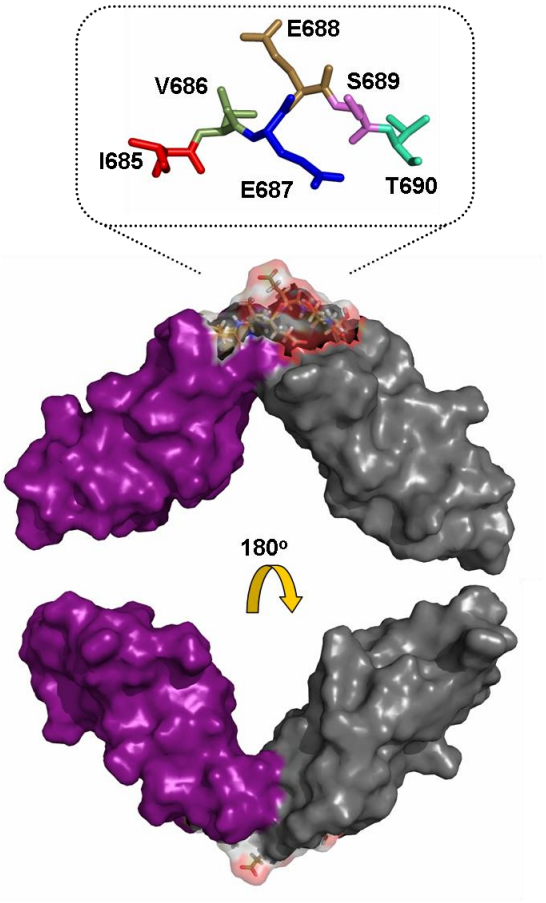


Figure 4.22 Stick and surface representation of the linking sequence of FH-11-12. The corresponding residues are labeled when shown on their own on the stick representation and are shown in partial transparency in the context of the whole molecule in the surface display.

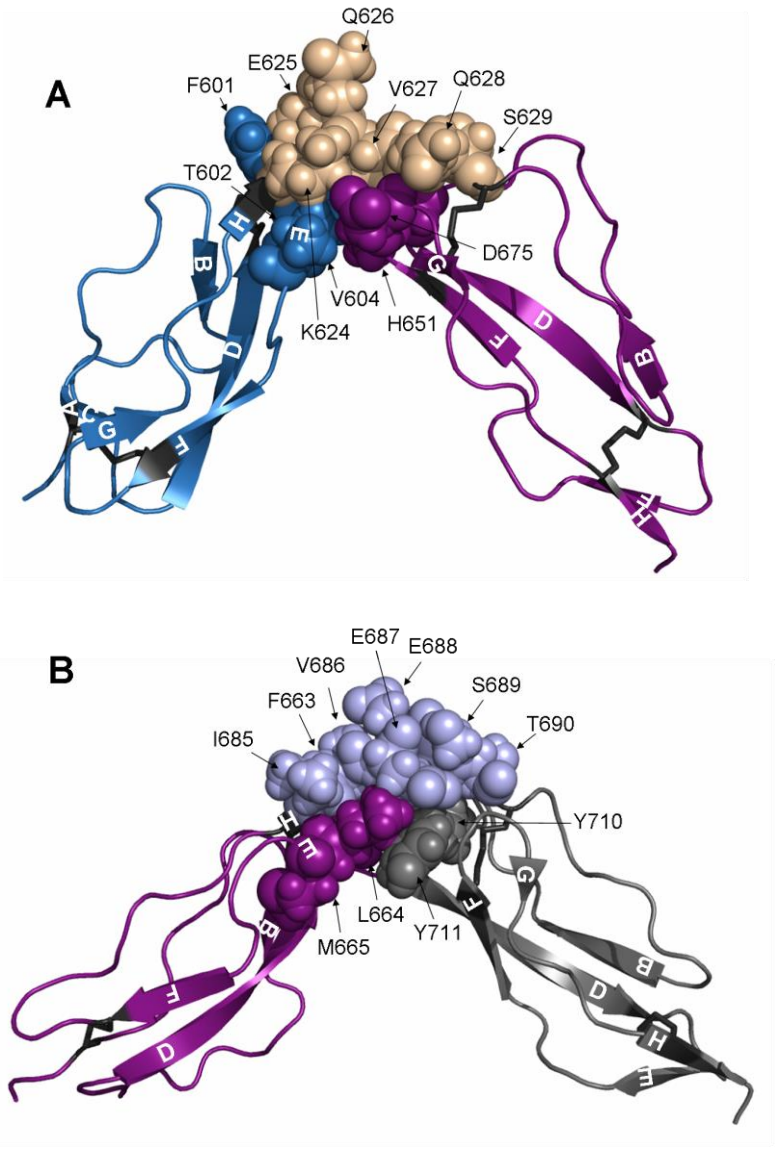


Figure 4.23 Representation of the two structures in which the residues that participate in the tight interfaces are represented as spheres and are labeled. (A) FH-10-11. Residues that belong to FH-10, the linker and FH-11 are shown in blue, beige and purple respectively. (B) FH-11-12. Residues from FH-11, the linker and FH-12 are shown in purple, light blue and grey respectively.

Figures 4.24 and 4.25 show examples of ^{13}C-NOESY data depicting long-range intermodular NOEs that define the relative orientations of the modules for FH-10-11 and FH-11-12.

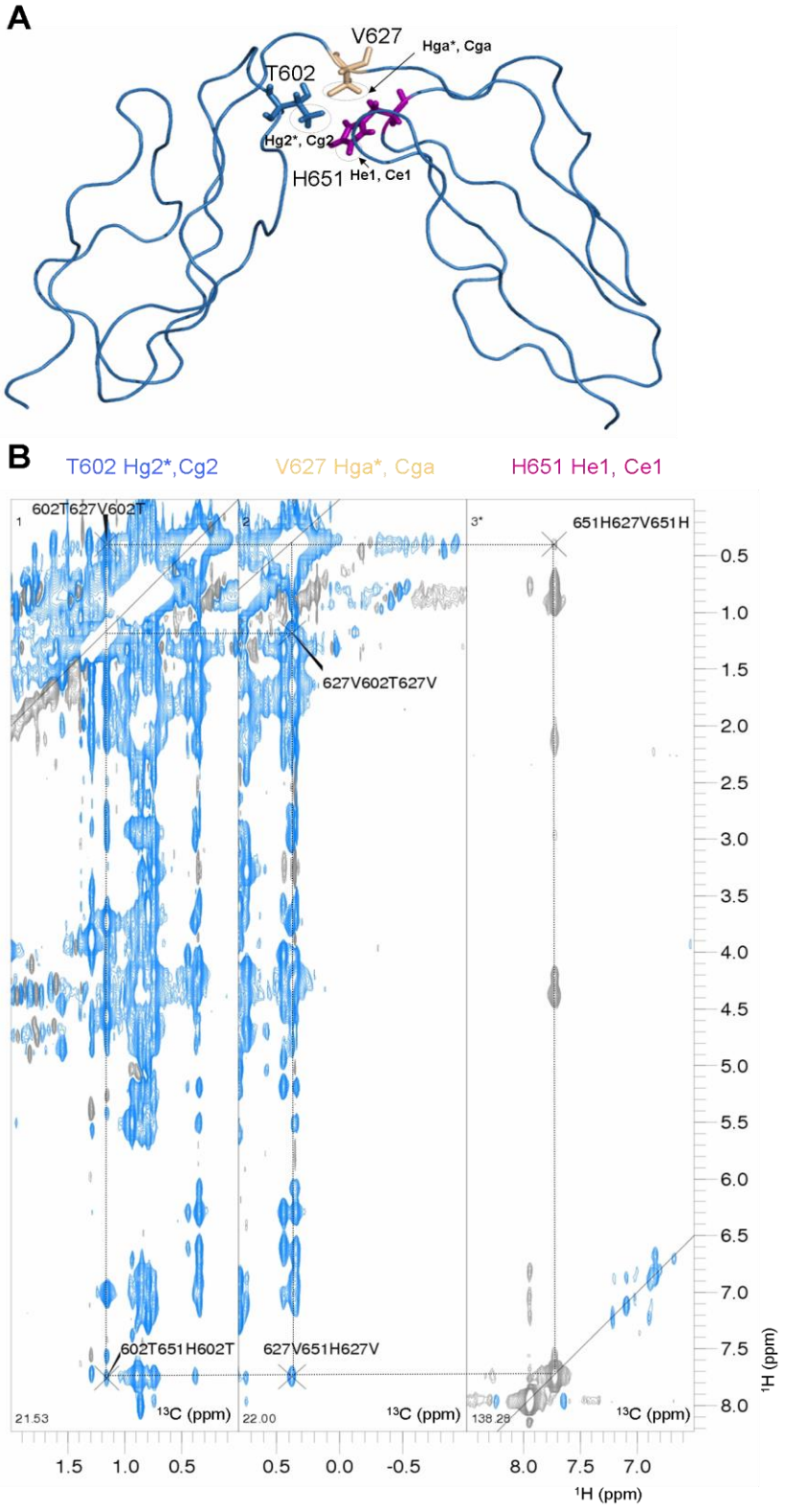


Figure 4.24 Example of intermodular NOEs for FH-10-11. (A) Representation of the structure in which three of the residues that participate in the tight interface, their sequence numbers and the atoms chosen for this demonstration are labeled. (B) ^{13}C-NOESY strips of atoms of the interface as indicated on top of each strip. These atoms correspond to the root resonance. Horizontal and vertical lines show the connectivities between NOE cross-peaks that arise between the atoms.

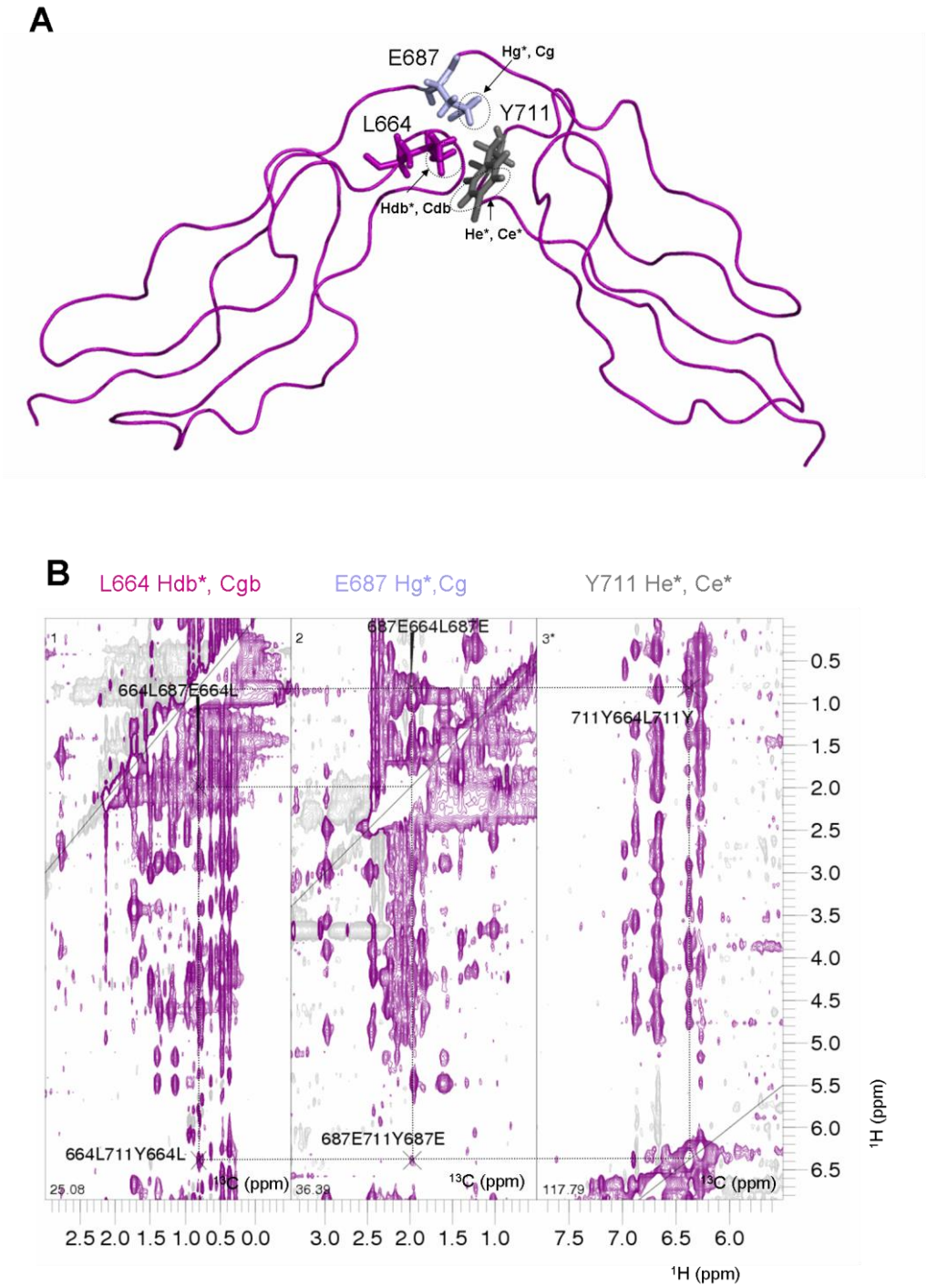


Figure 4.25 Example of intermodular NOEs for FH-11-12. See legend to figure 4.24 for details.

In conclusion, the smaller number of NOEs observed between FH-10 and 11 (compared to FH-11 and 12) is entirely consistent with the difference in the number of residues participating in the intermodular interfaces (and surface areas buried). The nature of the residues involved also contributes to the number of NOEs

observed. The NOE-based structure calculations suggest that both module pairs converge equally well. The large number of NOEs between modules and linker (for both pairs) is important in this respect. Overall, the network of over 100 NOEs contributing to the definition of intermodular orientations (in both pairs) supports the contention that these calculated structures are not over-restrained. Nonetheless, a reliance on NOEs alone might be unwise and hence further characterisation of these module pairs was carried out using SAXS (see Chapter 5)

4.1.7 Description of the FH-10-11 and FH-11-12 electrostatic and hydrophilic/hydrophobic surfaces

FH engages more efficiently with C3b bound to a self-surface when polyanions, such as GAGs are present, through the use of two binding sites that cooperate to recognize specific surface polyanions. The regions of FH primarily responsible for binding to polyanions are FH-6-8 and FH-19-20. Interactions with the multiple negatively charged groups of GAGs are probably favored by complementary electropositive surfaces on domains 6-8/19-20. Indeed, in a recent model of the FH:C3b complex, domains 6-8 and 19-20 are proximal to the hypothetical GAG-rich surface to which the target C3b would be tethered, and the central modules form a compact bent-back structure. It is however unclear where these central modules would reside, relative to other parts of the complex, nor if, or how, they might interact with a GAG- or sialic acid -coated surface. Thus it is interesting to consider the electrostatic surface of the new bi-module structures. Note, however, that although FH-13 in the previously solved NMR structure (43) has a striking electropositive patch on one face, it does not appear to interact with heparin.

Electrostatic analysis of the surfaces of the structures of FH-10-11 and FH-11-12 suggest it is in fact unlikely that the FH-10-12 region would contribute to favorable interactions with a polyanionic surface. One face of FH-10-11 is mainly negatively charged while the other face of the molecule exhibits a dispersion of positive and negative charge with a mini-electropositive patch (containing the sidechains of ^{582}Arg, ^{583}Lys, ^{584}Lys and ^{598}Lys) (see Fig. 4.26 B). Both faces of FH-

11-12 are predominantly negatively charged due mainly to plentiful Glu residues (Fig. 4.26 C).

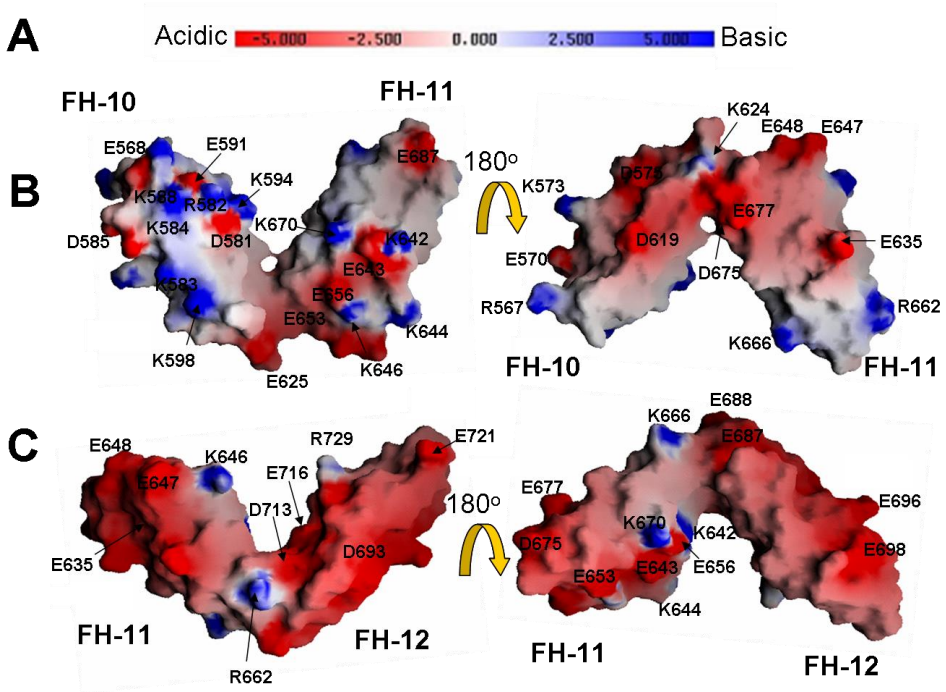


Figure 4.26 GRASP (139) electrostatic representations of the two structures (two views rotated by 180° about the x-axis). (A) Scale of electrostatic surface potential: Red = negative charge, blue = positive charge, range -5/+5 kT. (B) Representation of FH-10-11 in which charged residues are indicated on the surface. (C) Representation of FH-11-12 in which the residues responsible for the electrostatic properties are indicated on the surface.

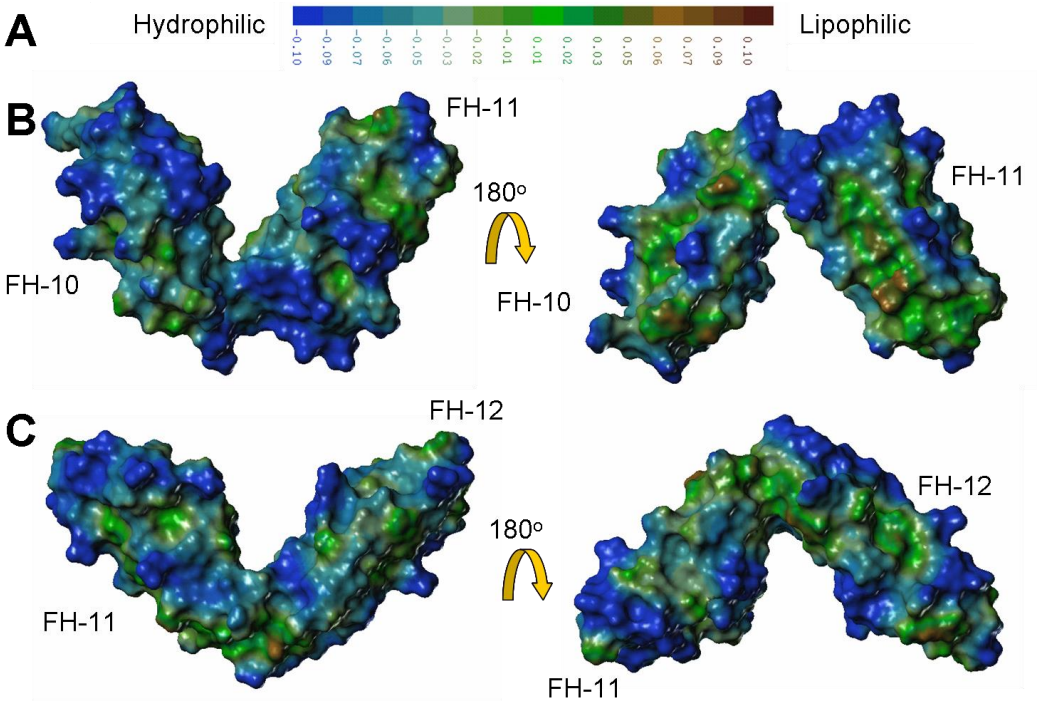


Figure 4.27 MOLCAD (140) lipophilic representation of the two CCP pairs (two views rotated by 180° about the *x*-axis in each case). (A) Hydrophilic to Hydrophobic scale. Blue corresponds to the most hydrophilic while brown to the most lipophilic. (B) Hydrophilic-lipophilic surface features for FH-10-11. (C) Hydrophilic-lipophilic surface features for FH-11-12.

According to the MOLCAD lipophilic surface representation of FH-10-11 and FH-11-12 (Fig. 4.27), there appears to be no prominent lipophilic patches on the faces of the two module pairs.

4.1.8 Overlay of FH-10-11 and FH-11-12 with other CCPs

The closest-to-mean structures of FH-10, 11 and 12 were compared to a set of individual experimentally derived CCP structures from the complement system (see Appendix section C3 for summary of the extent of structural similarity of FH-10, 11 and 12 with one another and with other solved CCP module structures). The comparison was based upon alpha-carbon RMSD values using the structural alignment program Combinatorial Extension (143). The FH-10 structure does not overlay particularly well with structures of other FH CCPs (backbone RMSD > 2 Å in all cases), while, amongst known structures it is most similar to CCP 3 of MCP with which it shares only 20 % identity (with the exception of an inserted proline-rich loop following Cys-II in MCP the backbone overlay was 1.4 Å.) FH-11 and 12

on the other hand, are quite similar to one another (backbone RMSD ~1.6 Å) and overlay well with FH-18 and FH-19 (e.g. backbone RMSD is 1.1 Å for FH-11(-12) versus FH-19 which share 33 % sequence identity). FH-12 overlays very well with CR2 CCP2 (backbone RMSD is 1.4 Å and sequence identity is 44 %). FH-10, 11 and 12 all belong to cluster "B" according to Soares *et al.* (144). Figures 4.28-4.30 shown below illustrate the best overlaying partners in each case.

4.1.8.1 Overlay of FH-10 with CCP 3 of MCP

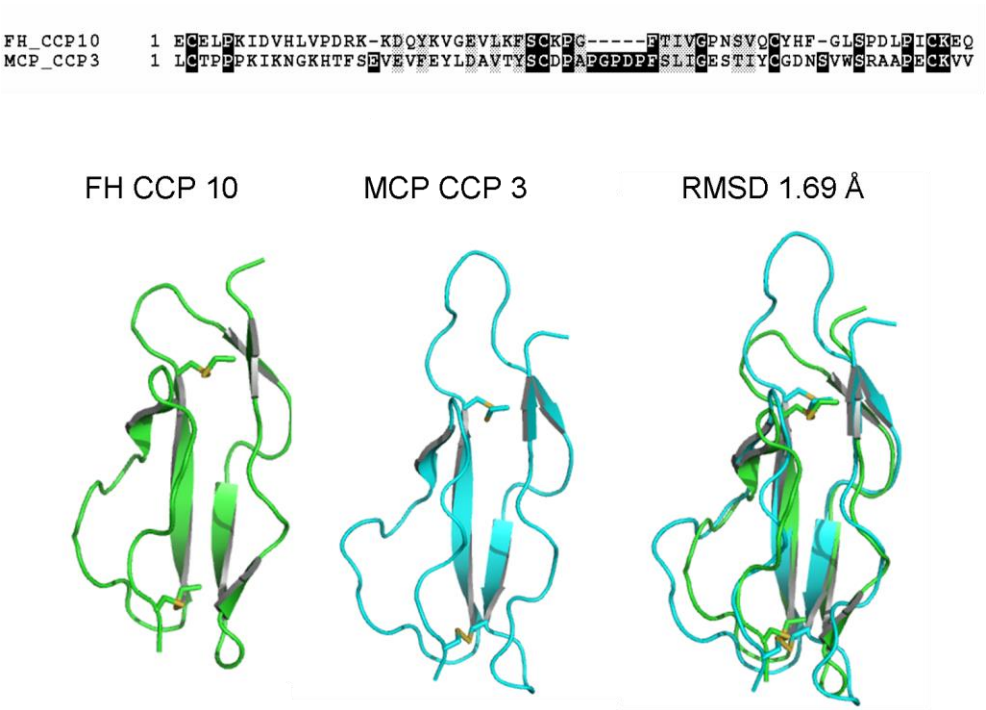


Figure 4.28 Upper panel: Sequences of FH-10 and CCP 3 of MCP in which identical residues are highlighted and similar residues shaded in grey. Lower panel: structures in cartoon representations of the individual CCPs as well as an overlay of the two. Disulfide-bonded Cys residues within each CCP are shown with sticks.

4.1.8.2 Overlay of FH-11a and 11b with FH-19

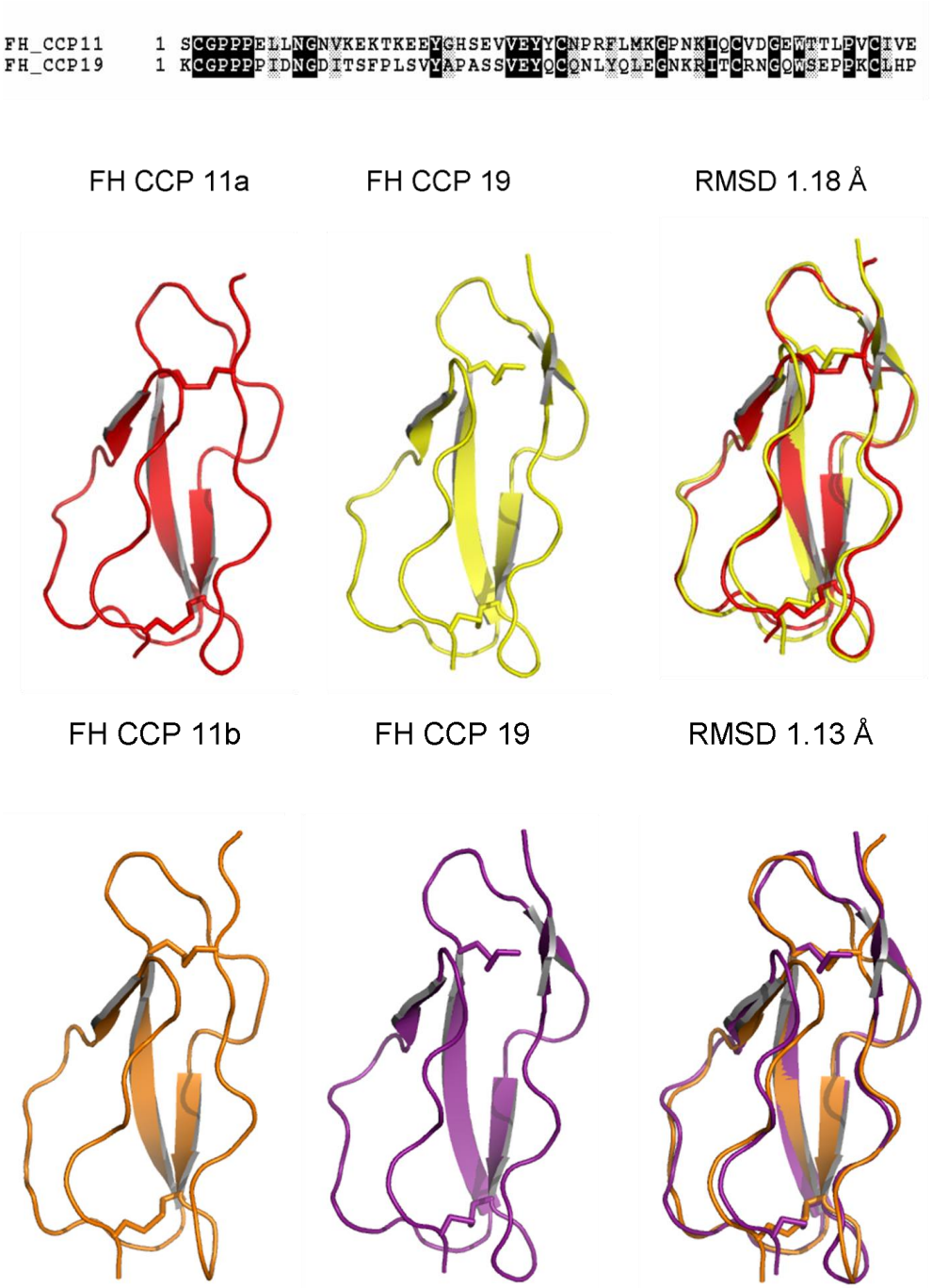


Figure 4.29 Upper panel: Sequences of FH-11 and FH-19 in which identical residues are highlighted and similar residues shaded in grey. Lower panel: structures in cartoon representations of the individual CCPs as well as an overlay of the two. Disulfide-bonded Cys residues within each CCP are shown with sticks.

4.1.8.3 Overlay of FH-12 with CCP 2 of CR2

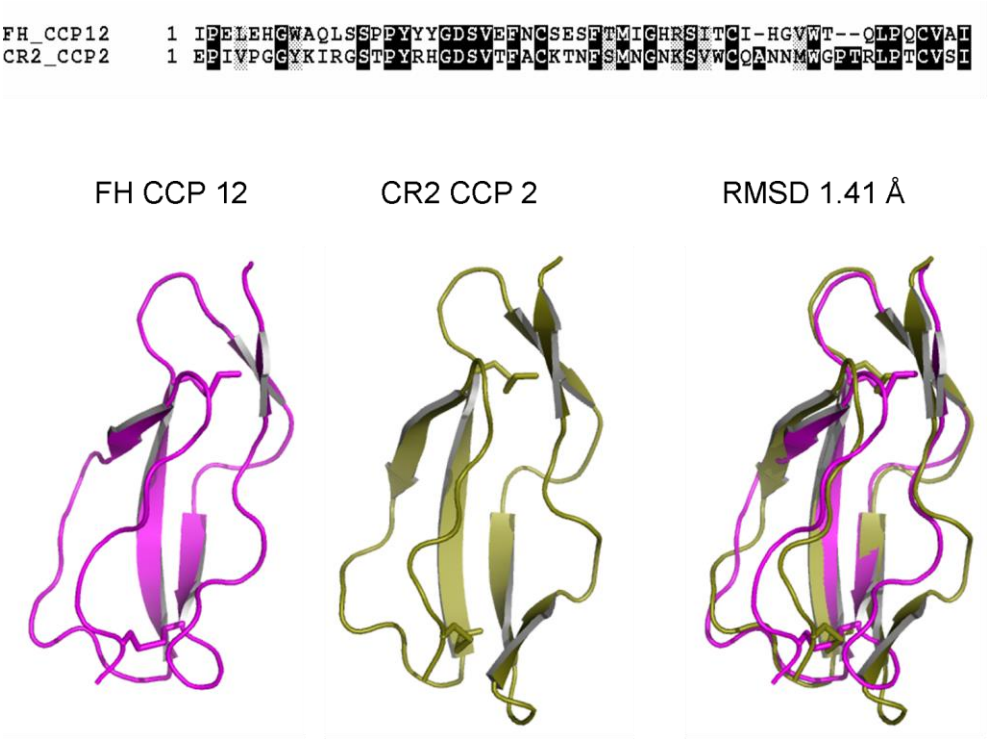


Figure 4.30 Upper panel: Sequences of FH-12 and CCP 2 of CR2 in which identical residues are highlighted and similar residues shaded in grey. Lower panel: structures in cartoon representations of the individual CCPs as well as an overlay of the two. Disulfide-bonded Cys residues within each CCP are shown with sticks.

4.1.8.4 Overlay of FH-11a of FH-10-11 with FH-11b of FH-11-12 and FH-12a of FH-11-12 with FH-12b of FH-12-13

The figures below (Figs. 4.31 and 4.32) show the backbone overlays of the closest-to-mean structures of FH-10-11, FH-11-12 and FH-12-13 (43) on their mutual CCP modules.

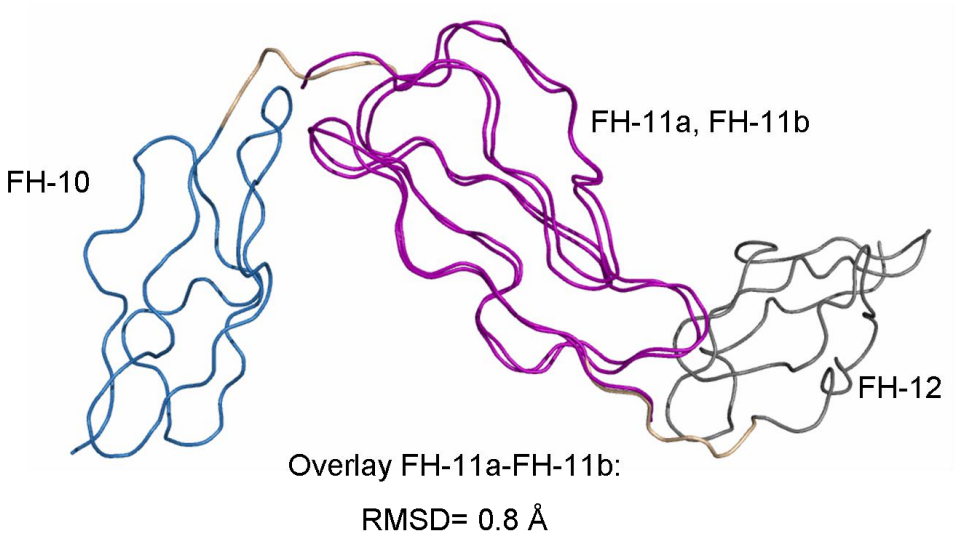


Figure 4.31 Overlay of the closest-to-mean structures of FH-10-11 and FH-11-12 in ribbon representation. The linkers are shown in beige.

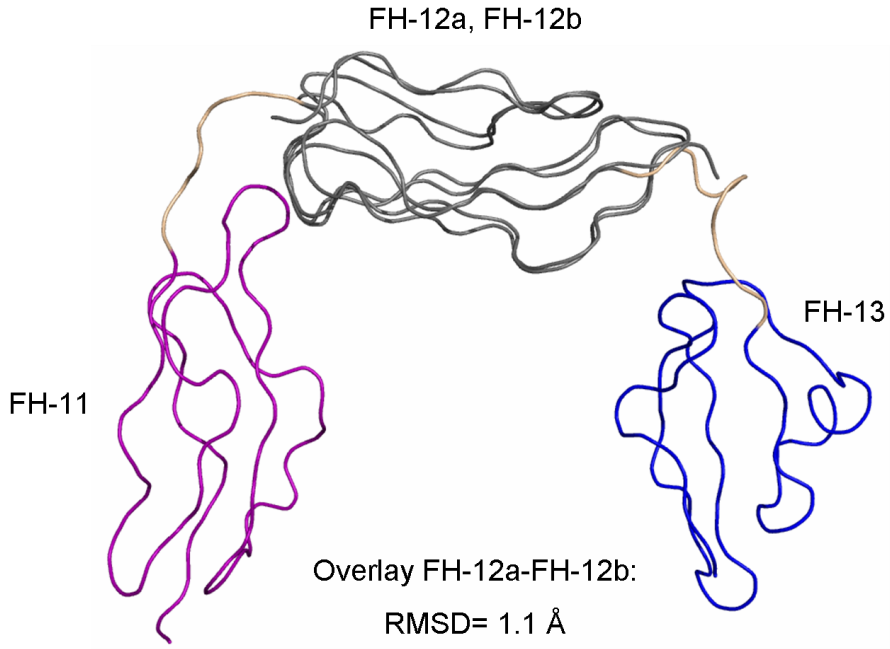


Figure 4.32 Overlay of the closest-to-mean structures of FH-11-12 and FH-12-13 in ribbon representation. The linkers are shown in beige.

FH-11 of FH-10-11 overlays well with FH-11 of FH-11-12, with an RMSD value of 0.8 Å (*i.e.* when both structures are superimposed on the mutual CCP module 11). Similarly, FH-12 of FH-11-12 overlays relatively well with FH-12 of FH-12-13 with an RMSD value of 1.1 Å.

These models – that assume non-neighbouring CCP modules have no influence on remote intermodular junctions - were validated against SAXS data (see below).

4.2 Conclusions

The central region of FH has remained, to date, little-explored territory in terms of structural biology, with the exception of a study combining the high-resolution structure of FH-12-13 and SAXS of FH-11-14 and FH-10-15. This study however was unable to distinguish between two compact models of FH-10-15 (43).

The current work describes new high-resolution structural information obtained on the central region of FH. Both module pairs are structurally similar to FH-12-13 (which has a tilt of ~80^{o}), but different than FH-1-2 (which has a tilt of ~23^{o} (39) FH-2-3 (tilt of ~15^{o} (39)) and FH-19-20 (~11^{o} (27)).

Longer linkers were initially considered a likely source of flexibility in full-length FH (43). With this in mind, it is worth mentioning that the structure of CCPs 1-2 of complement receptor 2, which are connected through an eight-residue long linker, exhibited a side-by-side conformation (intermodular tilt angle ~142^{o}) (151). Another example of a similar arrangement was observed for CCPs 2-3 of factor B, which are joined by a six-residue long linker (tilt ~189^{o}) (152). Finally, the six-residue linker that connects CCP 2 to CCP 3 of C2b (C2b is important for C4b binding in C3 convertase formation of the classical pathway) promotes a side-by-side arrangement (tilt ~218^{o}) (153).

Thus, based now on three examples, longer linkers can - in addition to promoting side-by-side structures and providing the potential for greater flexibility - result in specifically oriented structures (like the ones presented in this work). These could likely contribute to the biological diversity and specificity.

The high-resolution structural information for FH-10-11 revealed the presence of a His (^{651}His in module 11) sidechain in the miniature hydrophobic core at the interface between FH-10-11. Interestingly, FH-12-13 also has a His (^{773}His in module 13) residue at an equivalent position in space, and in the sequence (Figs. 4.33 and 4.34). In general the pKa for His sidechains is ~6; however in the cases of FH-10-11 and FH-12-13 this was not measured. Additional similarities in these two

interfaces were observed. In particular, the His residues of FH-11 and 13 are closely packed with Thr residues (^{602}Thr and ^{724}Thr) from FH-10, or FH-12, respectively. Moreover ^{627}Val from the linker in FH-10-11 and ^{747}Ile from the linker in FH-12-13 make similar contributions to their respective interfaces. Finally, in the NMR-derived structure of FH-15-16 a His (^{878}His, although in this case it is located on the other end of the double-module structure compared to FH-10-11 and FH-12-13) residue is also buried in an interface that involves hydrophobic residues (Figs. 4.33 and 4.34).

With these interfacial His residues in mind, it may be significant that a low-resolution structural study, focusing on the conformational and oligomerization properties of FH, showed that these are NaCl concentration and pH-dependent. For example, it was shown that at low pH, 5.4, FH exists mostly in a compact conformation. FH became more elongated at pH 9.4 (52). It is possible, therefore, that change in the protonation state of the aforementioned His residues (^{651}His in module 11 and ^{773}His in module 13, located at critical positions between modules) - from positively charged, polar and hydrophilic below its pKa to neutral and hydrophobic above its pKa - could affect intermodular orientations and hence the overall architecture of the whole of the central region.

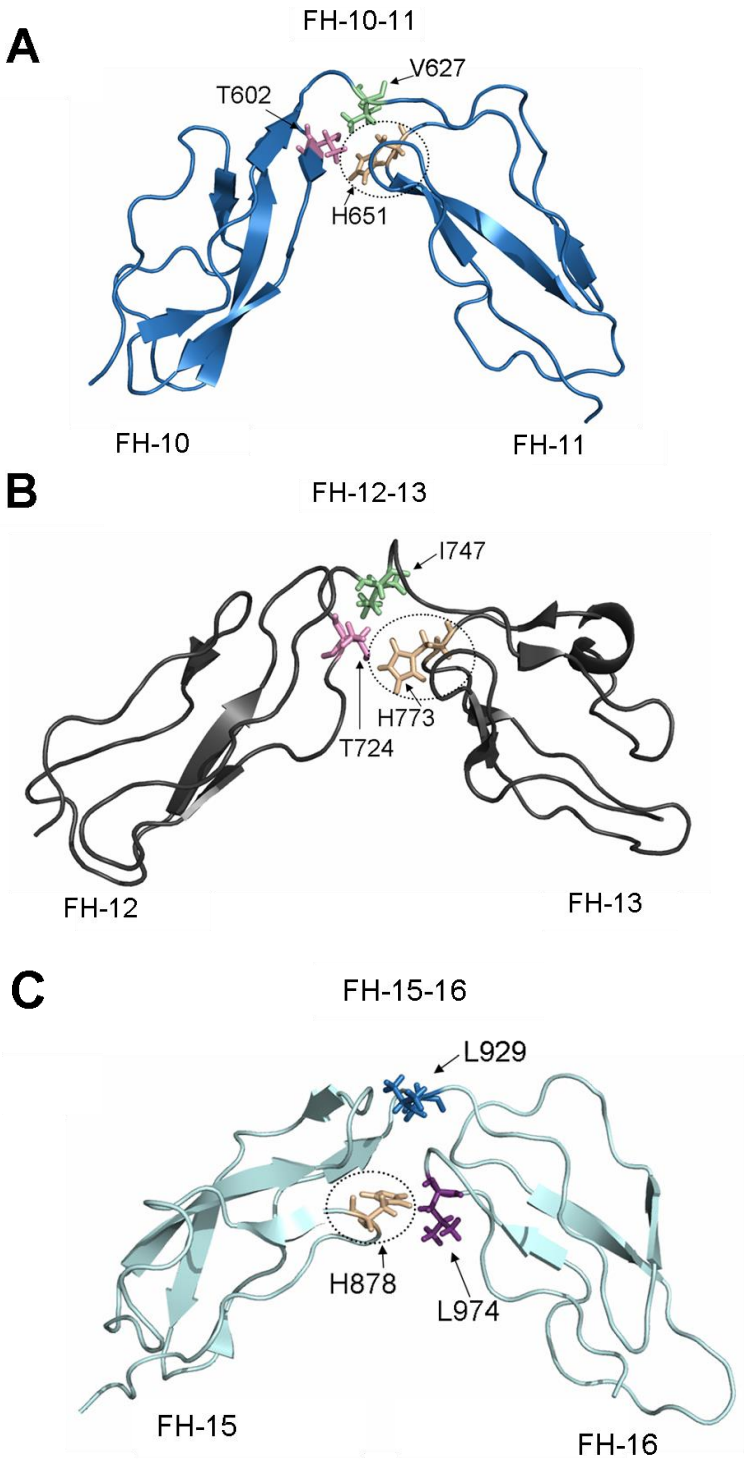


Figure 4.33 Cartoon representation of the structures of (A) FH-10-11, (B) FH-12-13 and (C) FH-15-16. In (A) and (B) the three residues predominantly responsible for the similarity of the interfaces of FH-10-11 and FH-12-13 are shown in sticks and are color-coded. The His residues that occupy an equivalent position in their corresponding module pair are highlighted. In (C) the H878 involved in the interface is highlighted with and hydrophobic residues of the interface are indicated.

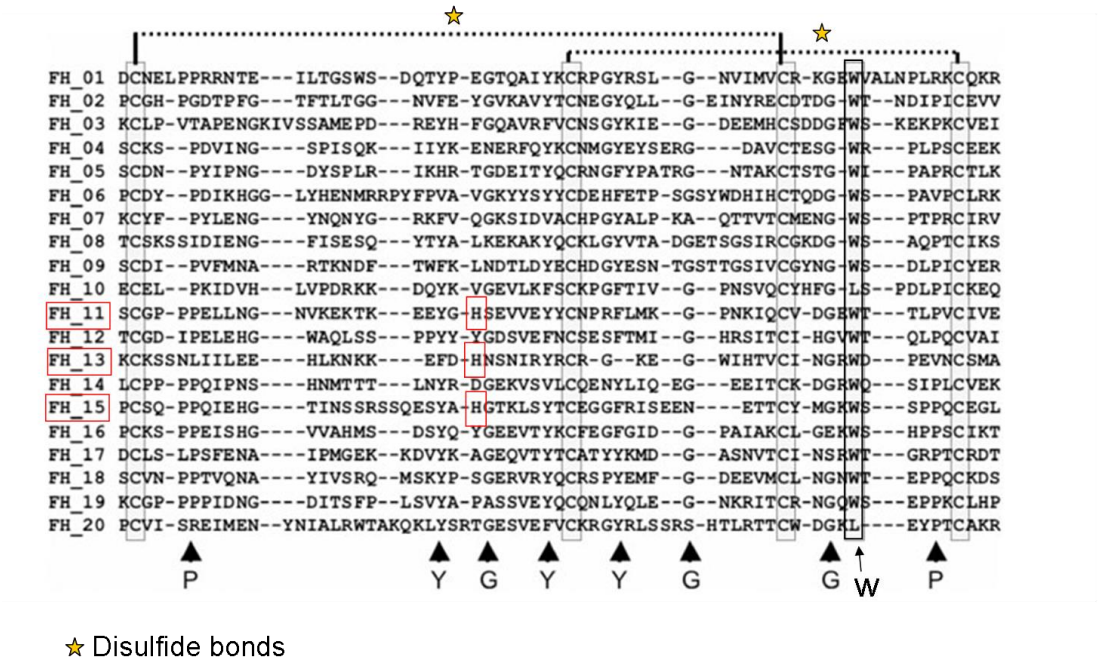


Figure 4.34 Multiple sequence alignment of the 20 CCPs that compose FH. Red boxes highlight the CCP modules and their corresponding His residues that are involved in the interfaces of FH-10-11, FH-12-13 and FH-15-16 (figure modified from (38)).

The new structures determined in this study allow for reinterpretation of the SAXS data for FH-10-15 and help distinguish between the two-currently existing-models (see also Fig. 4.30). SAXS data as well as the resultant modelling for the central region of FH will be presented in the next chapter, Chapter 5, while a hypothesis of how this region is important for the biological action of FH will be discussed in Chapter 7.

CHAPTER 5

Small-angle X-ray scattering-based studies of the central region of FH & analysis of flexibility of FH-10-11 and FH-11-12 by NMR[$]

[$] Parts of this chapter have been published in (149).

Overview

The central region of FH is emerging as an architecturally critical element of the protein that is essential for its full biological activity. A previous study (43) which combined the NMR-derived structure of FH-12-13 with interpretation of SAXS data collected on longer segments of FH showed that FH-10-15 adopts a compact arrangement. Since there was no evidence for co-existence with the more extended conformational possibilities that are typical of other multiple-CCP module segments, this compact conformation appears to be stable and could be rigid. It is noteworthy that within this region FH-12-13 proved to be inflexible as judged by both NMR and SAXS (43).

This study was, however, unable to distinguish between two different compact structural models for FH-10-15 ((43), Fig. 5.1)). In particular it did not reveal the tilt angles between other module pairs (besides FH-12-13) and only hinted at the existence of pairs with high tilt angles in which modules are folded back alongside their neighbours. Were such arrangements to exist they could have profound albeit indirect effects on those functions of FH that require collaboration between its N- and C-terminal ends. The newly determined solution structures of FH-10-11 and FH-11-12 provided an opportunity for further efforts to combine NMR and SAXS to probe the structure and dynamics of the central region of FH.

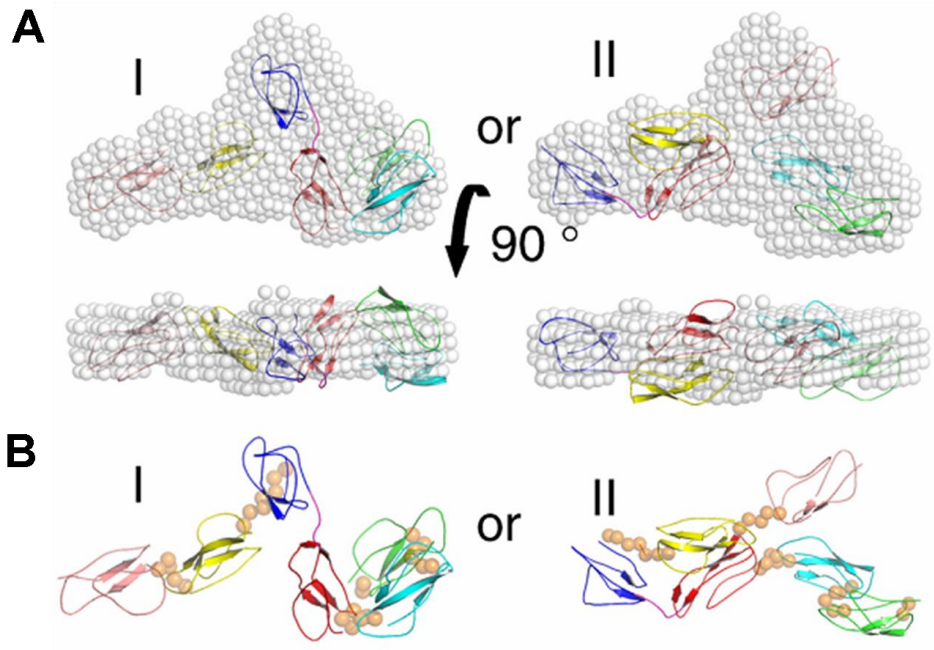


Figure 5.1 Previous models of FH-10-15 inferred from SAXS and the solution structure of FH-12-13. (A) Overlay of the two most typical BUNCH-derived models of FH-10-15 with the DAMMIF-derived *ab initio* shape envelopes (I and II). (B) The two most typical BUNCH-derived models of FH-10-15 (I and II). The cartoons are colour coded as follows: FH-10, green; FH-11, cyan; FH-12, red; FH-13, blue; FH-14, yellow and FH-15, pink. Orange spheres represent linker residues (Figure and legend adapted from (43)).

Initially SAXS data were used to validate the NMR-derived structures of FH-10-11 and FH-11-12 presented in Chapter 4. Subsequently SAXS was used to confirm the lack of flexibility within these pairs that had been inferred from the relaxation data. Finally the high-resolution structures derived from NMR (along with SAXS data collected on FH-8-9) were used to help interpret SAXS data collected on FH-10-15 and FH-8-15.

Taken together with the difficulties that had been encountered in producing samples of FH-13-14 and FH-14-15, these studies were interpreted in terms of a highly compacted structure for FH-12-15 and a dependence of FH-14 on its neighbours for stability.

5.1 Introduction to SAXS

SAXS is a technique for structural analysis performed in solution. In a typical SAXS experiment, the sample is exposed to a collimated monochromatic X-ray beam. The scattered intensity I is recorded as a function of the scattering vector

s=4πsinθ/λ where 2θ is the angle between the incident and the scattered beam (Fig. 5.2). The scattering from the solvent is measured separately and then subtracted to give the data of interest. This background-corrected intensity *I(s)* is proportional to the scattering of a particle averaged over all orientations. Several 1-D curves can be extracted from the *I(s)* information, from which important parameters can be obtained related to the molecule's size, oligomeric state and overall shape. These parameters include the molecular mass (MM), radius of gyration (R_g), hydrated particle volume (V_p), and maximum particle diameter (D_{max}) (154).

From the scattering patterns, *ab initio* shape reconstruction, validation of high-resolution structural data, rigid-body modeling of larger complexes using known NMR substructures and analysis of flexibility, are all feasible through widely-used methods.

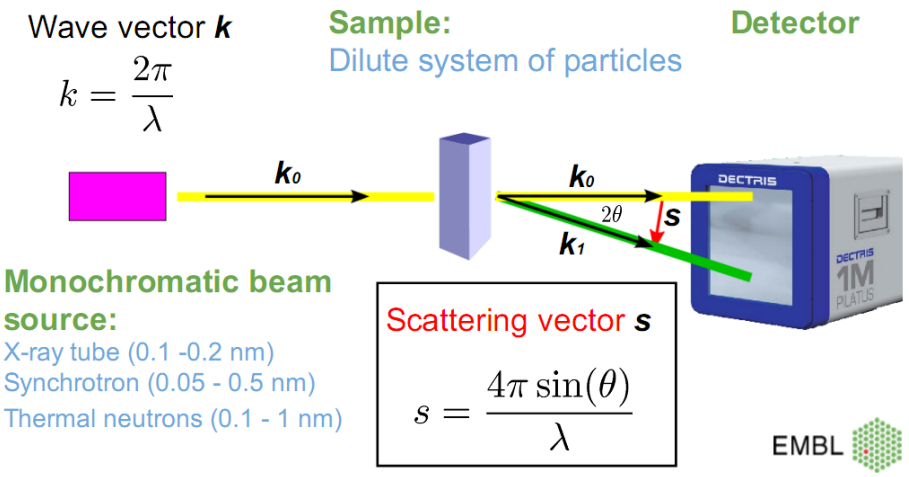


Figure 5.2 Representation of the scattering experiment; k_0 corresponds to the incident wave vector, k_1 to the final wave vector and 2θ to the so-called scattering angle (the angle between the incident and the scattered beam) (figure from Dr H. D. T. Mertens).

5.2 SAXS-derived parameters for FH-8-9, 10-11, 11-12, 10-12 and 8-15

Analysis of parameters extracted from scattering profiles and real-space distance distribution functions, *p(r)*, revealed monomeric proteins under conditions used for data collection (Table 5.1).

Bi-modules FH-8-9 and FH-10-11 yielded similar R_g values of ~2.2 nm, and D_{max} values of ~7.5 nm, while FH-11-12 appears to be, overall, more extended with

R_g and D_{max} of 2.4 nm and 8.3 nm respectively (Table 5.1). The positively skewed *p(r)* functions of both bi-modules (Fig. 5.3) are characteristic of two-domain "dumbbell" shaped structures, with the first maximum representing a common average intra-module distance of ~1.7 nm and the second maximum representing an approximate average separation of the centers of mass of the modules, ~3.5 nm for FH-8-9 and ~3.8 nm for both FH-10-11 and FH-11-12 (Fig. 5.3). Thus intra-modular distances are consistent with the well-established dimensions of CCP modules.

Construct	Rg^{autoRg}, nm	Rg^{GNOM}, nm	D_{max}, nm	V_p, nm^3	MM_{SAXS}, kDa	V_{DAM}, nm^3	MM_{DAM}, kDa
FH-8-9 (13.7 kDa)	2.1 ± 0.1	2.2 ± 0.1	7.5 ± 0.5	21 ± 5	15 ± 5	23 ± 5	11 ± 5
FH-10-11 (13.9 kDa)	2.2 ± 0.1	2.2 ± 0.1	7.5 ± 0.5	22 ± 5	16 ± 5	21 ± 5	10 ± 5
FH-11-12 (14.1 kDa)	2.4 ± 0.1	2.3 ± 0.1	8.3 ± 0.5	25 ± 5	19 ± 5	25 ± 5	12 ± 5
FH-10-12 (20.6 kDa)	2.7 ± 0.1	2.7 ± 0.1	9.5 ± 0.5	33 ± 5	20 ± 5	47 ± 5	24 ± 5
FH-8-15 (54.7 kDa)	3.5 ± 0.1	3.7 ± 0.1	12.5 ± 0.5	92 ± 10	62 ± 5	93 ± 10	47 ± 5

Table 5.1 FH SAXS parameters. Rg^{autoRg} and Rg^{GNOM} are the radius of gyration estimated from the SAXS data using the automated Guinier analysis routine and GNOM, respectively. D_{max}, V_p, V_{DAM}, MM_{SAXS} and MM_{DAM} are the maximum particle dimension, hydrated particle volume, dummy-atom model total excluded volume, molecular-mass estimated from *I(0)*, and molecular-mass estimated from the dummy-atom model volume, respectively. Data shown is that averaged, or merged and extrapolated to infinite dilution, using the SAXS profiles recorded at several concentrations. The molecular mass calculated from the sequence is shown in parentheses.

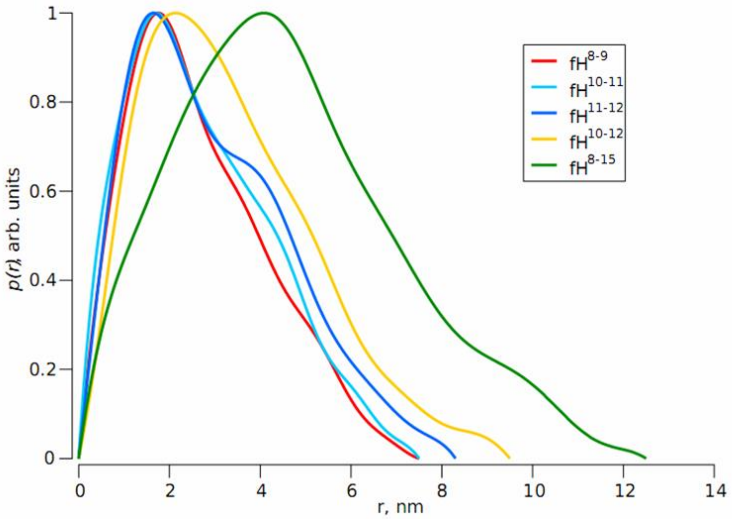


Figure 5.3 SAXS data for FH constructs. Pairwise interatomic distance distributions, *p(r)* for the FH constructs; FH-8-9 (red), FH-10-11 (light blue), FH-11-12 (dark blue), FH-10-12 (orange) and FH-8-15 (green).

5.3 Validation by SAXS of the NMR-derived structures of FH-10-11, FH-11-12 and of the NMR-derived, concatenated, FH-10-12 model

The NMR structures calculated were fit to the experimental SAXS data using the program CRYSOL (155). All 20 individual NMR conformers of the FH-10-11 ensemble provide a good fit to the scattering data with χ values between the data and the model of $1.3 < \chi < 1.6$ (Fig. 5.4).

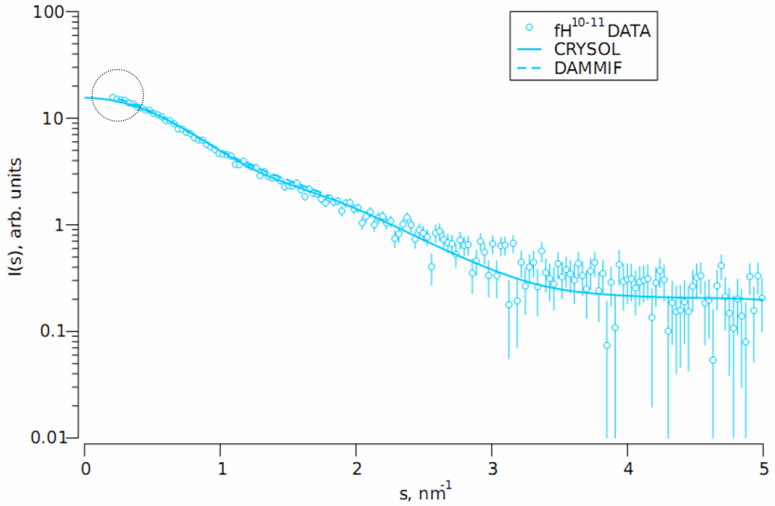


Figure 5.4 Fit of the best NMR conformer of FH-10-11 (solid lines) and best DAMMIF model (broken lines) to the FH-10-11 data (circles) (circles highlight the slight deviation on the curves observed form the fitting to the SAXS data).

The individual NMR conformers in the FH-11-12 ensemble, provide a less good but still acceptable fit with χ values between the data and the model of $1.6 < \chi < 2.2$ (Fig. 5.5).

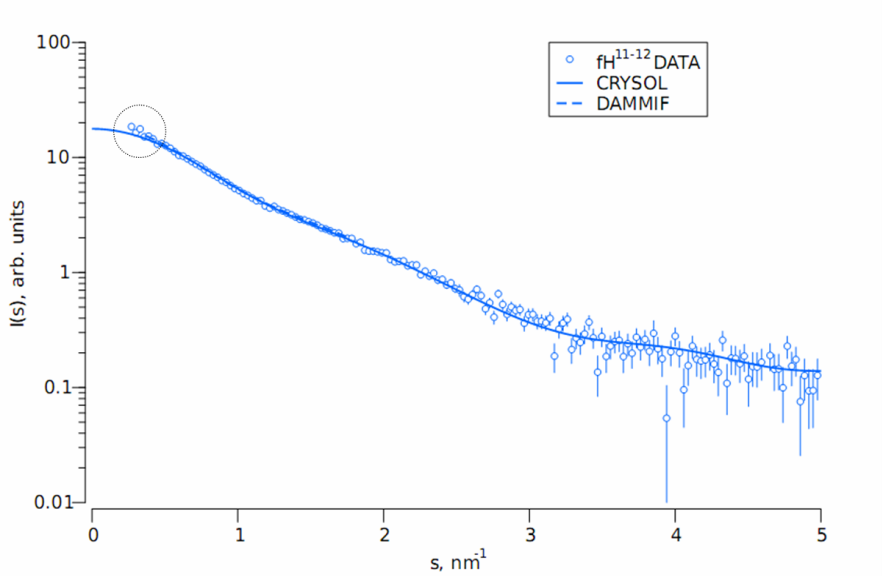


Figure 5.5 Fit of the best NMR conformer of FH-11-12 (solid lines) and best DAMMIF model (broken lines) to the FH-11-12 data (circles) (circles highlight the slight deviation on the curves observed form the fitting to the SAXS data).

In an *ab initio* approach, the average structures of both FH bi-modules were reconstructed from SAXS data using the bead-modeling program DAMMIF (156). Ten independent reconstructions were performed for each pair, and within each set highly similar shapes were obtained as measured by low (<1.0) normalized spatial discrepancies (NSD) (157). The discrepancies between the generated models and the experimental data were good ($0.9 < \chi < 1.2$) in both cases (Figs. 5.4 and 5.5). The total excluded dummy atom volume of the model was consistent (in each case) with what was expected, yielding comparable estimations of MM (Table 5.1). The average *ab initio* shapes were superimposed with the corresponding NMR ensembles of FH-10-11 and FH-11-12 using the program SUPCOMB (157), and good spatial agreement was observed (Fig. 5.6 A and B).

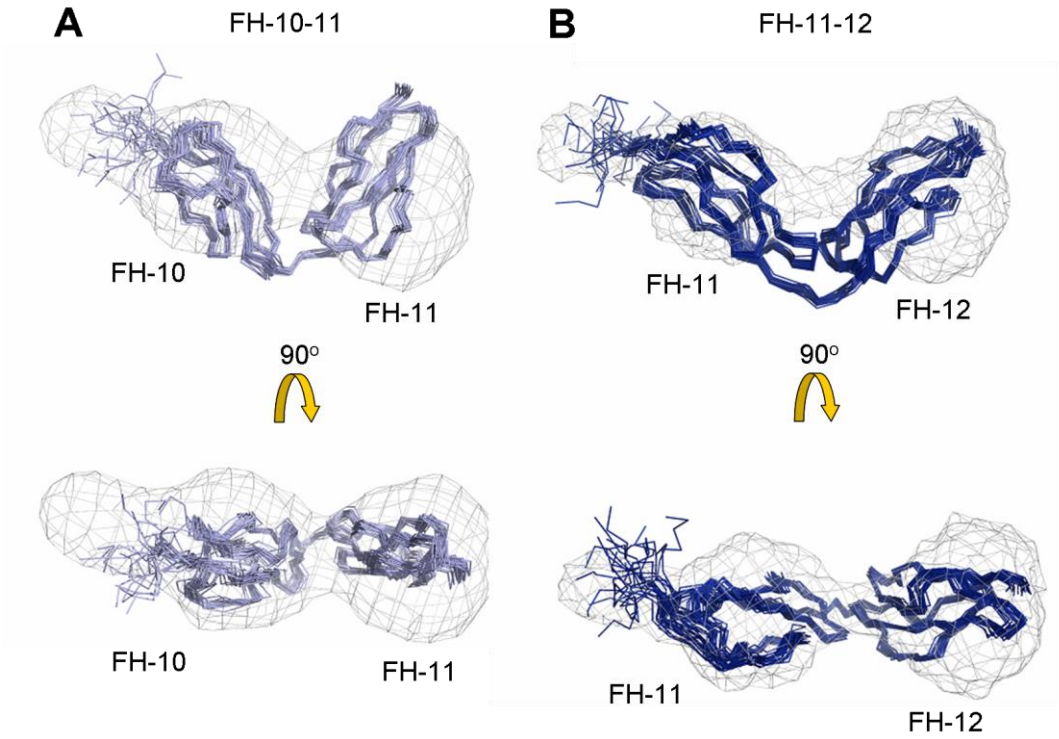


Figure 5.6 Overlay of the NMR ensembles of (A) FH-10-11 and (B) FH-11-12 with the most representative *ab initio* DAMMIF models, superimposed using the program SUPCOMB. Models FH-10-11 and FH-11-12 are colored light blue and dark blue respectively and *ab initio* shape envelopes are shown as mesh surfaces (two views are shown for each following 90° rotation about the x axis).

A model of the FH-10-12 structure was derived by concatenating the average structures of FH-10-11 and FH-11-12 thus generating a simple model for the tri-module FH-10-12, using the very similar independent structures of module 11 as an alignment template (in MODELLER). This model was then validated by fitting it to the SAXS data for FH-10-12 using CRYSOL (155) and provided an excellent fit (χ= 0.8) (Fig. 5.7).

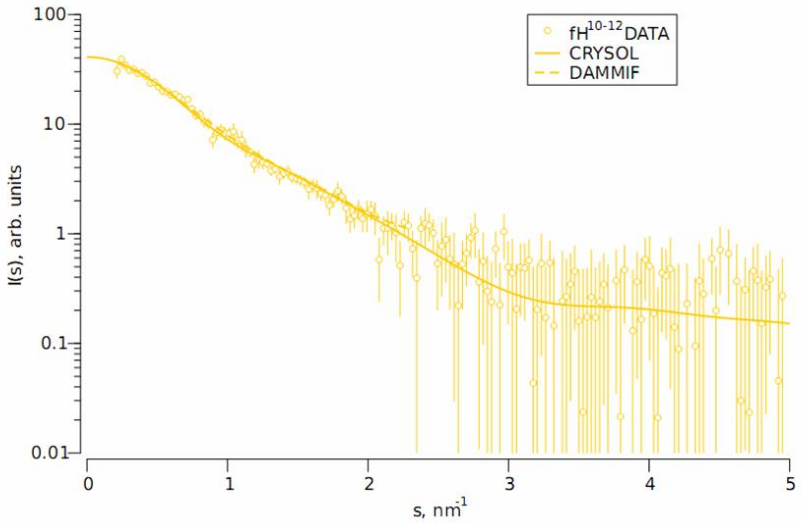


Figure 5.7 Fit of the concatenated NMR model for FH-10-12 (solid lines) and best DAMMIF model (broken lines) to the FH-10-12 data (circles).

In a separate *ab initio* approach similar to that described above for the bi-modules, the average solution structure of FH-10-12 was reconstructed from SAXS data using DAMMIF (156). As with the bi-modules, normalized spatial discrepancies were low ((NSD) =0.73 ± 0.12) and discrepancies between models and experimental data were low (χ =0.88 ± 0.01) (Fig. 5.7), while the total excluded dummy atom volume of the model was consistent with the MM (Table 5.1). Moreover the average *ab initio* shape superimposed well with the concatenated NMR model of FH-10-12 (SUPCOMB, (157)) (Fig. 5.8).

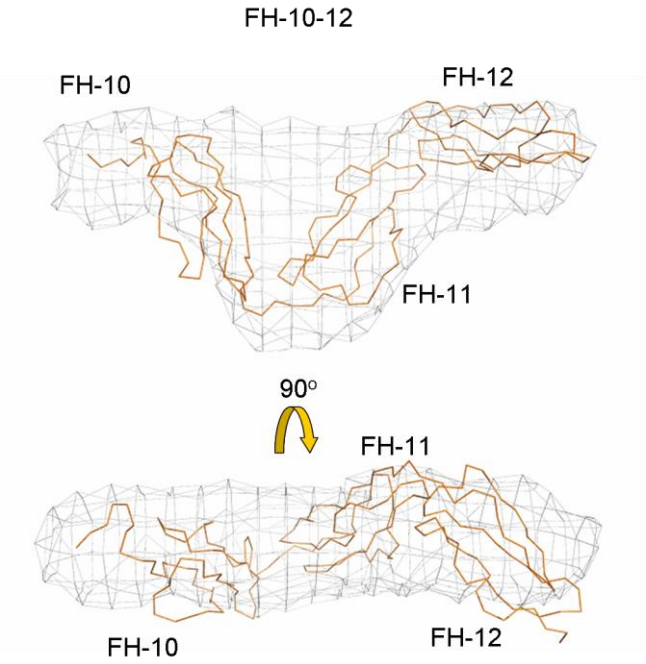


Figure 5.8 Overlay of the concatenated NMR model of FH-10-12 with the most representative *ab initio* DAMMIF model, superimposed using the program SUPCOMB. The FH-10-12 model is shown in orange while the *ab initio* shape envelope is shown as mesh surface (two views are shown for each following 90° rotation about the x axis).

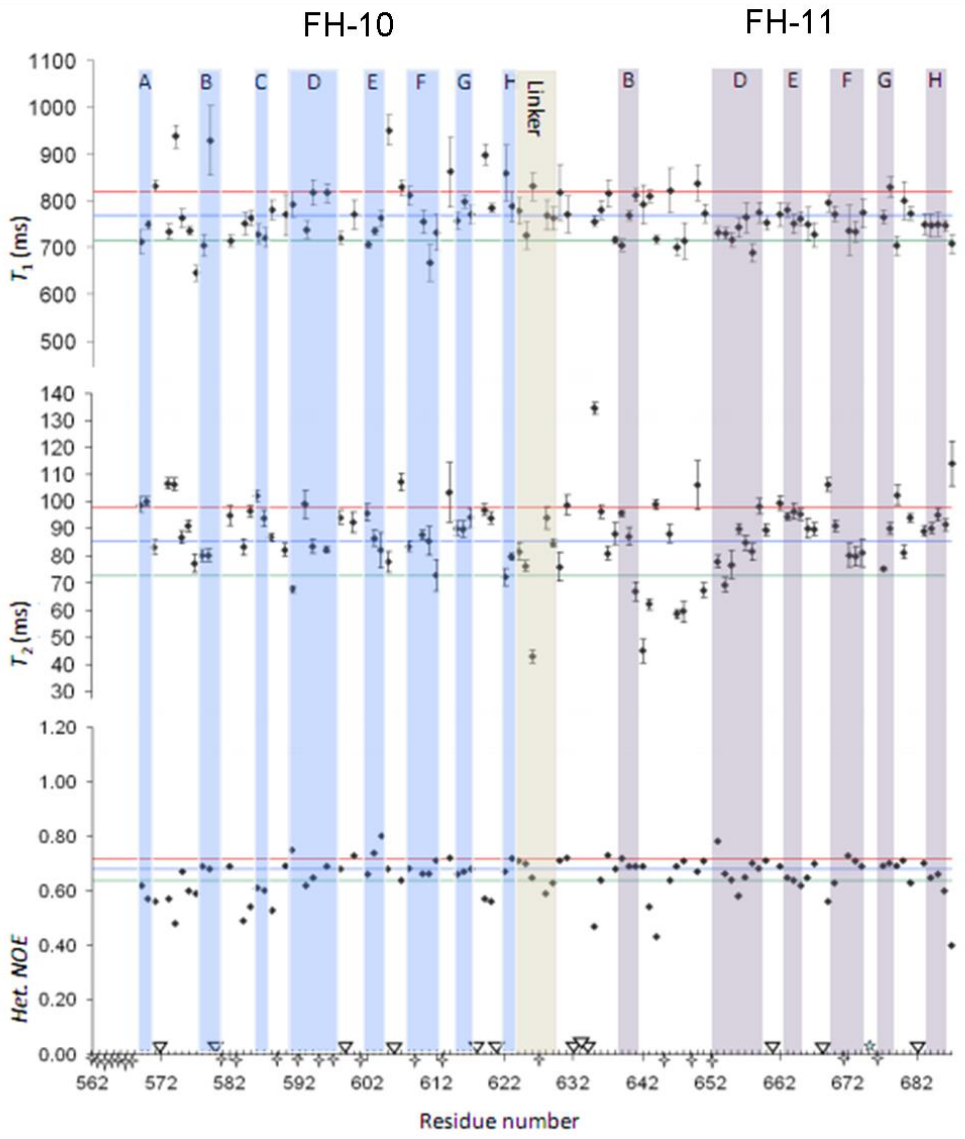
Studies of tri-molecular FH-10-12 showed that addition to FH-10-11 of module 12 or, from another perspective, addition to FH-11-12 of module 10, yielded only modest 0.1-0.3 nm increases in R_g and 0.8-1.2 nm increases in D_{max} (Table 5.1). Thus these three modules must form a non-linear structure.

5.4 Flexibility analysis of FH-10-11 and FH-11-12 by NMR

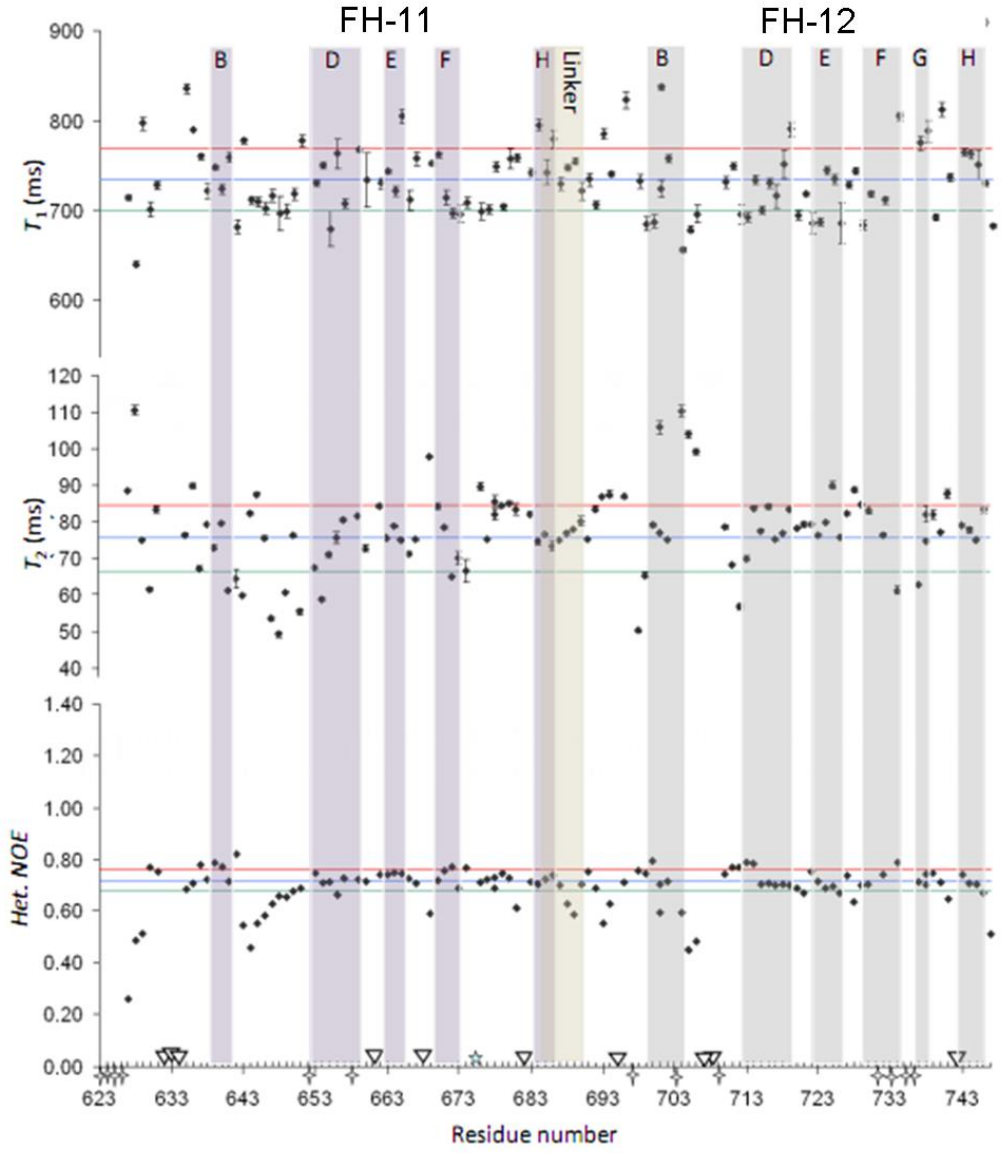
Measurements of ^{15}N spin-lattice (T_1), ^{15}N spin-spin (T_2) times and ^{1}H, ^{15}N NOEs were performed on both FH-10-11 and FH-11-12. Briefly, motions in the range of μs-ms can be investigated by T_1 and T_2 values while the latter are sensitive to fast motions. Motions in the nanosecond time scale can be probed by the heteronuclear NOE. Regions in the protein with flexible properties have small (<0.7) NOEs and larger than average T_1 and T_2 values. A T_1/T_2 ratio that exceeds the average by more than one standard deviation probably reflects a contribution to the line width by conformational exchange (158).

The corresponding values obtained after processing for each experiment were plotted against residue number. Peaks corresponding to prolines, residues that were

located in overlapped regions in the spectra and those that exhibited weak signals were not included in the analysis. The amide for residue ^{675}Asp could not be identified in any of the two constructs thus was also excluded from the analysis. The figures below, (Figs. 5.9 A and B), show the results obtained.



(A) FH-10-11



(B) FH-11-12

Figures 5.9 (A) and (B) ^{15}N Relaxation data for FH 10-11 and FH 11-12. Plots of the measured ^{15}N spin-lattice (T_1) and ^{15}N spin-spin (T_2) relaxtion times, and heteronuclear (^{1}H, ^{15}N) NOE values, plotted versus residue number. In each plot, the mean (shown in blue lines), +/- one standard deviation is indicated (shown in red and green lines respectively), calculated on the basis of residues with heteronuclear NOEs of 0.6 or above (159). Beta-strands and the position of the linker are shown by the labeled, shaded boxes. Inverted triangles on the x-axis indicate proline positions, star shapes denote residues that were excluded from the analysis due to overlap or low signal while a star shown in cyan marks the amide (^{675}Asp) that was not found in the HSQC.

For FH-10-11, mean values of T_1, T_2 and heteronuclear NOEs are similar in both modules (mean T_1 values of 777 ms and 760 ms for FH-10 and FH-11 respectively, mean T_2 values of 87 ms and 85 ms for FH-10 and FH-11 respectively and mean HetNOE values of 0.69 and 0.68 for FH-10 and FH-11 respectively).

These equal T_1 and T_2 averages are consistent with the notion that these modules do not move independently of one another but behave as a single fused entity. A similar observation was made in the case of FH-11-12 (mean T_1 values of 735 ms and 732 ms for FH-11 and FH-12 respectively, mean T_2 values of 74 ms and 77 ms for FH-11 and FH-12 respectively and a mean HetNOE value of 0.72 for both).

Notably, the linker regions are not particularly mobile (compared to other regions of the module pairs) on the timescales probed by relaxation measurements. Although the FH-11-12 linker includes two residues with lower than average NOEs, T_1 and T_2 values are unexceptional and other residues have average NOEs. In the FH-10-11 linker, one residue (Gln^{626}) has a low T_2, but this is in the middle of the six-residue linking sequence with its backbone in an exposed position (on the convex surface of the tilted modules) so it could be in conformational exchange independently of the rest of the linker.

In FH-10, relatively low ^{1}H, ^{15}N NOE values reveal ns-timescale mobility in the A-B and G-H loops as well as in the region between strands B and D that includes the short strand C. T_1 and T_2 values are unexceptional so there is little evidence for conformational exchange in FH-10 (this is despite the earlier observation that this module has very few slowly exchanging protons). The B-C loop also appears mobile (low HetNOEs) and residues within this loop belong to a small electropositive patch on one face of FH-10 (Fig. 5.10). The N-terminal region of FH-11 is more mobile (low hetNOEs, some high T_1 and T_2 values) in the FH-11-12 context (than in FH-10-11) as expected given the absence of module 10.

Interestingly, the B-D loop of FH-11 (within FH-10-11, lacking strand C) exhibits numerous residues with low T_2 values as well as two residues with very low heteronuclear NOEs. A similar pattern is observed in FH-11 within the FH-11-12 context, except that there are a significantly greater number of low heteronuclear NOEs. This is a mobile feature of module 11 in both pairs but it is stabilised to some extent in FH-10-11, presumably due to its participation in the interface with module 10. Likewise, in FH-12 there are several consecutive residues just prior to strand D with higher than average heteronuclear NOEs, which is again consistent with rigidification due to the compact interface with the preceding module. These are in

contrast to FH-12 strand B and its following residues that have the highest T_2 values in the protein and (for a strand) low heteronuclear NOEs.

In summary, several strands of evidence – similar average relaxation parameters on a per module basis, the rigidifying of loops close to intermodular interfaces, and the low levels of overall mobility in the linkers – reinforce inferences from structure calculations and from SAXS-based EOM analysis (see below), that both module pairs are overall rigid structures with little flexibility between modules.

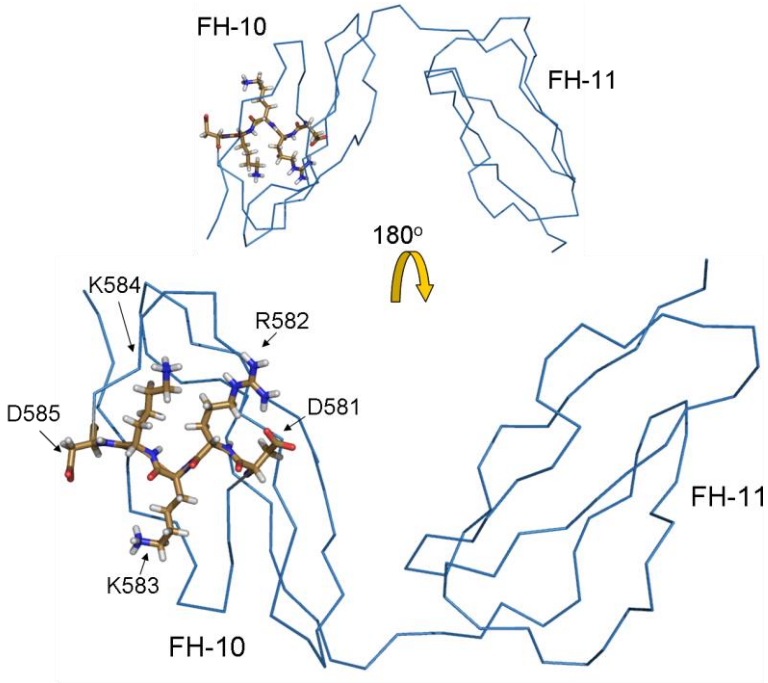


Figure 5.10 Ribbon representation of the closest-to-mean structure of FH-10-11 showing the region in module 10 that appears mobile on the timescales probed by NMR and that also forms a miniature electropositive patch. The corresponding residues are shown as sticks as indicated.

5.5 Flexibility analysis of FH-10-12 using the SAXS Ensemble Optimization Method (EOM)

The flexibility of construct FH-10-12 was studied by the SAXS-based EOM method which can assess the degree of compactness or extendedness. This method selects an ensemble of possible conformations from a pool of randomly generated models (10,000), using CRYSOL to calculate the theoretical scattering profiles and a genetic algorithm, GAJOE to select the representative set (160). In this way the solution behaviour of the sample relative to the random pool may be described.

When a selected ensemble shows a size-distribution profile as broad as that of the random pool it may be considered flexible while a narrower profile is indicative of a relatively rigid system.

For studying the flexibility of FH-10-12 the random pool of conformers was initially generated from FH modules 10, 11 and 12 with all linkers defined as flexible. The selected ensemble from the resultant size-distribution was less broad than that of the random pool and was skewed toward compact structures (Fig. 5.11 A). This suggests that the tri-module is not significantly flexible or extended in solution. In a subsequent calculation the FH-11-12 structure was treated as a rigid body while the FH-10-11 linker was allowed to be flexible. The R_g distribution from that calculation was narrow suggesting that the FH-10-11 linker is not significantly flexible in the FH-10-12 context (Fig. 5.11 C). When the orientations of FH-11-12 CCP modules are fixed, mainly compact 10-11 linker orientations (as also observed under NMR conditions) are selected in order to fit the SAXS data for FH-10-12.

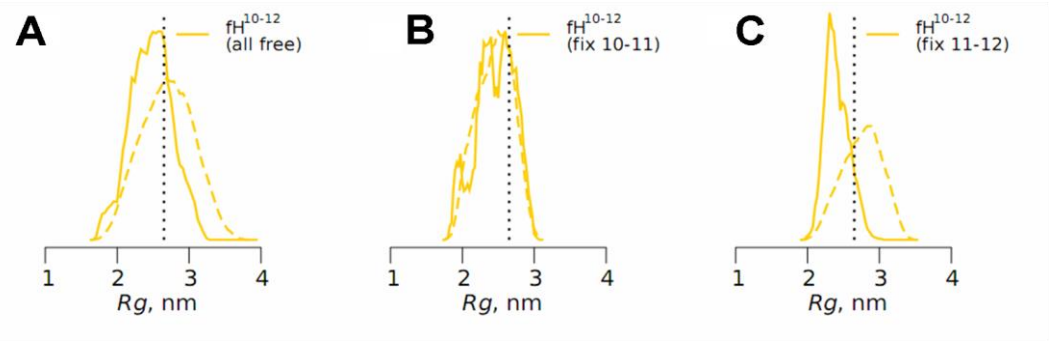


Figure 5.11 EOM analysis of FH-10-12. A vertical line indicates the experimentally determined R_g. Yellow dashed curves show the pools of randomly generated models, while solid curves the selected representative ensembles from within the pools. (A) All linkers were defined as flexible. (B) FH-10-11 defined as a rigid body while the FH-11-12 linker as flexible. (C) FH-11-12 defined as a rigid body while the FH-10-11 linker as flexible.

5.6 Modelling of FH-11-14 and FH-10-15 based on SAXS and NMR-derived structures

SAXS-based rigid body modelling was conducted for the larger FH constructs using merged models of the known NMR structures of FH-10-11, 11-12 and 12-13 (Fig. 5.12) the SAXS-derived data and the programs BUNCH and

CORAL (161), (162). This enabled us to optimise the spatial distribution of the CCPs in the corresponding SAXS envelopes and was performed specifically for domains 11-14, 10-15 and 8-15.

The NMR-derived model of FH-10-13 shown below (Fig. 5.12) demonstrates how the curvature of FH-10-11 does not follow the arc described by FH-11-13 resulting in an out-of-plane zigzag structure.

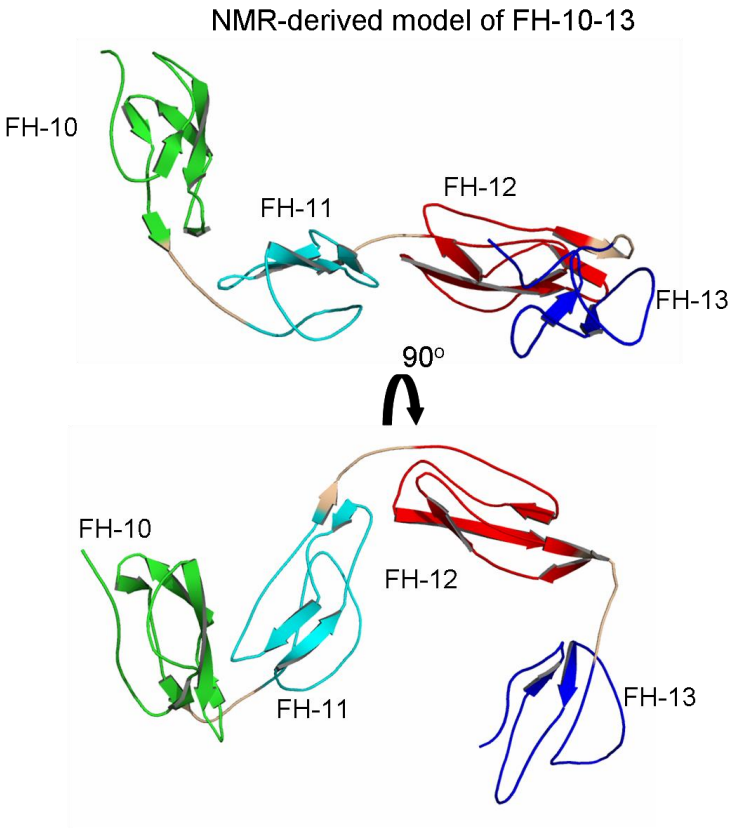


Figure 5.12 Cartoon representation of the NMR-derived concatenated model for FH-10-13. Two views are shown following 90° rotation about the x-axis. The linkers are shown in beige.

For the FH-11-14 region, the NMR-derived concatenated model of FH-11-13 was used in the modelling. The resultant model was in good agreement with previous findings (43) (Fig. 5.13). In this model FH-11-13 adopt a C-shaped orientation which is followed by an unusual and extreme tilt of the linker between FH-13 and FH-14 which in turn drives the orientation of FH-14.

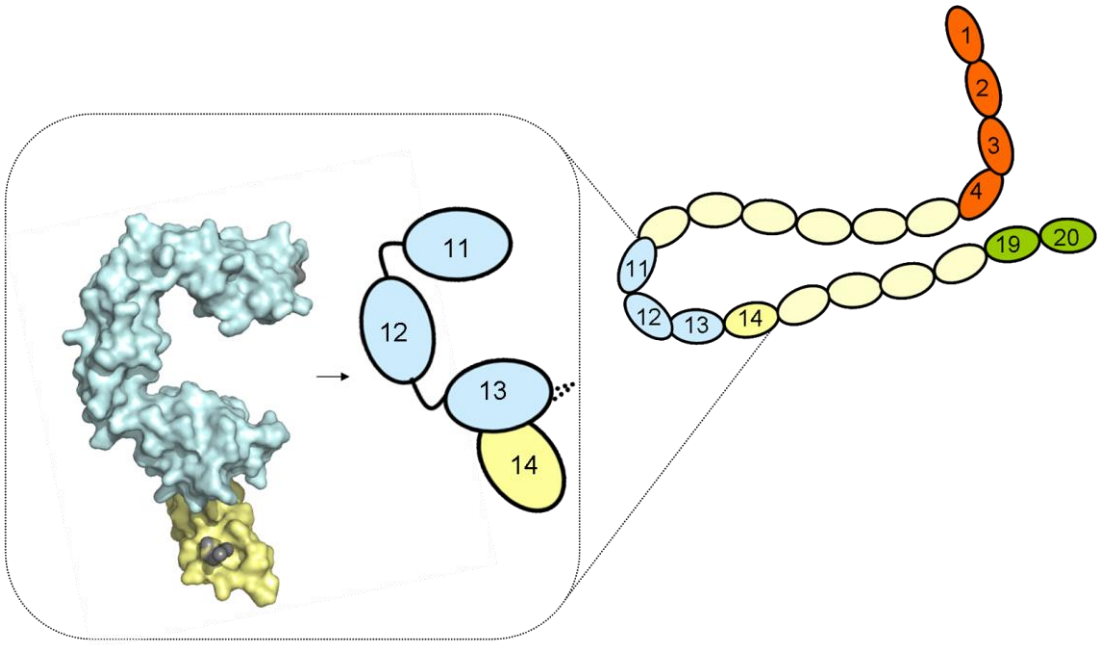


Figure 5.13 Modelling of the FH-11-14 region. Representation of the FH molecule: each CCP is shown as an oval; the FH-1-4 and 19-20 regulatory domains are shown in orange and green respectively. The NMR-derived FH-11-13 domain is shown in light blue and was used for the SAXS-derived modelling of the FH-11-14 region. FH-14 is shown in yellow (grey corresponds to dummy residues at the C-terminus of FH-14). The surface representation of the spatially optimized arrangement of FH-11-14 following modelling is shown colour-coded as already mentioned followed by a simplified version using oval shapes for each CCP module.

For SAXS-based modelling of the FH-10-15 region, the NMR-derived model of FH-10-13 was utilized. The model providing the best results (Fig. 5.14) was consistent with model I of the two possible previously determined, thus enabling distinction between the two (Fig. 5.15). In this model FH-10, 11, 12 and 13 follow a more open zigzag path with 90^{o} tilts. The extreme tilt between FH-13 and FH-14 results in a more compact globular "13-15" end as FH-14 and 15 appear to wrap around FH-13 and the 12-13 linker.

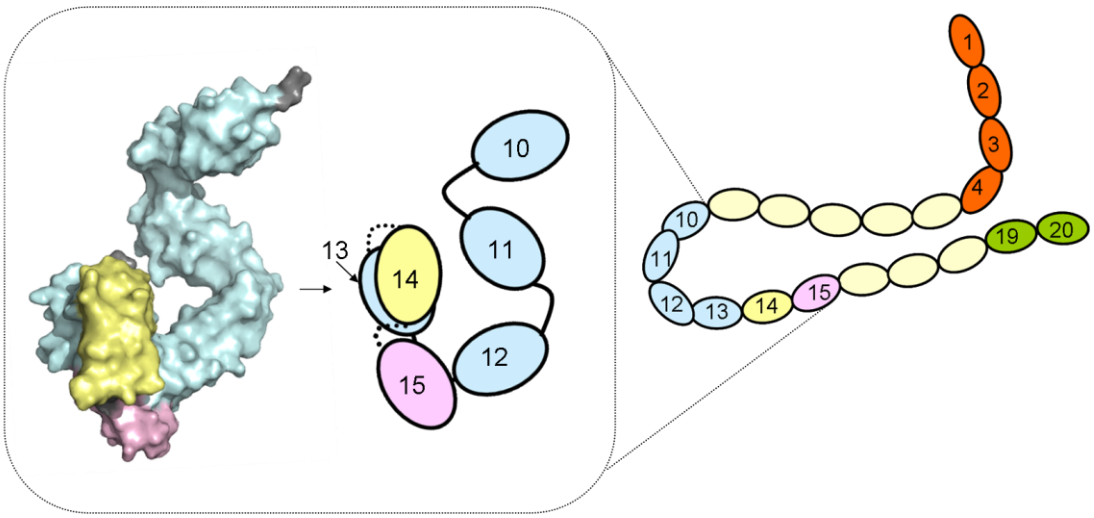


Figure 5.14 Modelling of the FH-10-15 region using the NMR-derived FH-10-13 model. See also figure 5.13 for explanation of the (similar) representations viewed. In this case the NMR-derived FH-10-13 is shown in light blue while FH-14 is shown in yellow and FH-15 in pink.

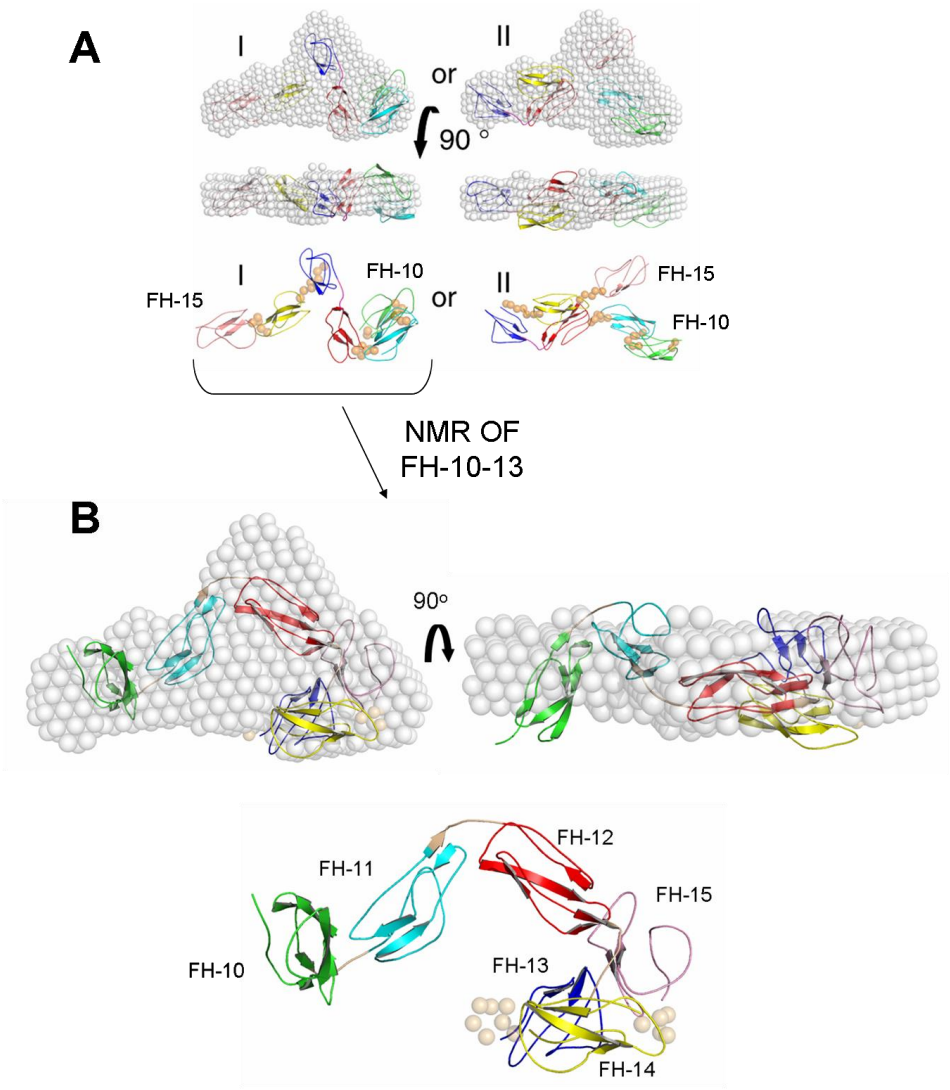


Figure 5.15 Modelling of the FH-10-15 region using the NMR-derived concatenated model for FH 10-13. (A) The two possible models of FH-10-15 as previously determined (43). (B) Model I from (A) was chosen as the representative based on the NMR structures of FH-10-11, 11-12 and 12-13. Representation of the SAXS-derived envelope (from model I) shown in spheres and the modelled CCPs within showed as cartoons as indicated on the figure. Note that in the newest model the positioning of the CCP modules within the envelope is the other way around compared to the initial model. The experimentally determined linkers from the NMR-derived FH-10-13 model are shown in beige whereas light-orange spheres represent linker residues modelled as a chain of dummy atoms. Two views are shown following 90° rotation about the x-axis. Finally the FH-10-15 modelled region is shown in cartoon representation without the SAXS envelope (CCPs are labeled as indicated).

5.7 SAXS data of FH-8-15 and flexibility analysis using the SAXS Ensemble Optimization Method (EOM)

The eight-module FH-8-15 construct yields SAXS parameters that are not compatible with a highly extended structure; its R_g and D_{max} values of 3.5 and

12.5 nm respectively are consistent with previous SAXS-based experiments (15), (43) (Table 5.1, Fig. 5.16). This reinforces the conclusion that the FH central modules contribute to a relatively compact structure core.

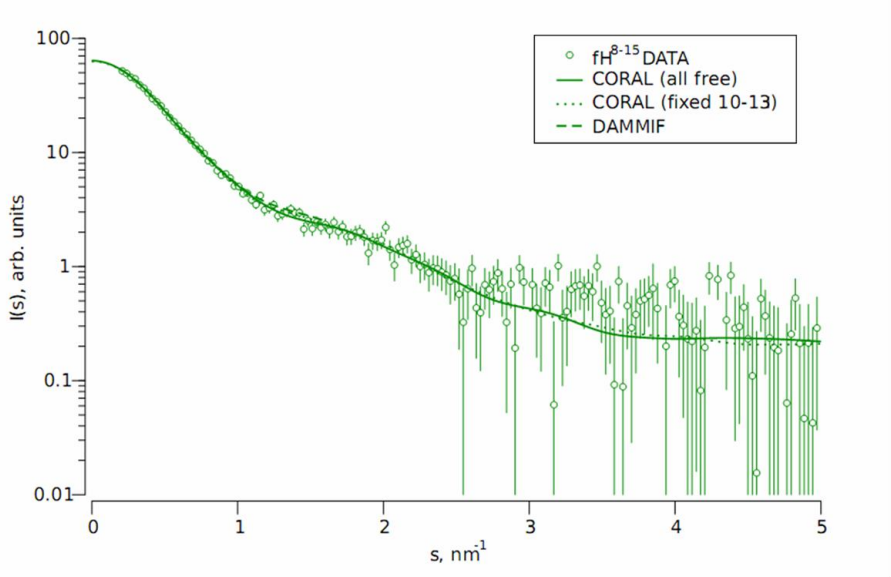


Figure 5.16 Fits of the best CORAL rigid body models of FH-8-15 with all linkers between CCP modules free (solid line) or with FH-10-13 fixed (dotted line) and the best DAMMIF model (broken line) to the FH-8-15 data.

The EOM analysis of FH-8-15 was conducted either with all linkers between modules defined as flexible, or with all linkers excluding 10-11, 11-12, 12-13 (43) treated as flexible. In the first case the resulting size-distribution of the selected ensemble is narrower than the random pool and skewed toward compact structures suggesting that this construct has limited flexibility (Fig. 5.17 A). In the second case however, a predominantly two-state system resulted. Significant populations of extended and more compact linker conformations are required in order to fit the experimental data (Fig. 5.17 B).

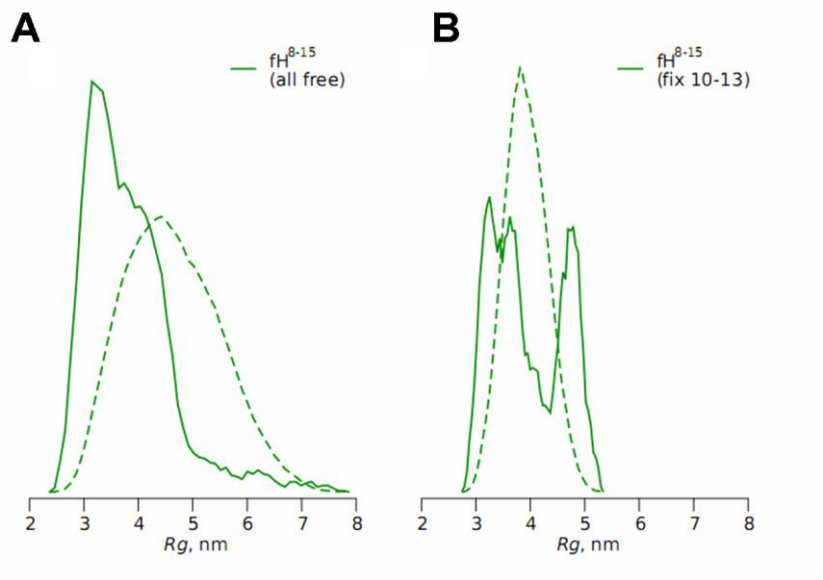


Figure 5.17 EOM analysis of FH-8-15. Green dashed curves show the pools of randomly generated models while solid lines the selected representative ensemble from within the pools. (A) All linkers defined as flexible. (B) FH-10-13 is defined as a rigid body while the FH-8-9, 13-14 and 14-15 linkers as flexible.

5.8 Conclusions

Both NMR-derived structures of FH-10-11 and FH-11-12 were validated by SAXS, since good fits were observed in both cases. The FH-10-12 model obtained by concatenating the closest-to-average structures of the two overlapping bi-modules provided an excellent fit to the SAXS data. The two-domain "V-shaped" orientations exhibited by both the NMR-derived FH-10-11 and FH-11-12 were further supported by the SAXS-derived parameters from real-space distribution functions *p(r)*. All in all it appears that FH-10-12 forms a rigid and not highly elongated structure.

EOM-based flexibility analysis of FH-10-12, using three approaches whereby either one or both linkers were defined as flexible, supported the inference that the 10-12 region forms a rather rigid system. This is in agreement with NMR-derived relaxation analysis conducted on the two pairs; no signs of inter-domain flexibility were observed on the time-scales probed by NMR.

The combination of the NMR structures of FH-10-11, 11-12 and the previously solved and SAXS-validated 12-13 resulted in models which were subsequently used for the SAXS rigid-body based modeling of the larger 11-14, 10-

15 and 8-15 regions. Specifically, modeling of the SAXS data of FH-11-14 using the 11-13 NMR model was in agreement with previous findings (43).

There were two possible models existing to date for the FH-10-15 region, from which model I was selected as the more realistic one following modeling using the NMR-derived 10-13 model. In this model one end is clearly more compact than the other (*i.e.* 10-11 *vs.* 14-15, figure 5.15. B), however up to now it was unclear which end was which. Since the NMR data for the "10-12" end are not suggestive of such a compact arrangement on that part of the molecule this clearly leaves the possibility that the "13-15" end is the one in which modules are tightly associated. In this optimized conformation of this region FH-10-13 adopts a relatively open arrangement followed by a much more compact conformation of FH-14 and 15, which actually wrap around FH-13; ultimately the three modules form a highly compacted supermodule.

Furthermore, EOM analysis of the FH-8-15 region using the "10-13 fixed" approach, (*i.e.* wherein all linkers excluding 10-11, 11-12 and 12-13 are treated as flexible), resulted in a predominantly two-state system. Since FH-10-15 appears compact and lacks flexibility this result would imply that FH-8-9 acts as a flexible end to the 8-15 region. What is more, the SAXS-derived parameters of FH-8-9 (Table 5.1) (indicative of a standard bi-module dumbbell shaped structure) reinforce the resultant unusual orientation of the FH-13-15 end in the central region of FH. For further considerations see DISCUSSION Section 7.1.

CHAPTER 6

SPR-based functional studies of the interaction of FH with PspC of *S. pneumoniae*

Overview

Successful pathogens must possess means of evading attack by the complement system (67). Some have evolved strategies to exploit the complement-regulatory abilities of host proteins (67). Many of these sequester or capture FH (69), that presumably then acts to prevent C3b amplification on the microbial surface. One such pathogen is *S. pneumoniae* (69). One of the ways by which it presumably escapes complement-mediated destruction is by expressing the gene for PspC, a surface-exposed protein and virulence factor that binds to FH (70), (84), (85).

Previous studies showed that the N-terminal region of PspC denoted PspCN is required for FH binding (88). In addition, an ITC measurement of the binding of full-length human FH (purified from plasma) to a 104-residues polypeptide representing the PspCN region yielded a $K_D < 1$ nM (88). Inconsistent data have been reported in the literature regarding which regions of the large (20-CCP) FH molecule are responsible for binding to PspC. In various studies, binding to one or more of several regions, including 6-10, 13-15, 8-11 & 19-20, have been reported (87), (90), (91).

It may be that microorganisms simply capture FH via its central or C-terminal modules such that FH modules 1-4 are able to engage with C3b and prevent it reaching the surface. Alternatively, or additionally, it could be that FH is held in such a way as to emulate its presumed conformation on a self-surface that – according to some hypotheses – renders it tenfold more potent. To investigate this issue, and thereby also potentially shed light on the mechanism of this key complement regulator, it was decided in the current work to pinpoint the regions of FH responsible for binding to the recombinant proteins PspCN (and also of PspCNR1- the N-terminal region of PspC, *i. e.* PspCN, but extended to the first of the two R repeats, see also figure 6.1-C). A further aim was to investigate the potential interaction of the complexes – both FH-PspCN and FH-PspCNR1 - with various (activated and inactivated) fragments of C3.

For these purposes, it was decided to perform surface plasmon resonance (SPR). The proteins used in the current work, were either recombinantly produced, or

purified from human plasma. In the latter case they were provided by Complement Technology (CompTech).

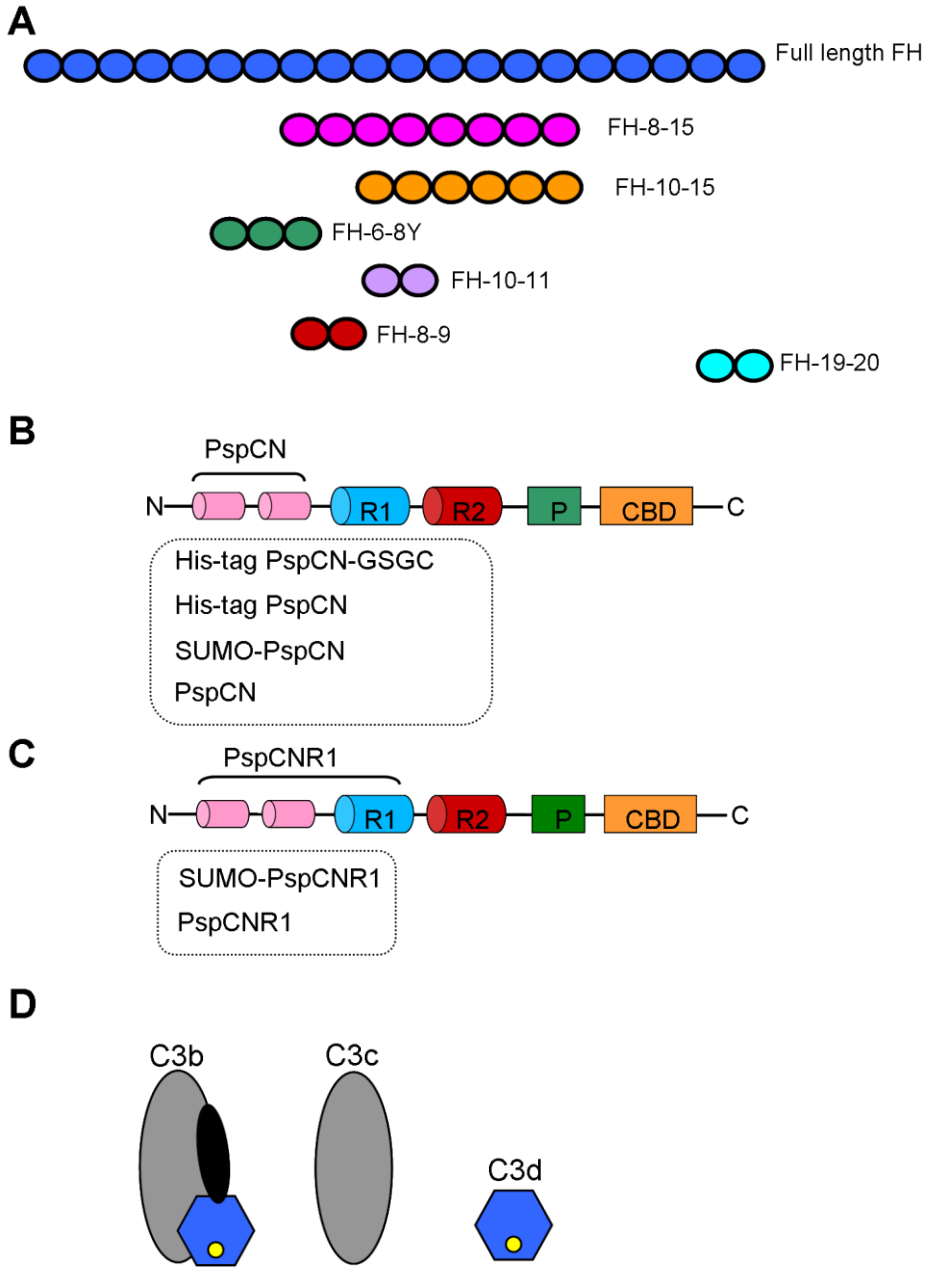


Figure 6.1 Summary of proteins employed in the current study of interactions between (fragments of): PspC, FH and C3. (A) Representation of full-length FH and the constructs used. Each oval represents a CCP module. Full-length FH is shown in blue and the FH segments 8-15, 10-15, 6-8Y, 10-11, 8-9 and 19-20 are shown in various colours. (B), (C) Representations of PspC in which the N-terminal region, the R1 and R2 repeats, the Proline-rich region (P) and the Choline-binding domain (CBD) are highlighted, and the various tagged and fusion versions listed (see text); (B) The fragments of the N-terminal region, designated PspCN, (C) The fragments designated PspCNR1. (D) Representation of the C3 fragments used for the current study; C3b, C3c and C3d.

For the binding studies of complement proteins with PspC fragments, three sets of SPR experiments were performed. In each experiment different versions of PspCN/PspCNR1 were used (see Fig. 6.1-B, -C), these being: His-tagged PspCN-GSGC (*i.e.* with the His tag at the N-terminus and a C-terminal artefact, GSGC, that was present for reasons described in Chapter 3 section 3.3), His-tagged PspCN (with no C-terminal artefact), a fusion of PspCN with N-terminal His-tagged SUMO (*i.e.* His-tag SUMO-PspCN), native PspCN obtained after cleavage of the N-terminal SUMO fusion partner, His-tag SUMO-PspCNR1, and finally native PspCNR1 obtained from cleavage of SUMO-PspCNR1. SUMO fusion proteins were produced by utilizing the pE-SUMOPro Kan vector (which is associated with higher yields and solubility) (107).

6.1 Binding of FH and FH fragments to His-tag PspCN-GSGC

This experiment was performed primarily to further map the FH regions responsible for binding to PspCN. In addition, it served to demonstrate data quality and reproducibility when the nitrilotriacetic acid (NTA) sensor chip was employed *i.e.* when immobilisation depended on the binding of the His sequence at the N-terminus of the PspCN –GSGC- fragment to an NTA-chelated Ni^{2+} atom on the chip. The affinity of the His-tag:Ni^{2+} interaction, *i.e.* K_D = 1 μM, was deemed adequate for most studies while this approach has the advantage over amine coupling that the His-protein molecules are immobilised in a uniform way (163). Values of K_D between FH constructs and immobilised PspCN-GSGC were not determined in this exploratory experiment, but the levels of response obtained (which correlate with the mass of analyte bound to the sensor chip surface) allowed a semi-quantitative comparison.

Thus, the NTA chip was charged with Ni^{2+} (see Section 2.5.1), then ~400 RUs of His-tagged PspCN-GSGC were immobilized, and the analytes (FH fragments) were flowed over the loaded chip surface. It should be noted that no successful regeneration conditions could be identified, therefore, after each cycle of charging - immobilizing – and analyte injection, the chip surface was stripped with sequential washes of EDTA, NaOH and SDS (see section 2.5.2).

In an initial experiment, all FH samples were injected in duplicate at concentrations of 0.5 μM. A good negative control would probably have been a recombinant version of FH-1-4 since there are no reports that this region interacts with PspC. The version however of FH-1-4 available in the lab had a C-terminal His-tag and was thus unsuitable. Figure 6.2 shows the sensorgram obtained for this experiment.

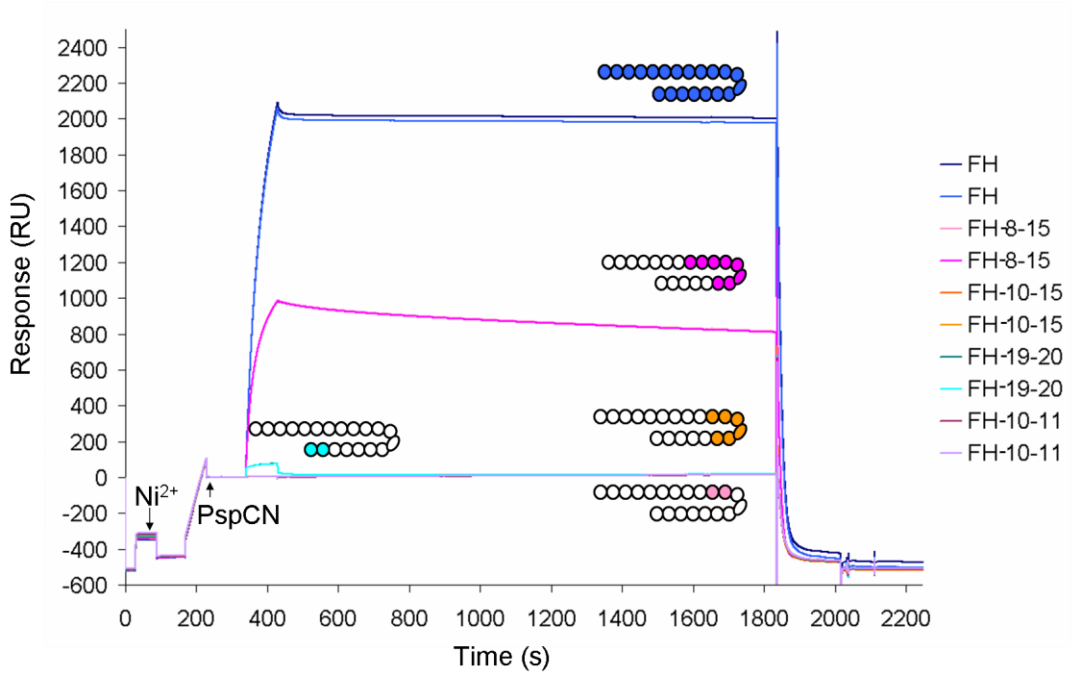


Figure 6.2 Surface plasmon resonance experiments to identify the PspCN binding regions of FH. Colour-coded background-subtracted duplicate traces are shown, see text for details. The parts of each trace corresponding to the charging of the NTA chip with Ni^{2+} as well as the immobilization levels of PspCN are also indicated on the figure.

The sensorgrams (Fig. 6.2) show that despite the need to strip the chip after each cycle of charging, immobilization and injection, reproducibility is excellent. Full-length FH shows a rapid association rate (k_{on}) followed by an extremely slow, practically undetectable, dissociation rate (k_{off}); this is indicative of a strong interaction with the PspCN. A similar observation was made for FH-8-15 although in this case dissociation appears slightly faster. On the other hand, FH-10-15 and FH-10-11 had much lower affinity (no detectable response at 0.5 μM) while FH-19-20 appeared to bind with low affinity (low response, fast off-rate). In summary, both full-length FH and FH-8-15 bind strongly to PspCN while, interestingly, negligible affinity exists for FH-10-15 and 10-11. These results clearly imply that FH-8-9 (*i.e.*

the difference between 10-15 and 8-15) is critical for binding to the bacterial protein fragment. As far as the weak interaction observed for FH-19-20 is concerned, this required further investigation.

6.2 Binding of FH fragments to His-tag PspCN

The previous experiment suggested that modules 8-9 of FH are critical in mediating the interaction with PspCN. In order to test this hypothesis further, the construct His-tagged PspCN, lacking the non-native C-terminal "GSGC" artefact, was used. In this experiment, FH-8-9 and FH-6-8Y were added to the list of FH segments to be tested while full-length FH and FH-8-15 were omitted. PspCN was once again immobilized through its N-terminal His-tag on the NTA sensor chip and the chip stripped between cycles as before. Once again, FH samples were injected in duplicate at 0.5 μM. The NTA chip was coated with ~90 RUs of His-tagged PspCN during each cycle. Figure 6.3 shows the results obtained.

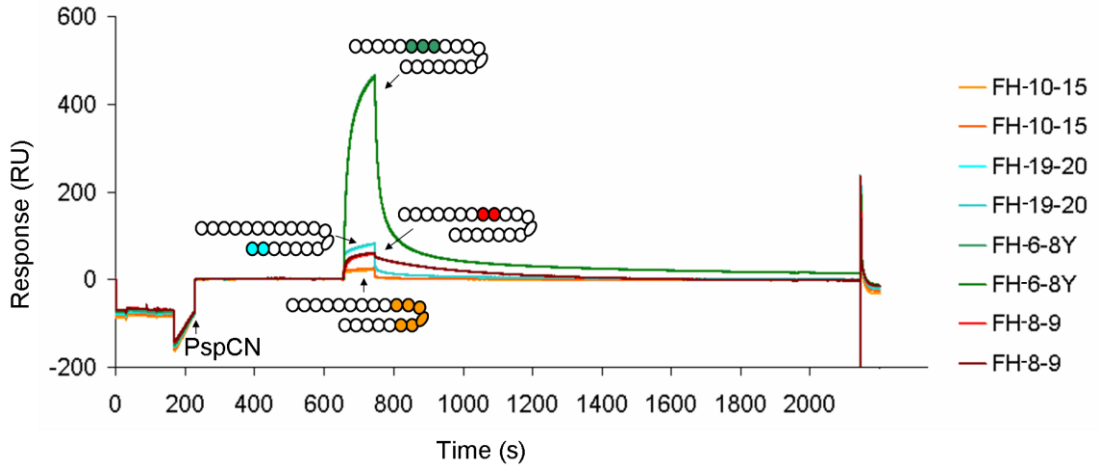


Figure 6.3 Surface plasmon resonance experiments to further investigate PspCN binding by various segments of FH. Background-subtracted duplicate traces (colour-coded) for FH are shown.

Reproducibility was, once again, very good. As before FH-10-15 exhibited negligible affinity and FH-19-20 exhibited a weak, fast on/fast off, interaction. Importantly, FH-8-9 showed a very much slower off-rate than FH-19-20. This is indicative of a strong interaction with PspCN while the lower response (compared to full-length FH or FH-8-15) is a consequence of its smaller size. FH-6-8Y exhibited a

superficially promising looking binding curve but the large response elicited suspicion. The size of a response is proportional to the number of bound molecules times their molecular mass, *i.e.* the total mass of protein on the chip surface. The theoretical analyte-binding capacity of the surface in RU is given by: RU_{max} = ligand level x MW analyte/ MW ligand x binding stoichiometry. Assuming a 1:1 binding stoichiometry and based on the corresponding MW of the ligand and the analyte and the 90 RU of ligand immobilized, the expected R_{max} for FH-6-8Y would be ~140 RU. This is clearly not the case as can be observed in the sensogram (a response of >400 RUs was obtained). This could be attributed to FH-6-8Y interacting non-specifically with the chip surface, and since background-subtracted data are shown (where the blank or background surface was not charged with Ni^{2+}), this interaction is likely to be with the Ni^{2+}. This was tested in the experiments described below (Section 6.3).

6.3 Binding of various fragments to His-tag SUMO-PspCN, His-tag SUMO-PspCNR1 and His-tag SUMO

The SUMO-fusion versions of PspC had not yet been tested for how they would interact with the FH fragments as well as with full length FH and C3b. As they would ultimately be used for the kinetic studies (in order to determine the affinity constants) with FH-8-9, FH-8-15 and full length FH (see sections 6.4-6.6) and because it was not known how these complement proteins would interact with SUMO alone, a preliminary experiment was performed in order to analyse binding profiles. His-tag SUMO-PspCN was immobilized on flow cell 2 of the NTA chip, His-tag SUMO-PspCNR1 on flow cell 3 and His-tag SUMO alone - as a control - on flow cell 4; flow cell 1 was the reference surface (and was not charged with Ni^{2+}). Full-length FH and the FH fragments 8-15, 10-15, 6-8Y, 8-9 and 19-20, as well as C3b, were then flowed over the chip. All samples were injected sequentially at 50 nM and then 500 nM (the chip was stripped in between injections) except for full-length FH, which was injected at 50 nM only. The results of this investigation are shown in figures 6.4-6.6.

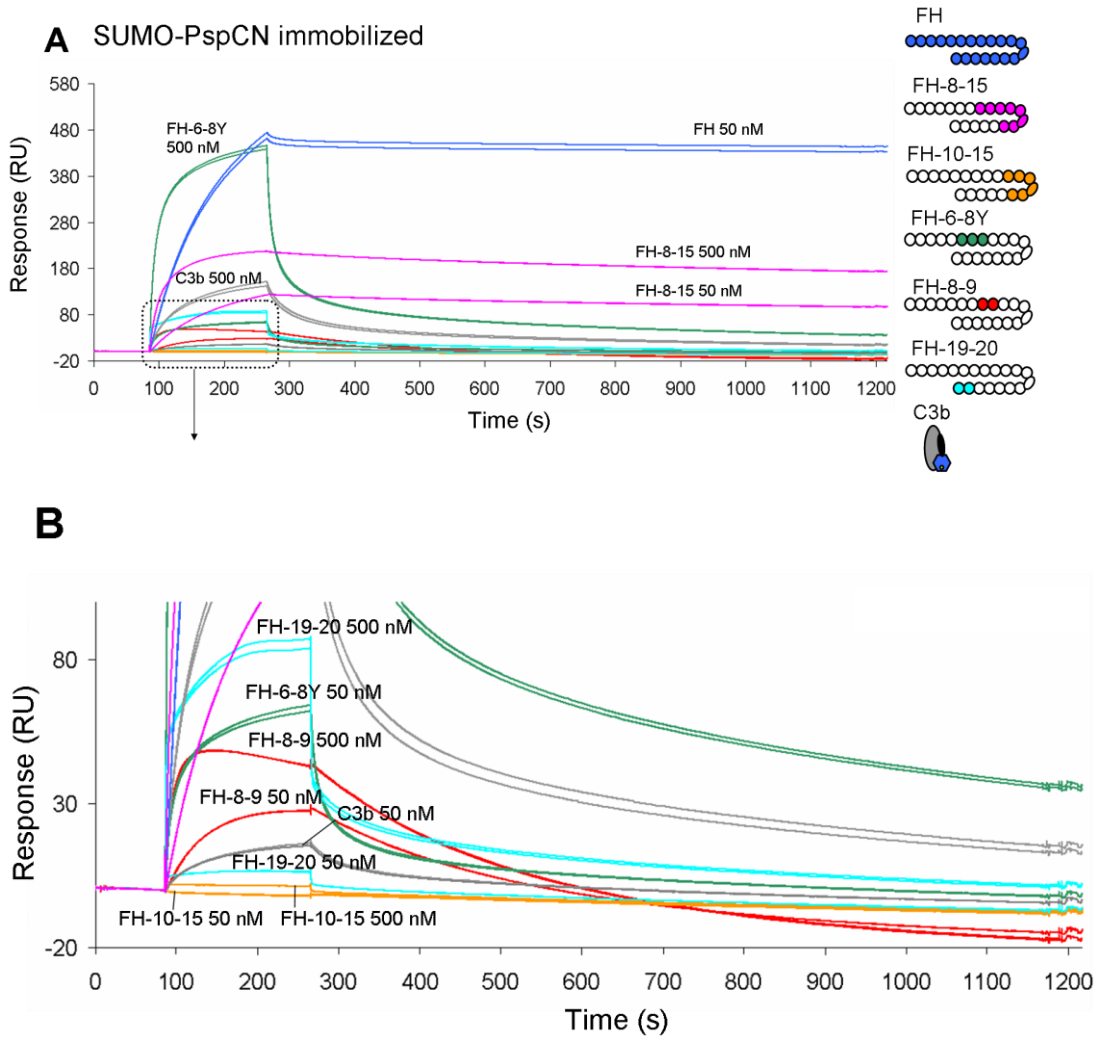


Figure 6.4 Surface plasmon resonance experiments to measure binding of various fragments to His-tag SUMO-PspCN. Duplicate traces were recorded (during a single experiment) at 50 nM and (except for FH) also at 500 nM. (A) Background-subtracted sensorgrams; traces are labeled. (B) Zoomed-in representation of selected background-subtracted sensorgrams not clearly distinguishable in (A).

In summary: (i) Full-length FH, FH-8-15 and FH-8-9 show binding profiles consistent - in terms of the strength of each interaction - with those obtained from previous experiments; thus affinity increases (*i.e.* K_D decreases) in the following order: FH-8-9, FH-8-15, FH. (ii) FH-10-15 showed no detectable interaction (as seen in previous assays); this FH segment could henceforth serve as a negative control. (iii) As before, a small but significant response was obtained with FH-19-20; (iv) a large response was observed in the case of FH-6-8Y; whether this could be attributed to binding to Ni^{2+} alone was tested further; (v) C3b also appeared to bind weakly.

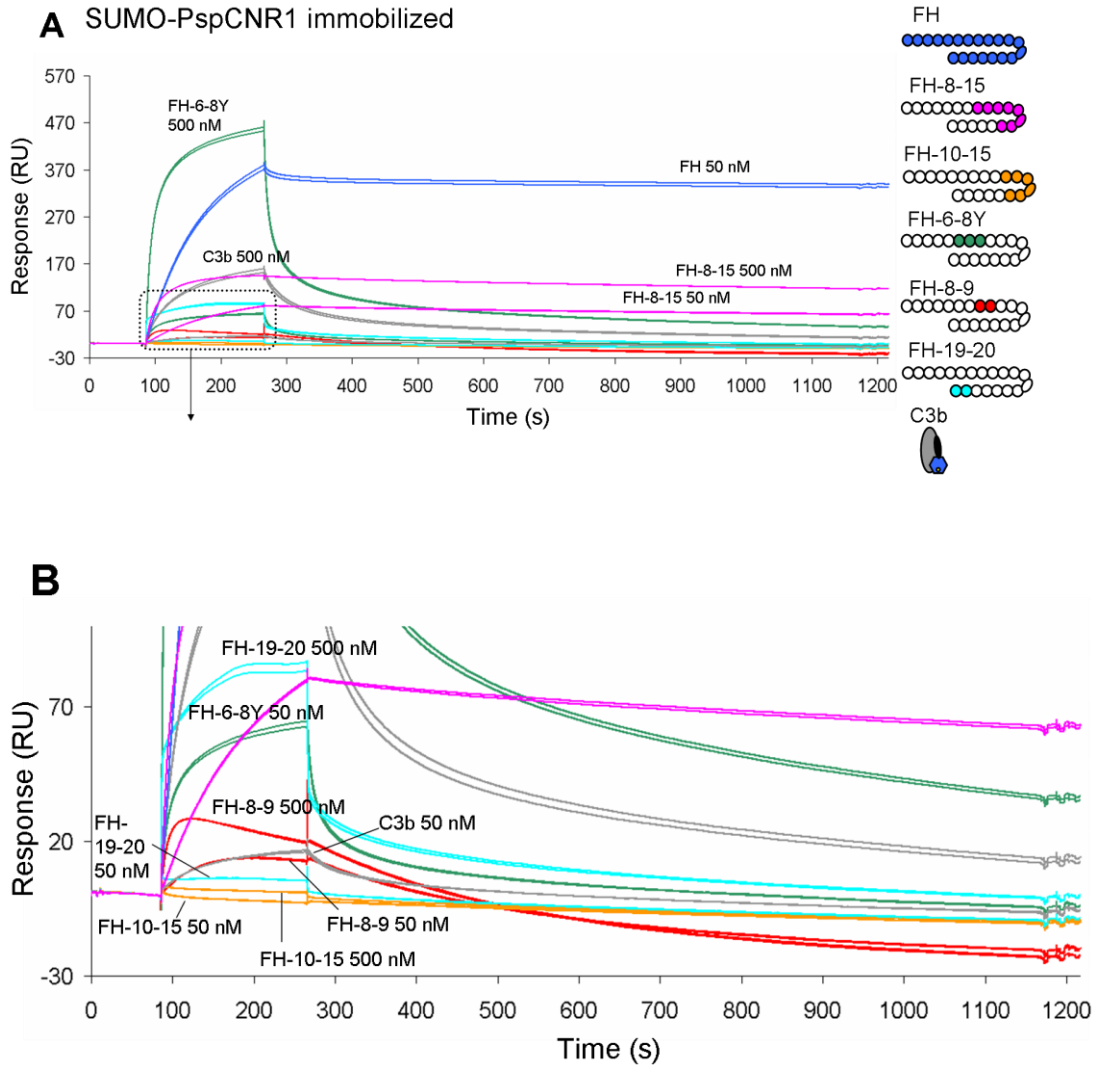


Figure 6.5 Surface-plasmon resonance experiments to measure binding to His-tag SUMO-PspCNR1. Duplicate traces were recorded - during a single-cycle experiment at 50 and 500 nM (full-length FH at 50 nM only). (A) Background-subtracted sensogram for the fragments; each curve corresponds to a sample as indicated on the figure. (B) Zoomed-in representation of selected background-subtracted sensogram not clearly distinguishable in (A).

The results from this assay were similar whether obtained with His-tag SUMO-PspCN or His-tag SUMO-PspCNR1. This implies that the R1 region does not affect the binding of PspC to the complement proteins used in this study.

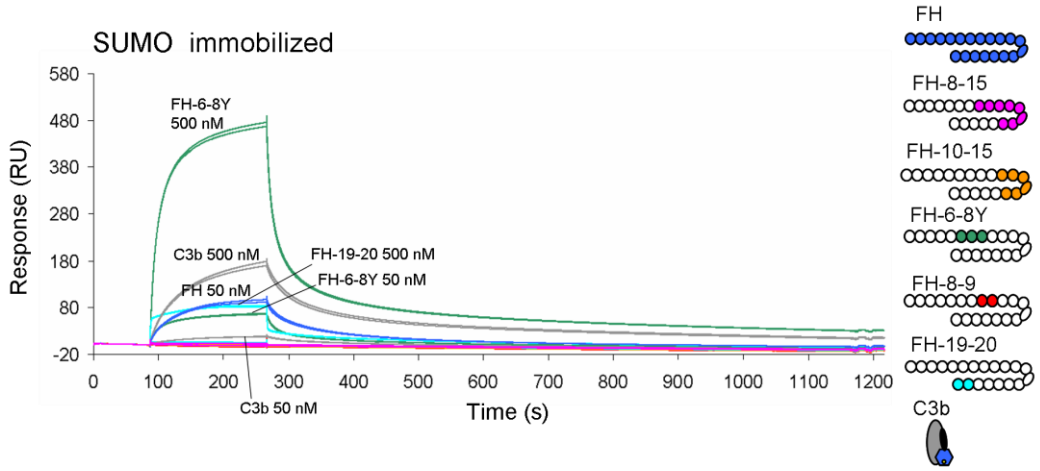


Figure 6.6 Surface plasmon resonance experiments to measure binding of various fragments to His-tag SUMO. Duplicate traces were recorded (during a single experiment) at 50 nM and (except for FH) also at 500 nM. Background-subtracted sensorgrams are shown.

It was observed that the following samples interacted with His-tag SUMO alone: FH, FH-6-8Y, FH-19-20 and C3b. In the case of FH however the profile differed significantly from that obtained when His-tag SUMO-PspCN/NR1 were immobilized as a fast off-rate and a smaller response were observed. For remaining samples, the profiles of their interaction with His-tag SUMO resemble those obtained when His-tag SUMO-PspCN/NR1 were used as ligands. This implies that either these samples interact with His-tag SUMO or with Ni^{2+} or both, rather than with PspCN/NR1.

To investigate further, the affinity of FH segments for the NTA chip charged with Ni^{2+}, in the absence of any immobilised His-tagged protein, was measured directly. For this purpose, approximately 55 RUs of Ni^{2+} were immobilized on flow cell 2 of the NTA chip and the analytes were injected at 500 nM with the exception of full-length FH that was tested at 50 nM. The resultant sensorgram is shown in figure 6.7.

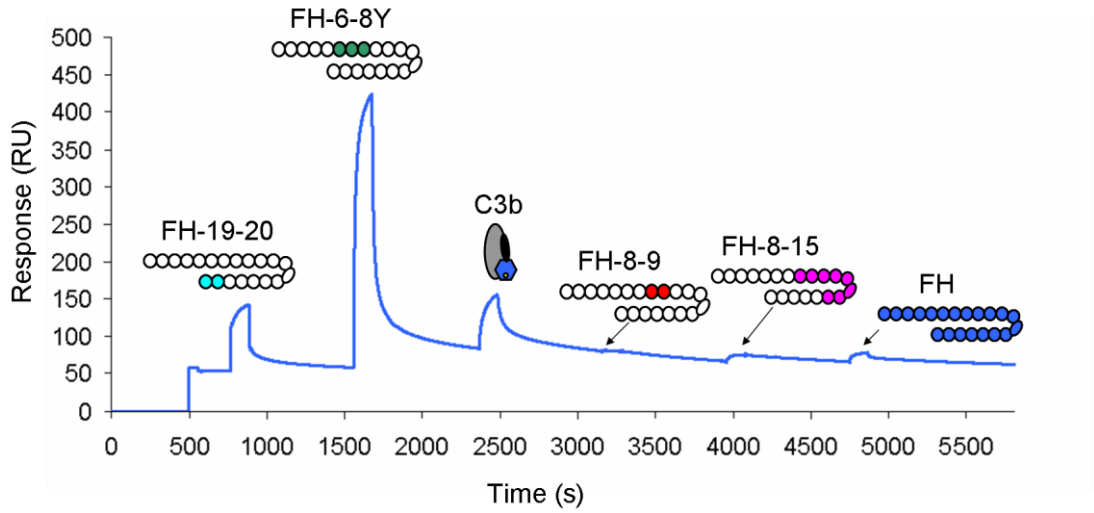


Figure 6.7 Use of SPR to investigate the binding of FH fragments to Ni^{2+}. All fragments, as indicated on sensogram, were injected at 0.5 µM concentrations over flow cell 2 of the NTA chip, which contained ~55 RUs $Ni^{2+}$with the exception of FH, which was injected at 0.05 µM.

This experiment demonstrated that FH-19-20, 6-8Y, and C3b interacted with immobilized Ni^{2+}. There is therefore no evidence that these proteins interact with His-tag SUMO or PspC. On the other hand, full-length FH, FH-8-15 and FH-8-9 showed very small, negligible or no detectable interactions with Ni^{2+}, confirming their chief interaction is with the PspC fragments. In total, these experiments established that the NTA sensor chip could be used for further kinetic and steady-state experiments, with immobilized SUMO serving as a control protein.

6.4 SPR-based studies of FH-8-9 binding to His-tag SUMO-PspCN and His-tag SUMO-PspCNR1

Based on the finding that FH-8-9 displayed strong affinity for PspCN (shown in section 6.3), further SPR was performed in order to obtain the dissociation constant. The His-tagged versions of SUMO-PspCN (125 RUs), SUMO-PspCNR1 (150 RUs) and SUMO (control) (60 RUs) were immobilised on flow cells 2, 3, and 4 respectively (flow cell 1 = reference). FH-8-9 was injected over all flow cells in a concentration series of: 0, 5 nM, 10 nM, 25 nM, 50 nM, 100 nM, 250 and 500 nM. After each injection the chip was stripped (see section 2.5.3) due to the slow off-rate of FH-8-9 (Figs 6.8-6.11).

6.4.1 Interaction of FH-8-9 with His-tag SUMO-PspCN

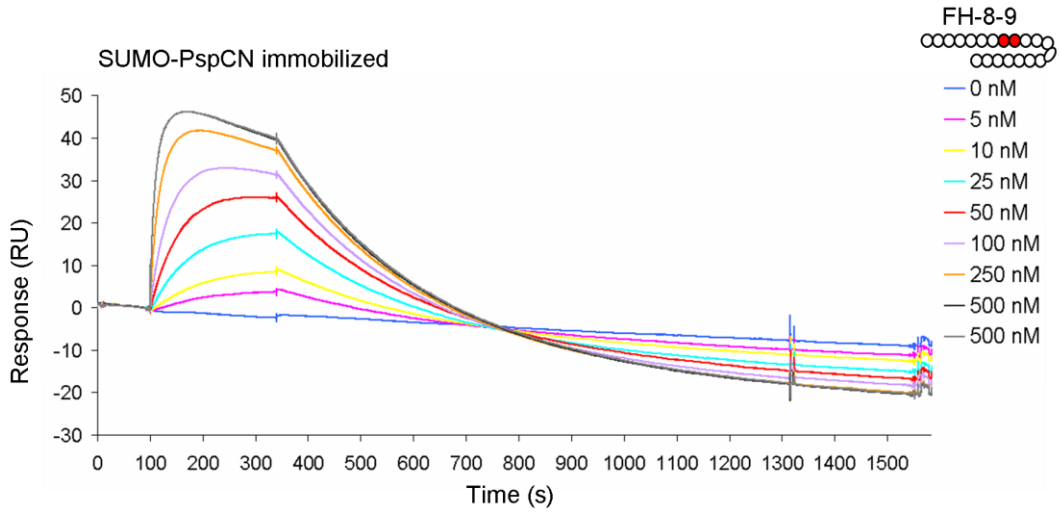


Figure 6.8 Use of SPR to measure the dissociation constant for the binding of FH-8-9 to SUMO-PspCN. Sensogram in which the background-subtracted traces are shown colour-coded for each of the concentrations flowed over flow cell 2 of the NTA chip as indicated on the figure.

The K_D obtained from this kinetic experiment (by fitting kinetic parameters from background-subtracted traces to a one-to-one Langmuir model) was 0.031 μM. The unusual shape of the curves may be explained by the possibility that the specific interactions of FH-8-9 with PspCN may be de-stabilising the interaction of the His-tag with the NTA chip, increasing the dissociation rate of PspCN from Ni^{2+} chip. This would then suggest an accelerated loss or leaching of the His-tagged PspCN from the NTA-Ni^{2+} chip surface when it is in complex with FH-8-9 resulting in the traces for the response to become negative instead of returning to baseline, and a maximal response that does not coincide with the end of the injection period (for higher concentrations of FH-8-9) as is normally observed. A steady-state affinity analysis was performed in addition to the kinetic experiments the result of which is shown in figure 6.9.

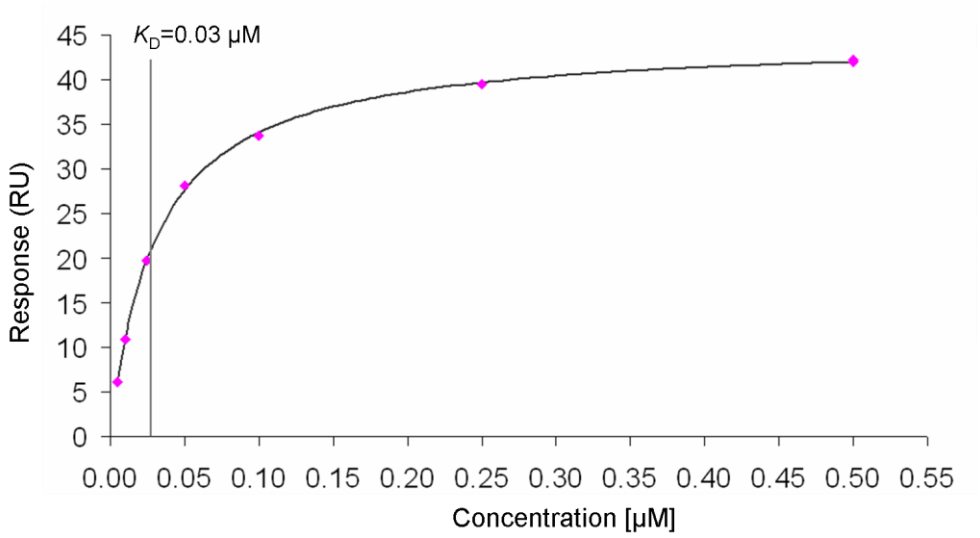


Figure 6.9 Plot of the SPR response obtained *vs.* analyte concentration after steady-state affinity analysis for FH-8-9 binding to His-tag SUMO-PspCN.

The measured K_D value from the steady-state affinity analysis was 0.03 +/- 0.0015 µM which is in excellent agreement with the value obtained from the kinetic experiments. Thus despite the atypical shapes of the profiles these measurements appeared to yield a useful and reproducible estimate of the dissociation constant.

6.4.2 Interaction of FH-8-9 with His-tag SUMO-PspCNR1

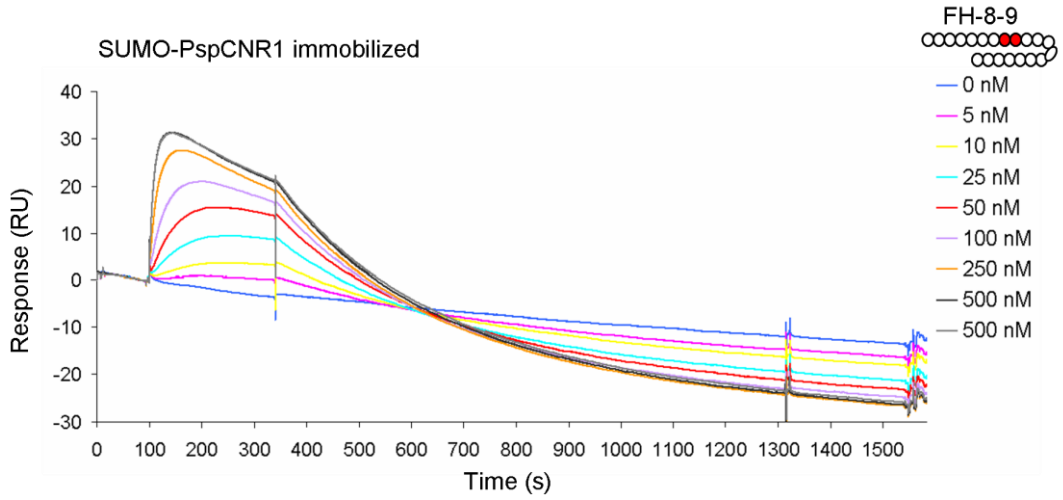


Figure 6.10 Use of SPR to measure the dissociation constant for the binding of FH-8-9 to His-tag SUMO-PspCNR1 by kinetic experiments. Sensogram in which the background-subtracted traces are shown colour coded for each of the concentrations flowed over flow cell 3 of the NTA chip as indicated on the figure.

The measured K_D value for the interaction of FH-8-9 with His-tag SUMO-PspCNR1 was 0.032 μM (obtained by fitting steady-state binding levels derived from background-subtracted traces to a one-to-one binding model). This result confirms that the R1 region of PspC does not contribute to the binding of PspCNR1 to FH-8-9. The failure of maximal response to coincide with the end of the injection time, and below-zero readings on return to baseline, are more pronounced versions of those observed in the previous experiment when SUMO-PspCN was immobilized. In addition to the kinetic experiments, steady-state affinity analysis was also performed (see Fig 6.11).

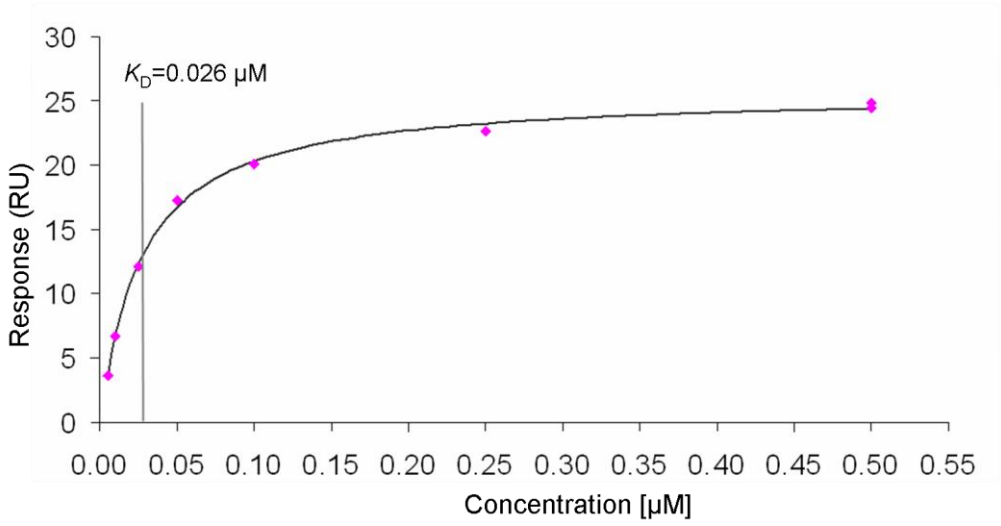


Figure 6.11 Plot of the response obtained *vs.* analyte concentration after steady-state affinity analysis for FH-8-9 binding to His-tag SUMO-PspCNR1 by SPR.

The obtained K_D value following this analysis was 0.026 +/- 0.0025 μM and is in very good agreement with the K_D value resulting from the kinetic analysis.

6.5 Kinetic studies of FH-8-15 with His-tag SUMO-PspCN and His-tag SUMO-PspCNR1

For determination of dissociation constants and kinetic parameters, a similar experimental setup to that described for FH-8-9 was employed. FH-8-15 was injected at concentrations of: 0, 0.25 nM, 0.5 nM, 1 nM, 5 nM, 10 nM, 25 nM, 50 nM, 75 nM and 100 nM (the chip was stripped after each injection) (see Figs. 6.12-6.13).

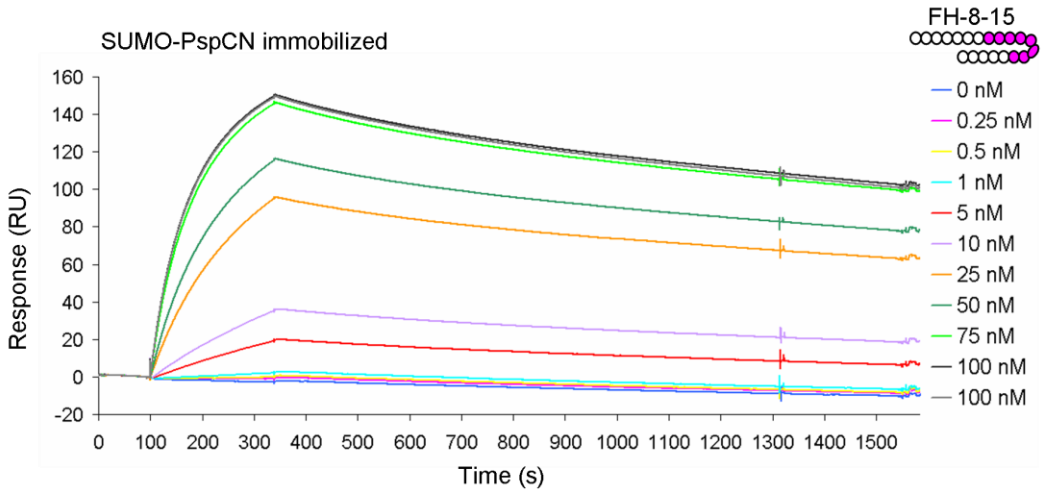


Figure 6.12 Use of SPR to measure dissociation constants for binding of FH-8-15 to His-tag SUMO-PspCN by kinetic experiments. Background-subtracted traces are shown colour coded for each of the concentrations flowed over flow cell 2 of the NTA chip as indicated on the figure.

The K_D value obtained (by fitting kinetic parameters from background-subtracted traces to a one-to-one Langmuir model) was 1.7 nM suggesting that FH-8-15 binds strongly to PspCN. Note that a disproportionate increase in RUs - compared to the remaining concentrations - is observed between 10 and 25 nM injections. The most likely explanation for this is error during the preparation of the dilutions of the samples and requires repetition of the experiment. The fitting of the parameters to a 1:1 binding model was performed by using just the five lowest concentrations (0.25, 0.5, 1, 5 and 10 nM) (data not shown), in addition to the full range presented above. The K_D value obtained following this approach was 0.75 nM.

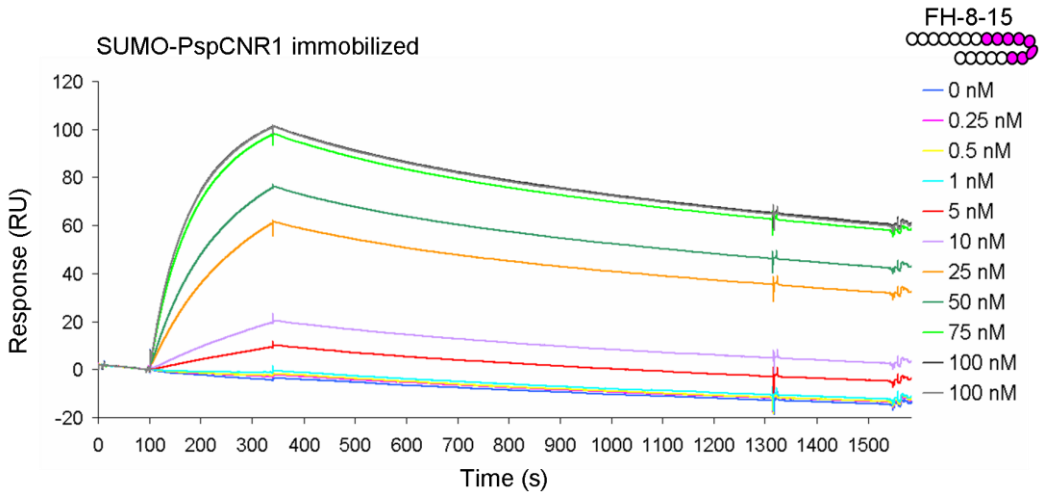


Figure 6.13 Use of SPR to measure the dissociation constant for the binding of FH-8-15 to His-tag SUMO-PspCNR1 by kinetic experiments. Sensogram in which the background-subtracted traces are shown colour coded for each of the concentrations flowed over flow cell 3 of the NTA chip as indicated on the figure.

The K_D value acquired for the interaction of FH-8-15 with His-tag SUMO-PspCNR1 was 2.3 nM. As in the case of His-tag SUMO-PspCN the fitting to a 1:1 binding model was also carried out by using the five lowest concentrations. This gave a K_D value of 1.2 nM (data not shown).

The tight interactions of FH-8-15 with PspCN/NR1 did not allow for a steady-state analysis to be performed in order to obtain K_D values (as was done in the case of FH-8-9) as in order for equilibrium to be reached a long experimental time and significant amounts of material would be required.

6.6 Kinetic studies of full length FH with His-tag SUMO-PspCN and His-tag SUMO-PspCNR1

Full-length FH has been reported to bind to PspCN with a sub-nM affinity based on ITC. To obtain an SPR-derived K_D estimate, the experimental approach previously described for FH-8-9 and FH-8-15 was followed. Samples were injected at concentrations of: 0, 0.25, 0.5, 1, 5, 10, 25, 50 and 75 nM (see Figs. 6.14-6.15).

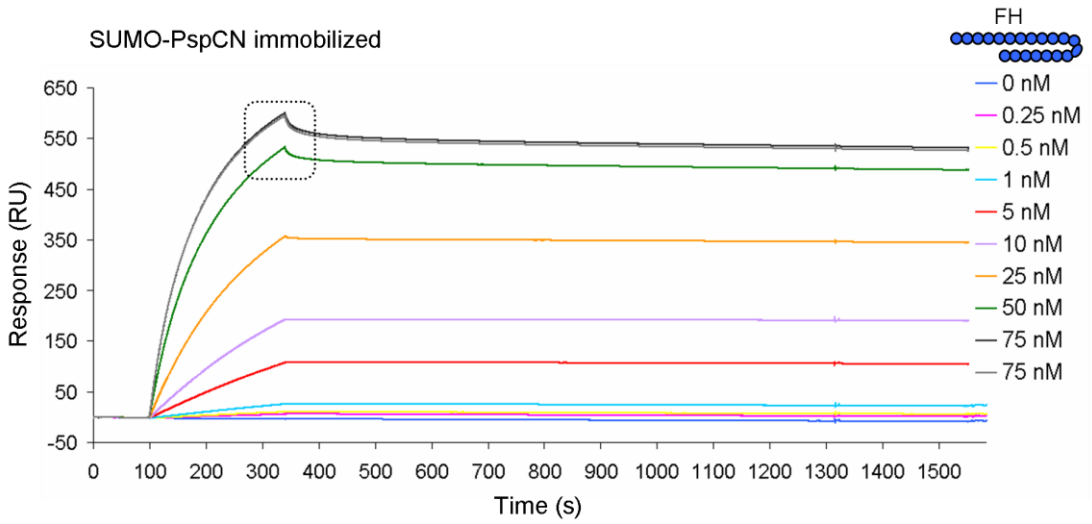


Figure 6.14 Use of SPR to measure the dissociation constant for the binding of FH to His-tag SUMO-PspCN by kinetic experiments. Sensogram, in which the background-subtracted traces are shown, colour coded for each of the concentrations flowed over flow cell 2 of the NTA chip. The "glitches" evident in traces for the two highest concentrations, 50 and 75 nM, are boxed.

The K_D value obtained (by fitting kinetic parameters from background-subtracted traces to a one-to-one Langmuir model) from this kinetic experiment was a very tight 0.15 nM. At the two highest concentrations, 50 and 75 nM, a "glitch" (highlighted with a box in figure 6.14) becomes evident corresponding to a low-affinity and transient interaction between FH and molecules or ions on the chip surface. This potentially corresponds to a small interaction with Ni^{2+} discussed earlier.

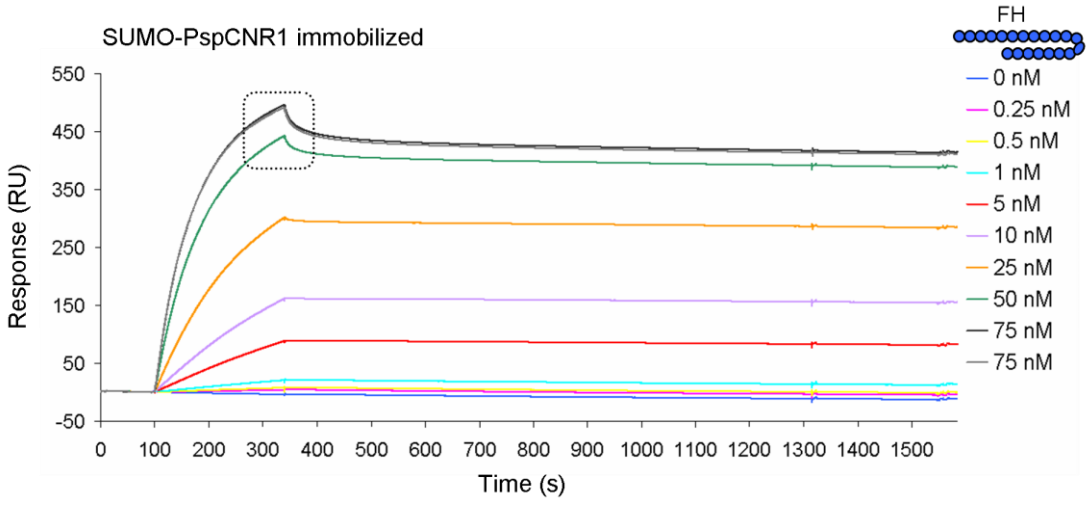


Figure 6.15 Use of SPR to measure the dissociation constant for the binding of FH to His-tag SUMO-PspCNR1 by kinetic experiments. Sensogram in which the background-subtracted traces are shown colour coded for each of the concentrations flowed over flow cell 3 of the NTA chip as indicated on the figure. For significance of the boxed region see legend to Fig 6.14.

The K_D value, according to kinetic analysis, of FH for PspCNR1 was 0.19 nM. This indicates that the R1 region of PspC essentially does not contribute to binding to FH, and is consistent with the non-participation of this region of the bacterial protein that was observed in binding FH-8-15. As in the previous experiment a putative transient and weak interaction with Ni^{2+} is manifested at the two highest concentrations (50 and 75 nM). As in the case of FH-8-15, a steady-state experiment could not be performed in order to obtain a K_D value (for reasons already mentioned, see section 6.5).

In summary, it may be concluded that FH associates very tightly with PspCN as demonstrated by a series of experiments, and the K_D of 0.15 nM. On the other hand, the importance of the interaction of FH with Ni^{2+} can be considered secondary in the context of the present work, but is of interest from the viewpoint of previously published work on the interaction of FH with bivalent ions (53), (164).

Table 6.1 summarizes the immobilization levels and K_D values obtained for the interaction of FH 8-9, FH 8-15 and FH with His-tagged SUMO-PspCN/NR1.

Ligand	Ligand loading (RU)	Analyte	K_D : Kinetic analysis	K_D +/- (SE) (μM): Steady-state affinity analysis
His-tag SUMO-PspCN	~125	FH-8-9	0.031 μM	0.030 (0.0015)
		FH-8-15 FH-8-15 low Concentration range	1.7 nM 0.75 nM	_
		FH	0.15 nM	_
His-tag SUMO-PspCNR1	~150	FH-8-9	0.032 μM	0.026 (0.0025)
		FH-8-15 FH-8-15 low concentration range	2.3 nM 1.2 nM	_
		FH	0.19 nM	_

[$](SE): Standard Error

Table 6.1 Summary of immobilization levels and K_D values obtained through kinetic and steady-state analysis for the binding of FH-8-9, 8-15 and full length FH to His-tag SUMO-PspCN and His-tag SUMO-PspCNR1.

The figure below (Fig. 6.16) shows a plot of the $1/K_D$ values obtained for the three FH segments *vs.* the two PspC constructs in a comparative manner.

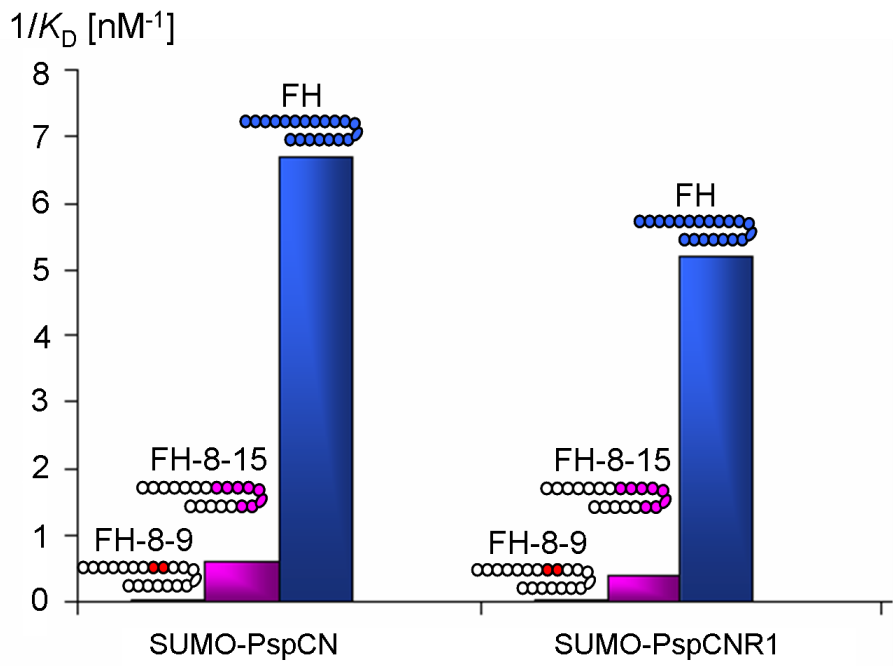


Figure 6.16 Bar-chart diagram summarizing K_D values determined for binding of full-length FH (shown in blue), FH-8-15 (shown in pink) and FH-8-9 (shown in red) to SUMO-PspCN and SUMO-PspCNR1.

6.7 Binding of FH, PspCN, PspCNR1, the FH-PspCN complex and the FH-PspCNR1 complex to C3 fragments: C3b, C3c and C3d

FH binds to C3b and this interaction is key to the protection of host cells against unwanted complement activation. It was interesting to investigate whether a complex consisting of FH-PspCN or FH-PspCNR1 retains affinity for C3b. Indeed, if PspC binds to the central region of FH, as suggested by the SPR studies, this would leave FH-1-4 and 19-20 free to interact with C3b thus helping *S. pneumoniae* to avoid being the target of C3b amplification and subsequent complement-mediated destruction. This work sets out to measure the affinity of FH-PspCN and FH-PspCNR1 complexes for C3b and for the smaller fragments generated by FI, namely iC3b, C3c and C3dg/C3d.

An experimental approach was designed on the basis of the previous observation that a 1:1 stable complex is formed between FH and PspCN/PspCNR1.

Thus C3b, C3c and C3d (all purchased by CompTech) were immobilized, via standard amine coupling (see Materials and Methods section 2.5.3), on a C1 sensor chip and dissociation constants of the interactions of FH alone, the FH-PspCN complex and the FH-PspCNR1 complex were determined by steady-state affinity experiments. The two PspC fragments used were devoid of tags or fusion partners. The immobilization levels achieved for each of the ligands (in RUs) on the carboxymethylated matrix-free C1 chip were as follows: 250 for C3b (in flow cell 2), 181 for C3c (in flow cell 3) and 37 for C3d (in flow cell 4) while flow cell 1 was the reference.

6.7.1 Binding of PspCN and PspCNR1 to C3b, C3c and C3d

In a control set of experiments PspCN and PspCNR1 were tested, on their own, for binding to C3, C3c and C3d (immobilized on flow cells 2, 3 and 4 respectvely). Figure 6.17 confirms a lack of any measurable interactions between PspCN and PspCNR1 with the C3 fragments.

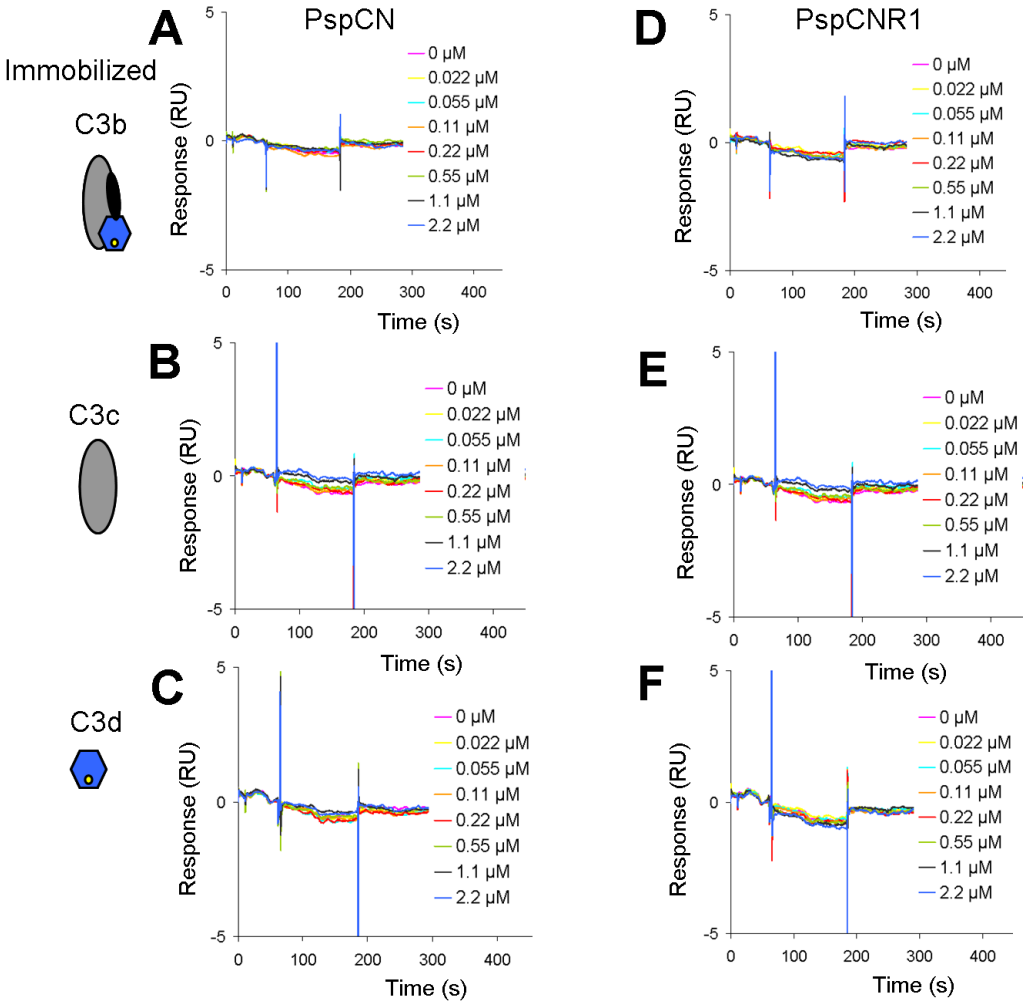


Figure 6.17 Use of SPR to confirm no appreciable binding of PspCN and PspCNR1 to the C3 fragments C3b, C3c and C3d. Background-subtracted sensorgrams are shown for PspCN and PspCNR1, colour-coded, as they were flowed at a range of concentrations over flow cells 2, 3, 4 which contained immobilized C3b, C3c and C3d respectively. (A) PspCN is flowed over C3b. (B) PspCN is flowed over C3c. (C) PspCN is flowed over C3d. (E) PspCNR1 is flowed over C3b. (F) PspCNR1 is flowed over C3c. (F) PspCN is flowed over C3d.

6.7.2 Binding of FH and the complexes FH-PspCN, FH-PspCNR1 to C3b

C3b was immobilized on flow cell 2 of the C1 sensor chip as previously described. An initial experiment with non-complexed FH is shown in figure 6.18.

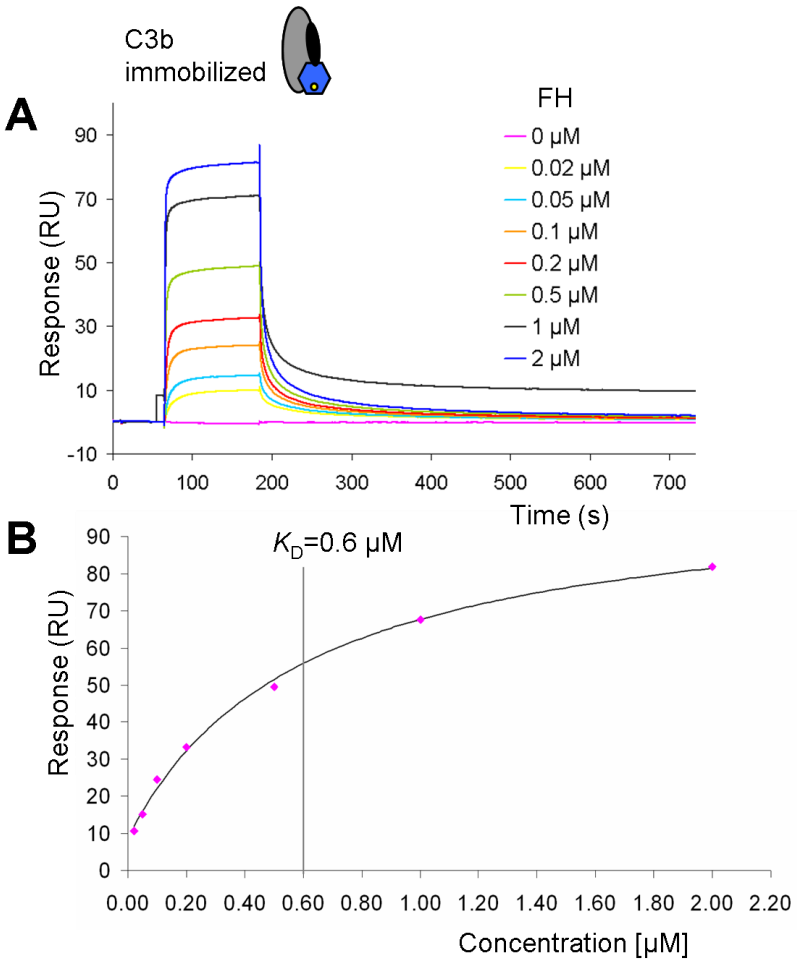


Figure 6.18 Use of SPR to confirm the dissociation constant for the binding of FH to C3b. (A) Background-subtracted sensogram is shown colour-coded for FH at a range of concentrations (0-2 µM). (B) Plot of the response obtained *vs.* analyte concentration fitted as described in Materials and Methods. The resultant K_D value is indicated on the plot.

The K_D value of 0.6 +/- 0.09 µM obtained is similar to that measured previously (0.59 +/- 0.04) µM when a C1 chip was used (30).

Subsequently a mixture containing a 1:1.1 ratio of FH to PspCN was prepared. A series of concentrations were flowed over the immobilised C3b yielding the response shown in figure 6.19.

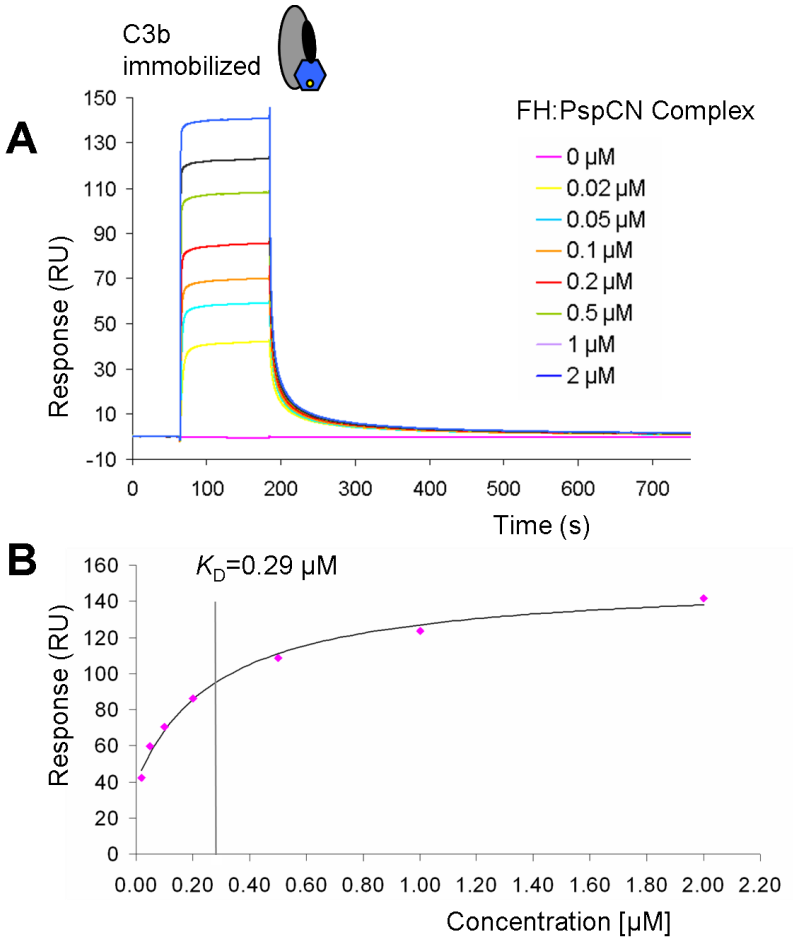


Figure 6.19 Use of SPR to measure the dissociation constant for the binding of the FH-PspCN complex to C3b by steady-state affinity. (A) Background-subtracted sensogram is shown colour-coded for the complex at a range of concentrations (all containing a 1:1.1 ratio of FH to PspCN). (B) Plot of the response obtained *vs.* analyte concentration. The resultant K_D value is indicated on the plot.

The K_D value for this interaction was determined to be 0.29 +/- 0.07 µM and is about two times stronger than the one observed for FH alone with C3b.

This measurement was repeated with the FH-PspCNR1 complex, as shown in figure 6.20.

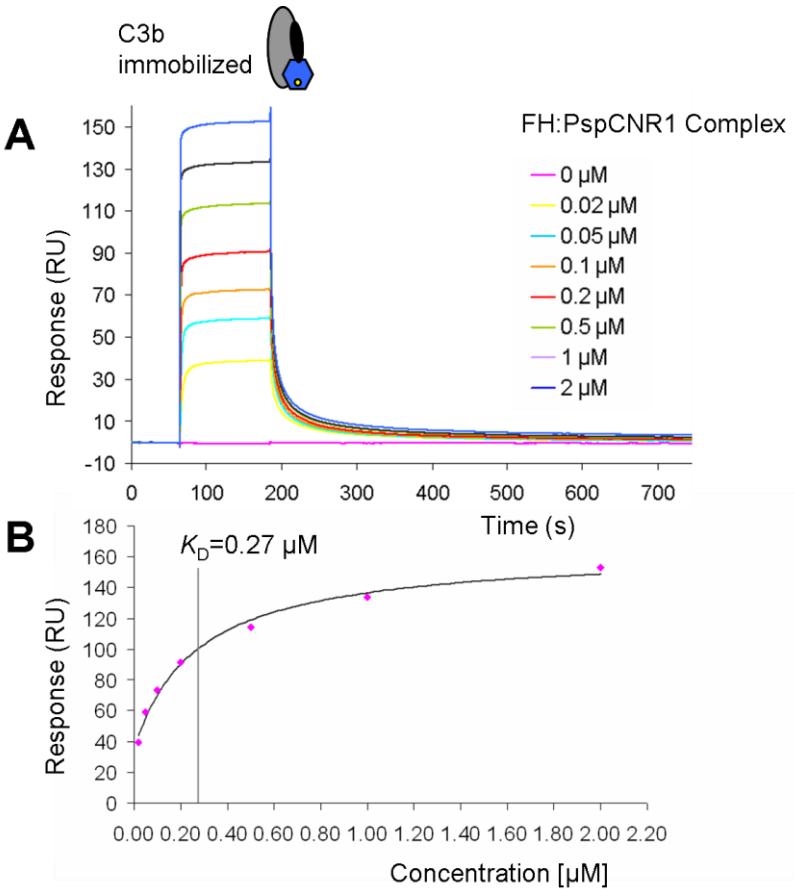


Figure 6.20 Use of SPR to measure the dissociation constant for the binding of the FH-PspCNR1 complex to C3b by steady-state affinity. (A) Background-subtracted sensogram is shown colour-coded for the complex at a range of concentrations (all containing a 1:1.1 ratio of FH to PspCNR1). (B) Plot of the response obtained *vs.* analyte concentration. The resultant K_D value is indicated on the plot.

6.7.3 Binding of FH and the complexes FH-PspCN, FH-PspCNR1 to C3c

The concerted action of FI and FH results in the cleavage of C3b to form inactive iC3b. This becomes further cleaved (by FI and CR1) into C3c and C3dg – the latter is cleaved, by one of several proteases, to give C3d. C3d remains tethered by to the surface while C3c is released. All previous studies have reported no appreciable binding of FH to C3c but there are mixed reports on the affinity of FH for C3d. The most recent results from this lab (C. Q. Schmidt, personal communication) indicate that FH does not bind to C3d (unlike the fragment FH-19-20).

In the first experiments, the affinity for C3c of the FH-PspCN and FH-PpsCNR1 complexes were compared with one another and with free FH (Fig. 6.21).

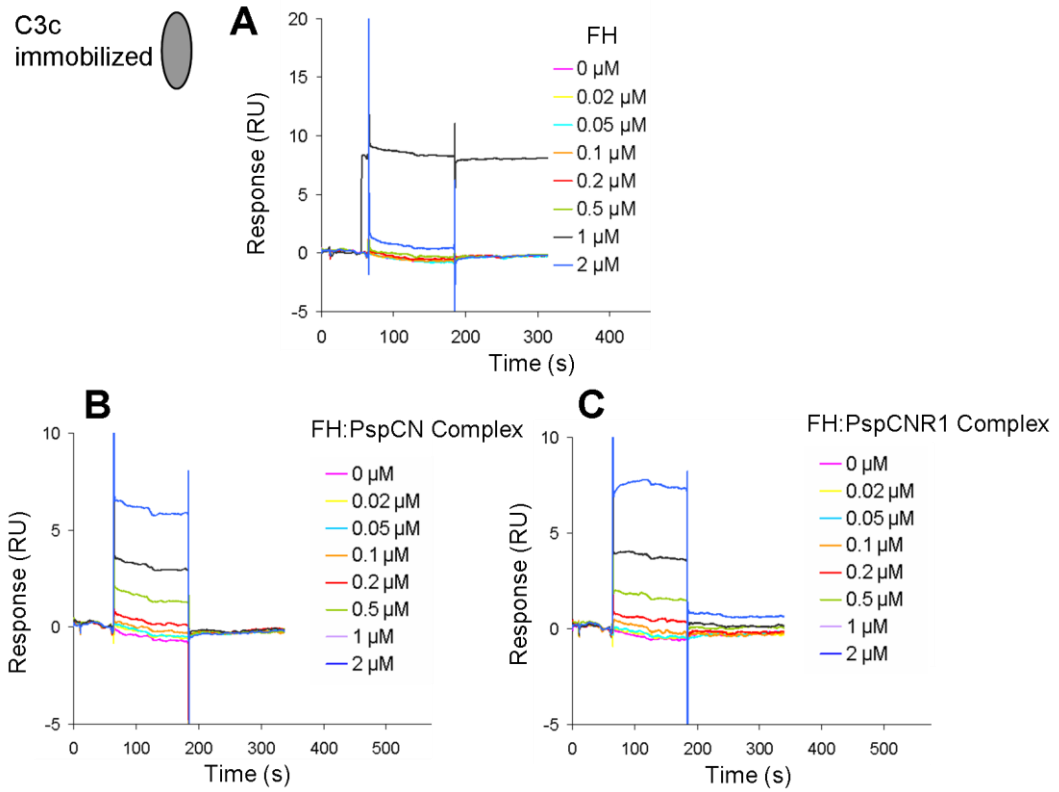


Figure 6.21 Use of SPR to compare the binding of FH, FH-PspCN and FH-PspCNR1 to immobilised C3c. (A) Background-subtracted sensogram (colour-coded) for FH over the concentration range 0-2 µM. (B) Background-subtracted sensogram is shown colour-coded for the complex FH-PspCN at a range of concentrations as indicated (with a 1.1:1 ratio of PspC to FH as before). (C) colour-coded for the complex FH-PspCNR1 at a range of concentrations as indicated (with a 1.1:1 ratio of PspC to FH as before). No appreciable binding to C3c was observed for FH alone or any of the complexes.

Consistent with previous findings FH does not seem to interact with C3c, and the same applies to the complexes FH-PspCN and FH-PspCNR1.

6.7.4 Binding of FH and the complexes FH-PspCN and FH-PspCNR1 to C3d

As alluded to above, only a weak interaction had been previously observed for FH binding to the C3d fragment in this lab. This is despite the fact that FH-19-20 alone can bind well to C3d and indeed has been crystallised in complex with both wild-type and a mutant version of FH-19-20. C3d corresponds almost exclusively to the thioester-containing domain (TED) and thus remains attached to a target surface, probably for days after the initial complement-activating event. Since the complexes FH-PspCN and FH-PspCNR1 bound with a higher affinity than does FH to C3b, as

shown in section 6.7.2 (above), the possibility of FH within the complex also interacting with C3d in a different manner to FH was investigated (see Figs. 6.22-6.24).

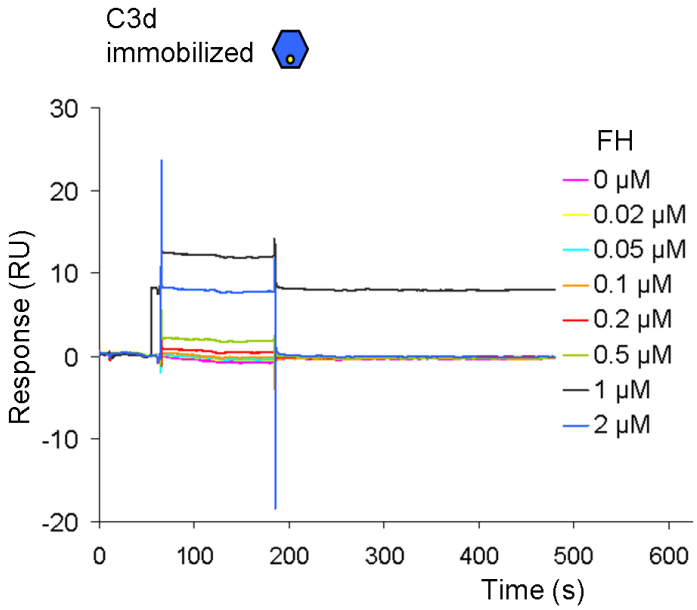


Figure 6.22 Use of SPR to study the binding of FH to C3d. Background-subtracted sensogram is shown colour-coded for FH at a range of concentrations (0-2 µM). C3d was immobilized on flow cell 4 of the C1 sensor chip.

In agreement with the aforementioned previous findings of our lab, no substantial interaction is observed, by SPR, between FH and C3d.

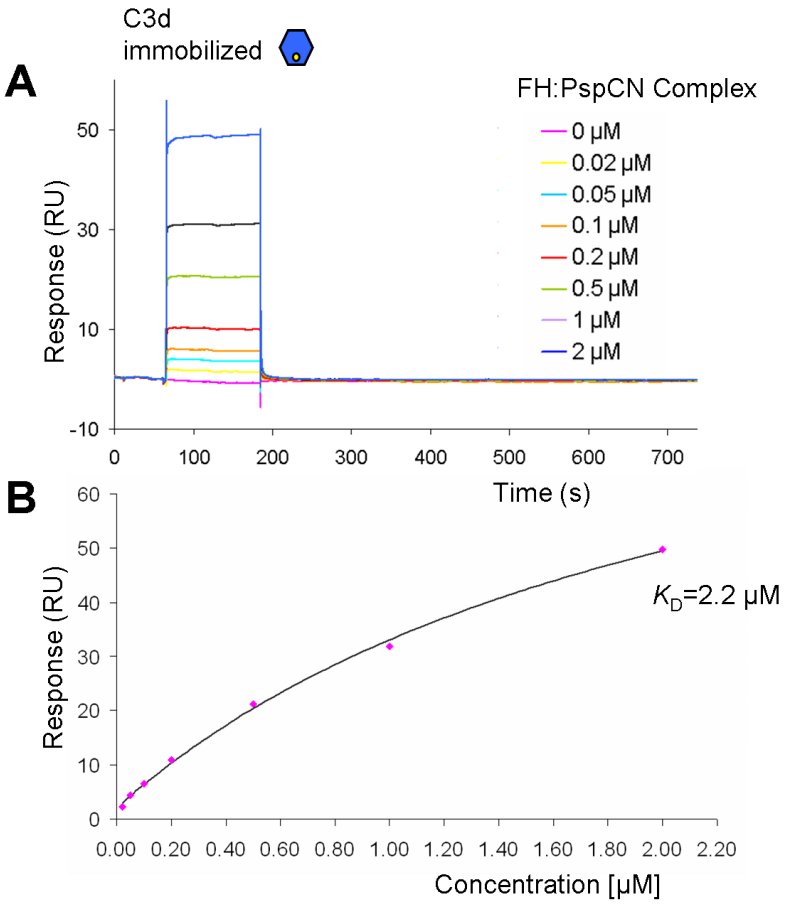


Figure 6.23 Use of SPR to measure the dissociation constant for the binding of the FH-PspCN complex to C3d by steady-state affinity. A 1:1.1 stoichiometry was used in preparing the samples of FH-PspCN. (A) Background-subtracted sensorgrams are shown (colour-coded according to concentration of complex) as shown. (B) Plot of the response obtained *vs.* analyte concentration. The resultant K_D value is indicated on the plot.

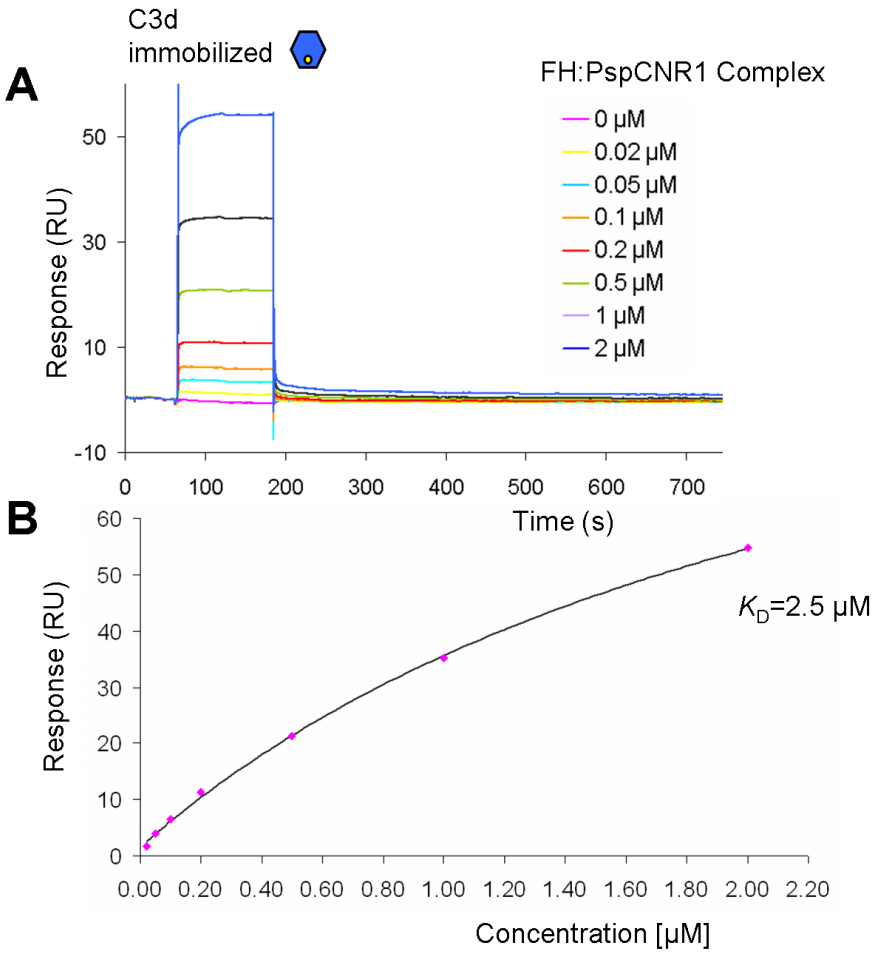


Figure 6.24 Figure 6.27 Use of SPR to measure the dissociation constant for the binding of the FH-PspCNR1 complex to C3d by steady-state affinity. A 1:1.1 stoichiometry was used in preparing the samples of FH-PspCNR1. (A) Background-subtracted sensorgrams are shown (colour-coded according to concentration of complex) as shown. (B) Plot of the response obtained *vs.* analyte concentration. The resultant K_D value is indicated on the plot.

The results clearly demonstrate the ability of both complexes, *i.e.* FH-PspCN and FH-PspCNR1 to bind C3d with K_D values of approximately 2.2 +/- 0.37 - 2.5 +/- 0.32 µM respectively. Since FH or PspC alone do not display such behavior, this result implies that binding to PspCN or PspCNR1 (presumably via its central modules) exerts an allosteric effect that alters the binding properties of the N- and/or C-terminal ends of FH.

These results together suggest that free, soluble, human FH has negligible affinity for the C3c and C3d parts of C3b although the FH involved in the original cleavage could presumably remain bound to iC3b (which corresponds to C3b, C3d and C3f) for some time via its C-terminus (only). In contrast, the FH-PspCN/NR1 complexes (presumably sequestered on the bacterial surface) can bind (presumably

via C-terminal modules) to the C3d portion of iC3b. Moreover, these complexes bind slightly more tightly (K_Ds of ~0.3 μM), compared to free FH (K_D of ~0.6 μM), to intact C3b *prior* to cleavage. This difference in affinity might be considered significant since measurements of the K_D for the FH-C3b interaction have been consistently in the region of 0.6-2.2 μM depending on the sensor chip used and the different C3b loading on the chip (30). The abovementioned observations are consistent with each other if one assumes that the bacterial protein alters the architecture of FH to reveal additional binding sites that are normally inaccessible.

The table below (Table 6.2) summarizes the K_D values obtained from the above analysis.

K_D +/- (SE)[$], (μM)			
	C3b	C3c	C3d
FH	0.6 (0.09)	-	-
FH-PspCN complex	0.29 (0.07)	-	2.25 (0.37)
FH-PspCNR1 complex	0.27 (0.07)	-	2.5 (0.32)
PspCN	-	-	-
PspCNR1	-	-	-

[$]Standard Error

Table 6.2 Summary of the K_D values obtained for the interaction of FH, PspC and the FH-PspC complexes with the C3 fragments.

6.8 Conclusions

The protein PspC confers on *S. pneumoniae* the ability to evade complement-mediated destruction by promoting sequestration of the complement regulator, FH (70). The principal binding sites on FH for its major ligands – C3b, the C3b.Bb complex, FI and glycosaminoglycans – all lie within the N-terminal seven CCPs, and the C-terminal four CCPs, out of the 20 CCPs that comprise FH (24), (26), (27), (29),

(30), (31), (32), (33), (40). The remaining modules of FH – *i.e.* 8-18 - appear to be devoid of any recognition sites for complement proteins although these modules may participate in binding to C-reactive (165), (166) and they contain eight N-glycosylation sites. In previous chapters and in published work (43) it was demonstrated that the central modules *i.e.* 10-13 feature strongly tilted modules and could form a hinge or strong bend in the structure of FH.

In vivo, cleavage of C3b by FI in the presence of FH and other cofactors results in the formation of first iC3b and, ultimately, the C3dg and C3d fragments. All these fragments contain the thioester domain (TED) and can remain covalently attached to the host surface, while the larger C3c fragment is released upon cleavage of iC3b to C3dg. Successive cleavages create or expose nascent protein-protein interaction sites. Previous studies (15), (57) suggested that FH binds to a C3b molecule (or possibly two C3b molecules) via simultaneous engagement of FH-1-4 with the MG domains (and the CUB and TED domains) of C3b, and of FH-19-20 with the TED only. The CUB domain is cleaved by FI, which is consistent with loss of binding of FH-1-4 to iC3b (31). The TED remains intact in all fragments, implying that FH could (after cleavage) remain bound transiently via its 19-20 domains that also have selective affinity for surrounding GAGs. But dissociation of FH from C3d (or iC3b), may normally be a virtually irreversible process since results presented here, and previous unpublished observations, suggest no or miniscule re-binding by intact FH to C3d. Thus the affinity for C3d resident in FH-19 and 20 must be masked or hidden in some way within intact FH. This could be important in regulating FH action and so warrants further study.

The work described in this chapter led to the following conclusions: FH and some of the FH fragments appear to interact with Ni^{2+}, complicating the exploitation of the metal affinity of His tags for immobilisation on SPR chips. The most pronounced effect was exchibited by FH-6-8. This is interesting from the point of view that a study showed that certain transition metals cause FH to aggregate and that this was attributed mainly to the FH-6-8 region. Specifically, the most important effect was observed with zinc and copper, followed by nickel, cadmium and iron (magnesium and calcium had little or no no effect). In addition, the aggregation caused was accompanied by loss of FH activity (53), (164).

Importantly, the present results clearly showed that full-length FH can bind to the N-terminal domain of PspC in an effectively irreversible manner. Notably, both FH-8-15 and FH-8-9 exhibit high affinity for PspCN and PspCNR1 while FH-6-8Y, FH-10-11 and FH-10-15 do not. The binding of FH-8-15 is – like FH – characterised by an extremely slow off-rate. The C-terminal CCP-module pair, FH-19-20, binds only weakly to PspCN. However it also appears to interact significantly with immobilized Ni^{2+} which makes it difficult for a clear conclusion to be drawn regarding the possibility that it could contribute to binding *via* an avidity effect in the context of full length FH.

The strength of the interaction between full length FH and the N-terminal region of PspC determined in this study is of comparable magnitude to the ITC-based study (K_D values for both are in the low-nM range) (88).

The results presented in this study may be further compared to the literature (however taking into account that the FH fragments used in most cases differed). In previous reports, ELISAs indicated that the FH-8-15 region was solely responsible for mediating interaction (91) which supports the tight interaction of FH-8-15 for PspCN/NR1 observed in this report. Ligand dot blotting localized the PspC on FH inclusively between modules 6-10 (90) while SPR suggested FH-8-11 (87) as the main site. Both of these studies support the strong interaction measured for FH-8-9 and PspCN/NR1 in the present study. This is in contrast to another study that concluded that fragments FH-8-10 and FH-8-12 lacked the ability to mediate binding (91). Finally, in a separate study it was shown that FH-11-15 did not show binding (87). This backs-up the lack of affinity of the FH-10-15 region in the present investigation.

In the next sets of experiments the interactions of the FH-PspCN/NR1 complexes with the C3 fragments resulting from the action of cofactors and FI were studied. Native FH showed affinity for C3b described by a K_D of 0.6 (+/- 0.09) μM on a C1 chip, but only minor interactions with C3c and C3d. Importantly, the FH-PspCN/NR1 complexes bound more tightly than free FH to C3b, with K_Ds of about 0.28 μM (+/- approximately 0.07) - as opposed to the 0.6 μM affinity exhibited by free FH (see above). While, as expected, the complexes did not bind to C3c, they did bind to C3d with K_D values in the region of 2.4 μM. This intriguing observation -

taken together with the previous work showing FH-19-20, but not full-length FH, binds to C3d - implies that the bacterial protein is able to bind in such a way as to reveal a cryptic binding site on FH for the TED portion of C3b.

CHAPTER 7

Discussion

7.1 The central modules of FH are arranged in a unique way that is important for biological activity

A range of evidence (15), (43), (49), (50) suggests that FH adopts a bent-back conformation such that binding sites on its N and C termini can engage cooperatively with the same molecular target(s). Such a bent-back conformation would be facilitated *via* the flexibility afforded by the 19 inter-modular linkers and would be stabilized by the extent and the nature of contacts between the 20 modules. While long-range contacts between distant CCPs may be important for FH architecture, the current work focuses on more local interactions amongst the central modules that lack binding sites for other complement proteins. This focus was dictated by the technical difficulties associated with working with full-length FH, which has never been crystallized, is probably flexible and dynamic, and is too large and complex to yield easily interpretable scattering patterns. We hypothesized that the central modules represented a little-explored region of the molecule that is probably critical for the dynamic architecture of FH. We further speculated that they may have a specific hinge-like role, allowing interchange between distinct conformations of FH in which a part-time cryptic binding site is either hidden or exposed. The ability to switch between such conformations could be important for the role played by FH as a self-surface selective regulator of complement.

The NMR-derived solution structures of CCP modules 10, 11 and 12 of FH exhibit typical CCP characteristics. They approximate to beta strand-rich prolate spheroids that have N- and C-terminal residues lying at opposite poles of the long axes. The pairs FH-10-11 and FH-11-12 are connected through six-residue long linkers. These, far from introducing flexibility as was originally thought likely, act as rather rigid spatial organizers. In both cases, the module-pairs adopt a fixed "V-like" conformation. Thus both pairs have defined tilts, of about 90-100^{o}, stabilized by a mini-hydrophobic core that is buried at the intermodular interface. Data collected using SAXS independently validated the two NMR-derived structures. Of note, a neighboring (overlapping) pair, FH-12-13, adopts a broadly similar conformational arrangement (43). FH has been reported to be pH- and salt concentration-sensitive (52). In this respect it is noteworthy that His sidechains occupying equivalent

positions appear in both the FH-10-11 and FH-12-13 interfaces. Furthermore a putative salt-bridge, ^{624}Lys-^{675}Asp occurs in FH-10-11 which may stabilize the corresponding interface.

The multitude of NOEs that support the abovementioned "V-like" structures of recombinant FH-10-11 and FH-11-12 tally with the lack of flexibility inferred from the NMR-relaxation studies at the particular timescales probed (ps, ns and microsecond). The resultant assumption that these bimodules would adopt the same compact and rigid bent structures in the FH-10-12 context was validated by SAXS analysis. The studies of FH-10-11, 11-12 and 12-13 culminated in the reconstruction of the FH-10-13 structure. When the three pairs are superimposed on their mutual modules (*i.e.* 10-11 and 11-12 on FH-11 and 11-12 and 12-13 on FH-12) they adopt an out-of-plane "S-like" conformation of length approximately 53 Å, and represent the most strongly tilted succession of four CCPs described to date. Its overall arrangement, with limited flexibility, supports the concept that its middle region will keep FH oriented in one or more preferred conformations so as to allow for the N- and C-terminal ends to be appropriately positioned in order to perform their regulatory functions.

In the light of these new data and through revisiting previous data that had originally been interpreted in terms of two possible arrangements of modules 10-15 (43), one model emerged as being more plausible. In this model (Fig. 7.1), modules 10, 11, 12 and 13 are maintained in the NOE and SAXS-determined open "S-shaped" trace with 90-100^{o} intermodular tilts. Despite these large tilt angles, these modules nonetheless form the more extended end of the FH-10-15 unit. At the other end, SAXS suggests that an 180^{o}-bend is imposed between modules 13 and 14 while modules 14 and 15 appear to wrap around module 13 and the 12-13 linker. Thus modules 12, 13, 14 and 15 together form the more globular end of the FH-10-15 unit.

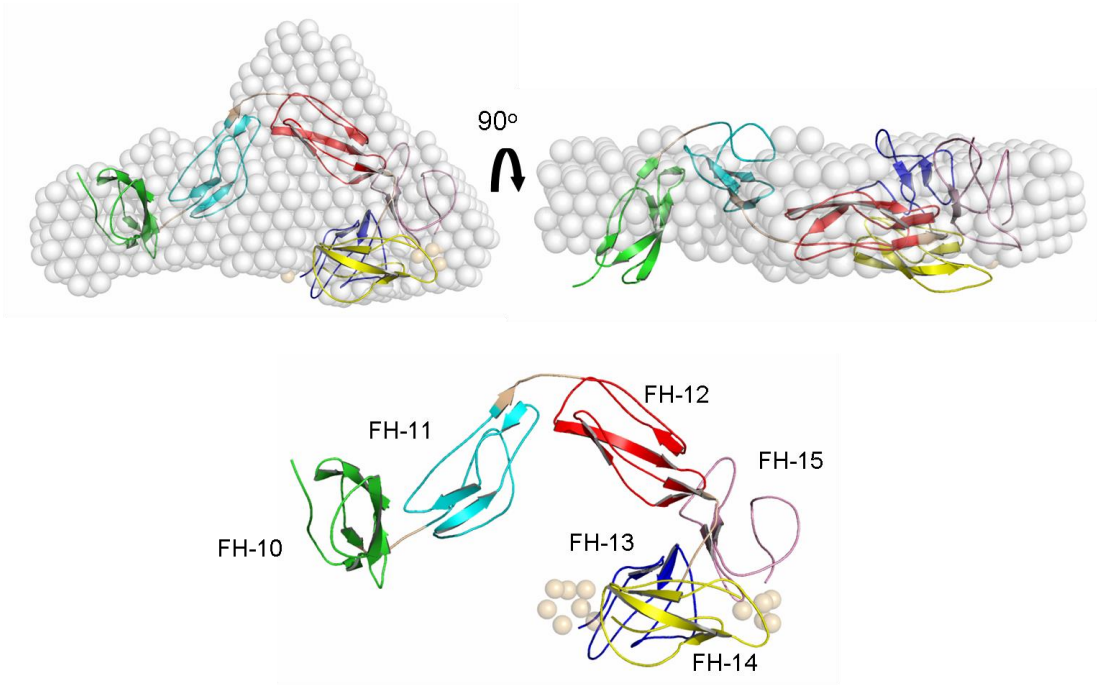


Figure 7.1 Upper panel: the most plausible model for the FH-10-15 region as emerged from SAXS and the NMR structures of FH-10-11, 11-12 and 12-13. Representation of the SAXS-derived envelope is shown in spheres and the modelled CCPs within showed as cartoons. The experimentally determined linkers from the NMR-derived FH-10-13 model are shown in beige whereas light-orange spheres represent linker residues modelled as a chain of dummy atoms. Two views are shown following 90° rotation about the x-axis. Lower panel: the FH-10-15 modelled region is shown in cartoon representation without the SAXS envelope.

Overall the FH-10-15 appears to lack much intermodular flexibility. The unusual positioning of module 14 (and also 15) with respect to the other modules may reflect the possibility that its stability depends upon one or more neighboring modules and/or linkers. In this respect it is noteworthy that previous attempts to produce folded and intact FH-13-14 and FH-14-15 samples proved unsuccessful. In addition, the FH-13-14 linker is the most variable amongst orthologues while FH-13 lacks the D-E loop and β-strand H whish are normally crucial for a stable interface to be formed with the subsequent CCP. These are in contrast to, for example, the structurally independent FH-8-9 pair that is easy to produce and in which both modules is properly folded under a range of conditions. An EOM analysis of FH-8-15 suggests these modules 8 and 9 correspond to a semi-flexible protrusion of FH-10-13, in stark contrast with FH-14-15. There is an example in the literature of a CCP module that is intrinsically disordered when produced as part of a recombinant bimodule (167). Finally, there are two examples in the literature of strand-swapping between adjacent CCPs creating fused modules (168), (169). In summary there is

literature precedence for CCP modules that are not autonomously folded and it is possible that FH-14 fits into this category. What is more, the existence of some higher order structure in the FH-13-15 region was inferred from a functional study aiming to locate the FH binding domains for PspC of *S. pneumoniae*. In this study FH-8-15 bound significantly followed, in magnitude, by FH-8-13. Interestingly FH-8-14 appeared to show decreased binding compared to FH-8-13 which would imply that module 14 was obstructing the binding site. This was no longer the case once FH-15 was present- in the FH-8-15 context (91).

Taking everything together a more sophisticated scenario can be envisaged regarding how FH may operate in a biologically relevant context. Assuming that the compact arrangement observed in *P. pastoris*-produced FH-8-15 and FH-10-15 is retained in the context of the full-length glycoprotein present in human plasma, then we may conclude that the central region of FH forms a unique structural unit. Moreover, assuming that our "S-shaped" model of FH-10-13 (built from concatenating module pairs and SAXS data for FH-10-12) is present in FH-10-15, then FH-13, 14 and 15 must be very compact indeed and it is reasonable to speculate that module 14 needs its neighbours for stability. The physiological role of this arrangement is perplexing. One possible scenario is that under certain conditions module 14 becomes dissociated from its neighbours and thereupon behaves like an intrinsically disordered domain (held together, presumably, by its disulfides) creating a flexible segment and thus allowing the central region to act as a hinge and FH to switch between closed and open conformations. In the former case (closed) FH would bind with relatively low affinity to C3b and C3d while in the latter (open) it could bind bivalently and with high affinity to a C3b molecule (shown in figure 7.2). The present study is likely to reflect the closed case (*i.e.* FH has weak affinity for C3b). More investigation is required to test whether flexional stress, or exposure to the electronegative environment found on a self-surface bearing the appropriate polyanionic markers, disrupts the 13-14 interface. Under these circumstances, module 14 would be likely to become intrinsically disordered, creating a very flexible segment within the FH molecule and allowing terminal regions to cooperate in binding to a target. The figure below (Fig. 7.2) summarizes the abovementioned points.

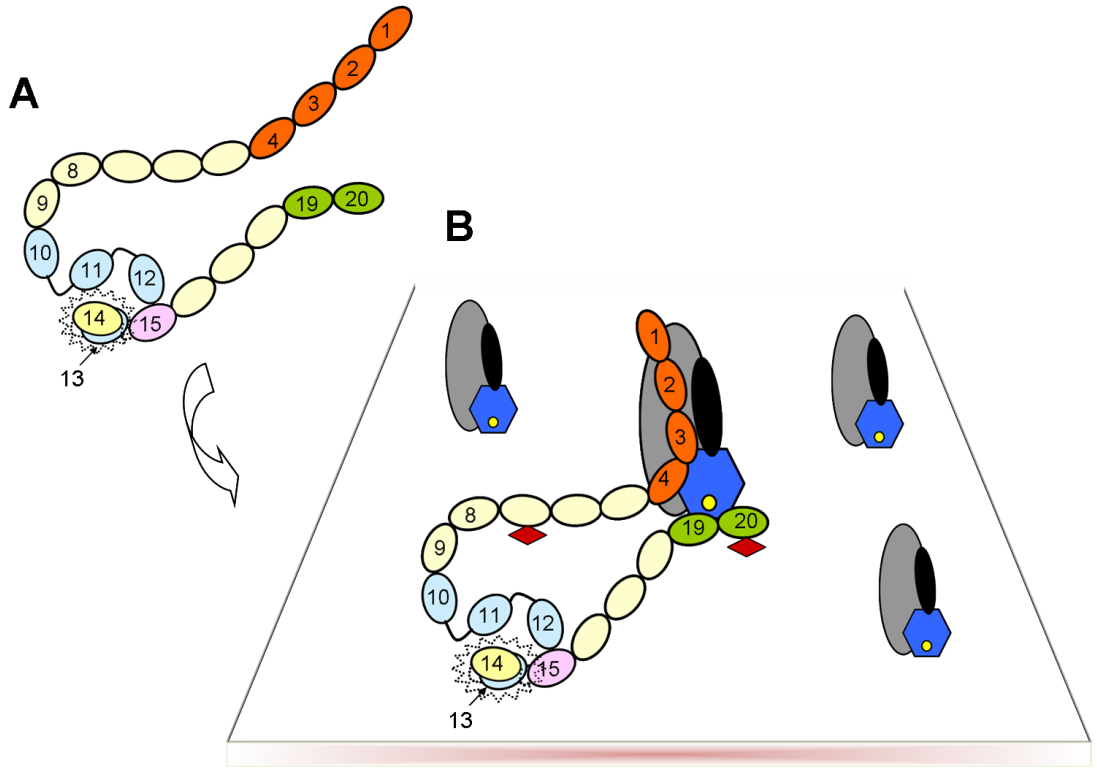


Figure 7.2 Representation of the FH molecule free in the closed conformation (A) and binding to surface-tethered C3b (anchored via its TED domain) in the context of self-surface bearing polyanionic markers (B), in the open conformation. In (B) FH-1-4 associate with the C3c-like part of C3b and the CUB domain while FH-19-20 bind to the C3d part of C3b (which equates to the TED domain). FH-7 and 20 associate with polyanionic markers (indicated with red diamonds). This mode of binding is facilitated by the FH-10-15 region in which all CCPs are shown in the optimal spatial organization-as determined through NMR and SAXS. A star highlights FH-14 that is putatively disordered and confers on the central region of FH the hinge-like ability thus allowing FH to adopt preferred and stabilised orientation(s) on a surface for optimal biological function.

7.2 The binding of PspC of *S. pneumoniae* to the central region of FH affects its interaction with C3b

The surface protein PspC of *S. pneumoniae* sequesters FH as a means to avoid immune killing. A proposed mechanism of action for FH of a separate study (92) was that FH binds to pneumococci through domains 8-11 and 19-20. The regulatory FH-1-4 region is then free to perform cofactor activity and thus evasion of complement by pneumococci takes place. The C-terminal region of FH can interact with polyanionic host-surface markers which promotes adhesion following which pneumococci exploit a FH-integrin complex for evasion into epithelial cells. Another

study using fragments corresponding to the 8-10, 8-12, 8-13, 8-14 and 8-15 FH regions concluded in the importance of both FH-13 and 15 in mediating the interaction with PspCN (91).

In the present study it was found that full length FH bound to PspCN and PspCNR1 of *S. pneumoniae* in a virtually irreversible manner. The possibility of there being more than a single binding site emerged from the fact that the full-length protein bound better than FH-8-15 (which still binds very tightly); however this requires further investigation. The FH-8-9 fragment also had a high affinity for the bacterial protein (although still lower than that of either FH or FH-8-15). Significantly, FH-10-15 did not show any discernible binding. The R1 domain of PspC appeared not to contribute to the interactions. This is all in agreement with two other separate studies (87), (88).

Thus the following can be concluded: PspC binds strongly through its N-terminus, primarily to domains 8-9 of FH. In the FH-8-15 context it may be that binding to FH-8-9 induces the exposure of a second binding site only available in the FH-8-15 context (given that FH-10-15 showed no binding). Alternatively, binding may be enhanced by the presence of the 9-10 linker or interface. Therefore, full-length FH binds in a practically irreversible manner mainly through modules 8-15.

The remarkably strong and largely irreversible interaction of full length FH with PspCN/NR1 meant that it was possible to study the behavior of a complex consisting of either FH-PspCN or FH-PspCNR1 together with C3b, C3c and C3d. A 1:1 stoichiometry of the FH-PspC complex had been observed in a separate study (88).

FH-PspCN and FH-PspCNR1 bound C3b about two times stronger compared to free FH (K_D values of 0.29 and 0.27 μM as opposed to 0.6 μM correspondingly). It is possible that the binding events stabilize a preferred, presumably functionally more active (as inferred from the higher affinity observed for the binding of the complex to C3b) conformation of FH, that may resemble the one adopted on a polyanionic marker-rich host surface. This possibility fits very well with the notion that FH has a closed conformation in which access to a cryptic C3b or C3d-binding site is occluded, in addition to an open conformation, stabilized by PspC, in which this binding site is accessible.

Neither the complexes, nor free FH, bound C3c (in which the CUB domain is cleaved while the TED domain is missing altogether). The former was an expected result, while the latter has already been demonstrated. Strikingly, the complexes showed significant binding to C3d with a K_D in the range of ~ 2 μM, while native FH interacts with C3d very weakly compared to FH-19-20. This observation implies that it is the C3b-binding site located in FH-19-20 that is cryptic and that PspC is able to bind in such a way as to reveal this cryptic binding site that is directed to the TED portion of C3b, or C3d. It is easy to envisage how this could be achieved given that PspC appears to bind to more than one discontinuous region of FH and so FH could effectively wrap around PspC.

The figure below (Fig. 7.3) shows a hypothetical model of PspCN-bound to FH engaging with C3b as resulted from the structural and functional assays. The FH-1-4 regulatory region remains free to operate in the complex. It is possible that PspCN mimics the effect that polyanionic markers found on self-surfaces have in terms of orienting FH in a preferred stabilized conformation (see also figure 7.2).

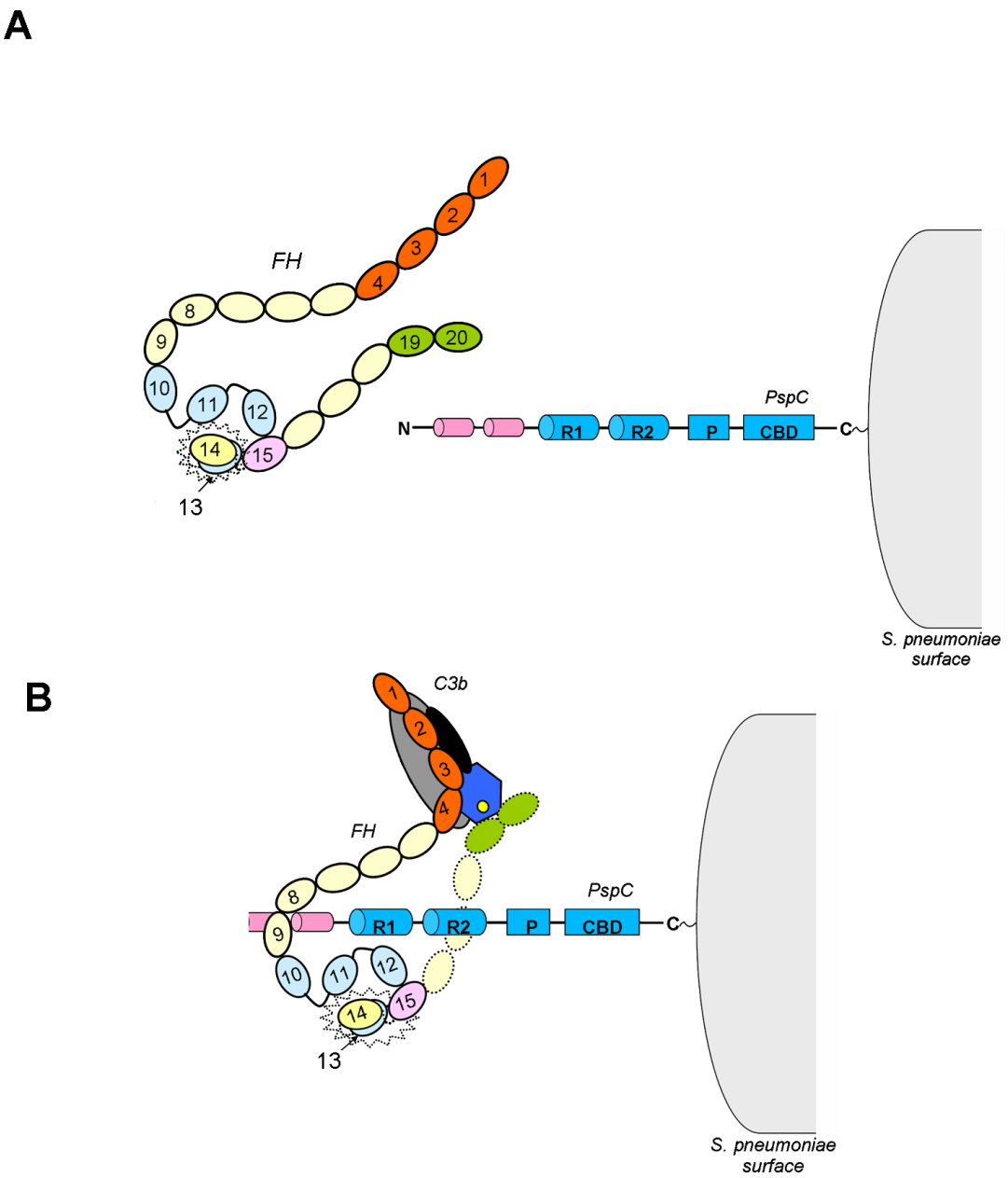


Figure 7.3 Hypothetical models of the FH and PspC molecules free and in complex. (A) FH and PspC are shown free and FH is in the closed conformation. (B) FH is shown binding to the N-terminus of PspC-following its sequestering by *S. pneumoniae* and is in a presumably preferred and open orientation, possibly the one adopted on a self-surface. Each oval represents a CCP module as indicated-FH-19-20 is shown in green. The FH-10-15 region is shown as modeled from the combination of NMR and SAXS. The regulatory FH-1-4 region is shown in complex with C3b while FH-19-20 interacts with the C3d part of C3b (*i.e.* the TED). A star highlights the putative hinge region within the central portion of FH.

BIBLIOGRAPHY

1. Nonaka, M. & Kimura, A. (2006). Genomic view of the evolution of the complement system. *Immunogenetics,* **58,** 701-713.
2. Medzhitov, R. & Janeway, C. (2001). Innate immunity. *N. Engl. J. Med.* **343,** 338-344.
3. Ricklin, D., Hajishengallis, G., Yang, K. & Lambris, J. D. (2010). Complement-a key system for immune surveillance and homeostasis. *Nat. Immunol.* **11,** 785-797.
4. Walport, M. J. (2001). Complement. First of two parts. *N. Engl. J. Med.* **344,** 1058-1066.
5. Gros, P., Milder, F. J. & Janssen, B. J. C. (2008). Complement driven by conformational changes. *Nat. Rev. Immunnol.* **8,** 48-58.
6. Pangburn, M. K. & Müller-Eberhard, H. J. (1980). Relation of putative thioester bond in C3 to activation of the alternative pathway and the binding of C3b to biological targets of complement. *J. Exp. Med.* **152,** 1102-1114.
7. DiScipio, R. G. (1992). Formation and structure of the C5b-7 complex of the lytic pathway of complement. *J. Biol. Chem.* **267,** 17087-17094.
8. Flyvbjerg, A. (2010). Diabetic angiopathy, the complement system and the tumor necrosis factor superfamily. *Nat. Rev. Endocrinol.* **6,** 94-101.
9. Janssen, B. J. C., Huizinga, E. G., Raaijmakers, H. C. A., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B. & Gros, P. (2005). Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature,* **437,** 505-511.
10. Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D. & Gros, P. (2006). Structure of C3b reveals conformational changes underlying complement activity. *Nature,* **444,** 213-216.
11. Fredslung, F., Jenner, L., Husted, L. B., Nyborg, J., Andersen, G. R. & Sottrup-Jensen, L. (2006). The structure of bovine complement component C3 reveals the basis for thioester function. *J. Mol. Biol.* **361,** 115-127.
12. Law, S. K. A. & Dodds, A. W. (1997). The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6,** 263-274.
13. Nishida, N., Walz, T. & Springer, T. A. (2006). Structural transitions of complement component C3 and its activation products. *Proc. Natl. Acad. Sci. USA.* **103,** 19737-19742.
14. Pangburn, M. K. & Müller-Eberhard, H. J. (1986). The C3 convertase of the alternative pathway of human complement. Enzymatic properties of the bimolecular proteinase. *J. Biochem.* **235,** 723-730.
15. Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Kavanagh, D., Mertens, H. D. T., Svergun, D. I., Johansson, C. M., Uhrin, D., Barlow, P. N. & Hannan, J. P. (2011). Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol.* **18,** 463-470.
16. Kirkitadze, M. D. & Barlow, P. N. (2001). Structure and flexibility of the multiple domain proteins that regulate complement activation. *Immunol. Rev.* **180,** 146-161.

17. Reid, K. B. & Day, A. J. (1989). Structure-function relationships of the complement components. *Immunol. Today,* **10,** 177-180.
18. Meri, S. & Pangburn, M. K. (1990). Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. *Proc. Natl. Acad. Sci. USA.,* **87,** 982-3986.
19. de Córdoba, S. R., Lublin, D. M., Rubinstein, P. & Atkinson, J. P. (1985). Human genes for three complement components that regulate the activation of C3 are tightly linked. *J. Exp. Med.* **161,** 1189-1195.
20. Sim. R. B. & DiScipio, R. G. (1982). Purification and structural studies on the complement-system control protein beta 1H (Factor H). *J. Biochem.* **205,** 285-293.
21. de Córdoba, S. R. & de Jorge, E. G. (2008). Translational Mini-Review Series on Complement Factor H: Genetics and disease associations of human complement factor H. *Clin. Exp. Immunol.* **151,** 1-13.
22. Weiler, J. M., Daha, M. R., Austen, K. F. & Fearon, D. T. (1976). Control of the amplification convertase of complement by the plasma protein beta 1H. *Proc. Natl. Acad. Sci. USA,* **73,** 3268-3272.
23. Ferreira, V., Pangburn, M. K. & Cortés, C. (2010). Complement control protein factor H: the good, the bad and the inadequate. *Mol. Immunol.* **47,** 2187-2197.
24. Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N. & Pangburn, M. K. (2006). Critical role of the C-terminal domains of factor H in regulating complement activation on cell surfaces. *J. Immunol.* **177,** 6308-6316.
25. Herbert, A. P., Urhin, D., Lyon, D., Pangburn, M. K. & Barlow, P. N. (2006). Disease-associated sequence variations congregate in a polyanion-recognition patch on human factor H revealed in 3D structure. *J. Biol. Chem.* **281,** 16512-16520.
26. Blackmore, T. K., Sadlon, T. A., Ward, H. M., Lublin, D. M. & Gordon, D. L. (1996). Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J. Immunol.* **157,** 5422-5427.
27. Herbert, A. P., Deakin, J. A., Schmidt, C. Q., Blaum, B. S., Egan, C., Ferreira, V. P., Pangburn, M. K., Lyon, M., Uhrin, D. & Barlow, P. N. (2007). Structure shows that a glycosaminoglycan and protein recognition site in factor H is perturbed by age-related macular degeneration-linked single nucleotide polymorphism. *J. Biol. Chem.* **282,** 18960-18968.
28. Ormsby, R. J., Jokiranta, T. S., Duthy, T. G., Griggs, K. M., Sadlon, T. A., Giannakis E. & Gordon, D. L. (2006). Localization of the third heparin-binding site in the human complement regulator factor H. *Mol. Immunol.* **43,** 1624-1632.
29. Pangburn, M. K., Atkinson, M. A. & Meri, S. (1991). Localization of the heparin-binding site in the human complement regulator factor H. *J. Biol. Chem.* **266,** 16847-16853
30. Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum B. S., Lyon, M., Uhrin, D. & Barlow, P. N. (2008). A New Map of Glycosaminoglycans and C3b Binding Sites on Factor H. *J. Immunol.* **181,** 2610-2619.

31. Wu, J., Wu, Y. Q., Ricklin, D., Janssen, B. J., Lambris, J. D. & Gros, P. (2009). Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat. Immunol.* **10,** 728-733.
32. Hellwage, J., Jokiranta, T. S., Friese, M. A., Wolk, T. U., Kampen, E., Zipfel P. F. & Meri, S. (2002). Complement C3d/C3b and cell surface polyanions are recognized by overlapping binding sites on the most-carboxyl terminal domain of complement factor H. *J. Immunol.* **169,** 6935-6944.
33. Jokiranta, T. S., Hellwage, J., Koistinen, V., Zipfel, P. F. & Meri, S. (2000). Each of the three binding sites on complement factor H interacts with a distinct site on C3b. *J. Biol. Chem.* **275,** 27657-27662.
34. Ripoche, J., Day, A. J., Harris, T. J. & Sim, R. B. (1988). The complete amino acid sequence of human complement factor H. *J. Biochem.* **249,** 593-602.
35. Esparza-Gordillo, J., Soria, J. M., Buil, A., Almasy, L., Blangero, J., Fontcuberta, J. & Rodriguez de Cordoba, S. (2004). Genetic and environmental factors influencing the human factor H plasma levels. *Immunogenetics,* **56,** 77-82.
36. Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B. & Campbell, I. D. (1991). Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.* **219,** 717-725.
37. Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B. & Campbell, I. D. (1991). Secondary structure of a complement control protein module by two-dimensional 1H NMR. *Biochemistry* **30,** 997-1004.
38. Schmidt, C. Q., Herbert, A. P., Hocking, H. G., Uhrin, D. & Barlow, P. N. (2008). Translational Mini-Review Series on Complement Factor H: Structural and functional correlations for factor H. *Clin. Exp. Immunol.* **151,** 14-24.
39. Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrin, D. & Barlow, P. N. (2008). Structure of the N-terminal Region of Complement Factor H and Conformational Implications of Disease-linked Sequence Variations. *J. Biol. Chem.* **283,** 9475-9487.
40. Pechtl, I. C., Neely, R. K., Dryden, D. T., Jones, A. C. & Barlow, P. N. (2011). Use of time-resolved FRET to validate crystal structure of complement regulatory complex between C3b and factor H (N terminus). *Protein Sci.* **20,** 2102-2112.
41. Schneider, M. C., Prosser, B. E., Caesar, J. J., Kugelberg, E., Li, S., Zhang, Q., Quoraishi, S., Lovett, J., Deane, J. E., Sim, R. B., Roversi, P., Johnson, S., Tang, C. M. & Lea, S. (2009). Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates. *Nature,* **458,** 890-893.
42. Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J., Jowitt, T. A., Clark, S. J., Tarelli, E., Uhrin, D., Barlow, P. N., Sim, R. B., Day, A. J. & Lea, S. M. (2007). Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.* **204**, 2277-2283.
43. Schmidt, C. Q., Herbert, A. P., Mertens, H. D. T., Guariento, M., Soares, D. C., Urhin, D., Rowe, A. J., Svergun D. I. & Barlow, P. N. (2010). The central Portion of Factor H (Modules 10-15) Is Compact and Contains a Structurally Deviant CCP Module. *J. Mol. Biol.* **395,** 105-122.

44. Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B. & Campbell, I. D. (1993). Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol.* **232,** 268-284.
45. Jokiranta, T. S., Jaakola, V. P.,Lehtinen, M. J.,Parepalo, M., Meri, S.&Goldman, A. (2006). Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J.* **25,** 1784-1794.
46. Morgan, H. P., Mertens, H. D. T., Guariento, M. Schmidt, C. Q., Soares, D. C., Svergun, D. I., Herbert, A. P., Barlow, P. N. & Hannan, J. P. (2012). Structural Analysis of the C-Terminal Region (Modules 18-20) of Complement Regulator Factor H (FH). *PLoS ONE* **7,** e32187.
47. Fernando, A. N., Furtado, P. B., Clark, S. J., Gilbert, H. E., Day, A. J., Sim, R. B. & Perkins, S. J. (2007). Associative and structural properties of the region of complement factor H encompassing the Tyr402His disease-related polymorphism and its interactions with heparin. *J. Mol. Biol.* **368,** 564-581.
48. Okemefuna, A. I., Gilbert, H. E., Griggs, K. M., Ormsby, R. J., Gordon, D. L. & Perkins, S. J. (2008). The regulatory SCR-1/5 and cell-surface-binding SCR-16/20 fragments of factor H reveal partially folded-back solution structures and different self-associative properties. *J. Mol. Biol.* **375,** 80-101.
49. Aslam, M. & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **309,** 1117-1138.
50. DiScipio, R. G. (1992). Ultrastructures and interactions of complement factors H and I. *J. Immunol.* **149,** 2592-9.
51. Nan, R., Gor, J. & Perkins, S. J. (2008). Implications of the progressive self-association of wild-type human factor H for complement regulation and disease. *J. Mol. Biol.* **375,** 891-900.
52. Okemefuna, A. I., Nan, R., Gor, J. & Perkins, S. J. (2009). Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J. Mol. Biol.* **391,** 98-118.
53. Nan, R., Gor, J., Lengyel, I. & Perkins, S. J. (2008). Uncontrolled zinc- and copper- induced oligomerisation of the human complement regulator factor H and its possible implications for function and disease. *J. Mol. Biol.* **384,** 1341-1352.
54. Jouvin, M. H., Kazatchkine, D., Cahour, A. & Bernard, N. (1984). Lysine residues, but not carbohydrates, are required for the regulatory function of H on the amplification C3 convertase of complement. *J. Immunol.* **133,** 3250-3 254
55. Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riande, S., Siret, L. & Bihoreau, N. (2007). Site-specific N-glycan characterization of human complement factor H. *Glycobiology,* **17,** 932-944.
56. Roversi, P., Johnson, S., Caesar, J. J., McLean, F., Leath, K. J., Tsiftsoglou, S. A., Morgan, B. P., Harris, C. L., Sim, R. B. & Lea, S. M. (2011). Structural basis for complement factor I control and its disease-associated sequence polymorphisms. *Proc. Natl. Acad. Sci. USA,* **108,** 12839-44.

57. Kajander, T., Lehtinen, M. J., Hyvärinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E., Meri, S., Goldman, A. & Jokiranta, T. S. (2011). Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. *Proc. Natl. Acad. Sci. USA,* **108,** 2897-902.
58. Zipfel, P. F. & Skerka, C. (1994). Complement factor H and related proteins: an expanding family of complement-regulatory proteins? *Immunol. Today* **15,** 121-126.
59. Kuhn, S., Skerka, C. & Zipfel, P. F. (1995). Mapping of the complement regulatory domains in the human factor H-like protein 1 and in factor H1. *J. Immunol.* **155,** 5663-5670
60. Hellwage, J., Kuhn, S. & Zipfel, P. F. (1997). The human complement regulatory factor-H-like protein 1, which represents a truncated form of factor H, displays cell-attachment activity. *J. Biochem.* **326,** 321-327.
61. Zipfel, P. F., Skerka, C., Hellwage, J., Jokiranta, S. T., Meri, S., Brade, V., Kraiczy, P., Noris, M. & Remuzzi, G. (2002). Factor H family proteins: on complement, microbes and human diseases. *Biochem. Soc. T.* **30,** 971-978.
62. Jozsi, M. & Zipfel, P. F. (2008). Factor H family proteins and human diseases. *Trends Immunol.* **29,** 380-387.
63. Heinen, S., Hartmann, A., Lauer, N., Wiehl, U., Dahse, H. M., Schirmer, S., Gropp, K., Enghardt, T., Wallich, R., Halbich, S., Mihlan, M., Schlotzer-Schrehardt, U., Zipfel, P. F. & Skerka, C. (2009). Factor H-related protein 1 (CFHR-1). Inhibits complement C5 convertase activity and terminal complex formation. *Blood,* **114,** 2439-2447.
64. Mihlan, M., Hebecker, M., Dahse, H. M., Halbich, S., Huber-Lang, M., Dahse, R., Zipfel, P. F. & Jozsi, M. (2009). Human complement factor H-related protein 4 binds and recruits native pentameric C-reactive protein to necrotic cells. *Mol. Immunol.* **46,** 335-344.
65. de Cordoba, S. R. & de Jorge, E. G. (2008). Translational Mini-Review Series on Complement Factor H: Genetics and disease associations of human complement factor H. *Clin. Exp. Immunol.* **151,** 1-13.
66. Zipfel, P. F., Hallström, T., Hammerschmidt, S. & Skerka, C. (2008). The complement fitness Factor H: Role in human diseases and for immune escape of pathogens, like pneumococci. *Vaccine,* **26S,** I67-I74.
67. Rooijakkers, S. H. M. & van Strijp, J. A. G. (2007). Bacterial complement evasion. *Mol. Immunol.* **44,** 23-32.
68. Lambris, J. D., Ricklin, D. & Geisbrecht, B. V. (2008). Complement evasion by human pathogens. *Nat. Rev. Microbiol.* **6,** 132-142.
69. Blom, A. M., Hallström, T. & Riesbeck, K. (2009). Complement evasion strategies of pathogens-Acquisition of inhibitors and beyond. *Mol. Immunol.* **46,** 2808-2817.
70. Jarva, H., Jokiranta, T. S., Würzner, R. & Meri, S. (2003). Complement resistance mechanisms of streptococci. *Mol. Immunol.* **40**, 95-107.
71. Lim, W. S., Macfarlane, J. T., Boswell, T. C., Harrison, T. G., Rose, D., Leinonen, M. & Saikku, P. (2001). Study of community acquired pneumonia aetiology (SCAPA) in adults admitted to hospital: implications for management guidelines. *Thorax,* **56,** 296-301.

72. Musher, D. M. (2000). In *Principles and Practice of Infectious Diseases* (Mandell, G. L., Bennett, J. E. & Dolin, R. D., eds) pp. 2128-2147, Churchill Livingstone, New York.
73. Brown, J. S., Hussel, T., Gilliland, S. M., Holden, D.W., Paton, J. C., Ehrenstein, M. R., Walport, M. J. & Botto, M. (2002). The classical pathway is the dominant complement pathway required for innate immunity to *Streptococcus pneumoniae* infection in mice. *Proc. Natl. Acad. Sci. USA*, **99,** 16969-16974.
74. Tu, A. T., Fulgham, R. L., McCrory, M. A., Briles, D. E. & Szalai, A. J. (1999). Pneumococcal Surface Protein A Inhibits Complement Activation by Streptococcus pneumoniae. *Infect. Immun.* **67,** 4720-4724.
75. Gordon, S. B., Irving, G. R. B., Lawson, R. A., Lee, M. E. & Read, R. C. (2000). Intracellular Trafficking and Killing of *Streptococcus pneumoniae* by Human Alveolar Macrophages Are Influenced by Opsonins. *Infect. Immun.* **68,** 2286-2293.
76. Paton, J. C., Rowan-Kelly, B. & Ferrante A. (1984). Activation of Human Complement by the Pneumococcal Toxin Pneumolysin. *Infect. Immun.* **43,** 1085-1087.
77. Paton, J. C. (1996). The contribution of pneumolysin to the pathogenicity of *Streptococcus pneumoniae*. *Trends Microbiol.* **4,** 103-6.
78. Dave, S., Brooks-Walter, A., Pangburn, M. K. & McDaniel, L. S. (2001). PspC, a Pneumococcal Surface Protein, Binds Human Factor H. *Infect. Immun.* **69,** 3435-3437.
79. Janulczyk, R., Iannelli, F., Sjöholm, A. G., Pozzi, G. & Björck, L. (2000). Hic, a Novel Surface Protein of *Streptococcus pneumoniae* That Interferes with Complement Function. *J. Biol. Chem.* **275,** 37257-37263.
80. Neeleman, C., Geelen, S. P. M., Aerts, P. C., Daha, M. R., Mollnes, T. E., Roord, J. J., Posthuma, G., van Dijk, H. & Fleer, A. (1999). Resistance to both complement activation and phagocytosis in type 3 pneumococci is mediated by the binding of complement regulatory protein factor H. *Infect. Immun.* **67,** 4517-4524.
81. Hammerschmidt, S., Talay, S. R., Brandtzaeg, P. & Chhatwal, G. S. (1997). SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol. Microbiol.* **25,** 1113-24.
82. Cheng, Q., Finkel, D. & Hostetter, M. K. (2000). Novel purification scheme and functions for a C3-binding protein from *Streptococcus pneumoniae*. *Biochemistry* **39,** 5450-5457.
83. Jarva, H., Janulczyk, R., Hellwage, J., Zipfel, P. F., Björck, L. & Meri, S. (2002). *Streptococcus pneumoniae* Evades Complement Attack and Opsonophagocytosis by Expressing the *pspC* Locus-Encoded Hic Protein That Binds to Short Consensus Repeats 8-11 of Factor H. *J. Immunol.* **168,** 1886-1894.
84. AlonsoDeVelasco, E., Verheul, A. F. M., Verhoef, J & Snippe, H. (1995). *Streptococcus pneumoniae*: Virulence Factors, Pathogenesis, and Vaccines. *Microbiol. Rev.* **59,** 591-603.

85. Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* **6,** 288-301.
86. Iannelli, F., Oggioni, M. R. & Pozzi, G. (2002). Allelic variation in the highly polymorphic locus pspC of *Streptococcus pneumoniae*. *Gene,* **284,** 63-71.
87. Hammerschmidt, S., Agarwal, V., Kunert, A., Haelbich, S., Skerka, C. & Zipfel, P. F. (2007). The Host Immune Regulator Factor H Interacts via Two Contact Sites with the PspC Protein of *Streptococcus pneumoniae* and Mediates Adhesion to Host Epithelial Cell. *J. Immunol.* **178,** 5848-5858.
88. Lu, L., Ma, Y. & Zhang, J. (2006). *Streptococcus pneumoniae* Recruits Complement Factor H through the Amino Terminus of CbpA. *J. Biol. Chem.* **281,** 15464-15474.
89. Luo, R., Mann, B., Lewis, W. S., Rowe, A., Heath, R., Stewart, M. L. Hamburger, A. E., Sivakolundu, S., Lacy, E. R., Bjorkman, P. J., Tuomanen, E. & Kriwacki, R. W. (2005). Solution structure of choline binding protein A, the major adhesin of *Streptococcus pneumoniae*. *EMBO J.* **24,** 34-43.
90. Dave, S., Pangburn, M. K., Pruitt, C. & McDaniel, L. S. (2004). Interaction of human factor H with PspC of *Streptococcus pneumoniae*. *Indian J. Med. Res.* **119,** 66-73.
91. Duthy, T. G., Ormsby, R. J., Giannakis, E., Ogunniyi, A. D., Stroeher, U. H., Paton, J. C. & Gordon, D. L. (2002). The Human Complement Regulator Factor H Binds Pneumococcal Surface Protein PspC via Short Consensus Repeats 13 to 15. *Infect. Immun.* **70,** 5604-5611.
92. Agarwal, V., Asmat, T. M. Luo, S., Jensch, I., Zipfel, P. F. & Hammerschmidt, S. (2010). Complement Regulator Factor H Mediates a Two-step Uptake of *Streptococcus pneumoniae* by Human Cells. *J. Biol. Chem.* **285,** 23486-23495.
93. Schneider, M. C., Prosser, B. E., Ceasar, J. J. E., Kugelberg, E., Li, S., Zhang, Q., Quoraishi, S., Lovett, J. E., Derane, J. E., Sim, R. B., Roversi, P., Johnson, S., Tang, C. M. & Lea, S. M. (2009). *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature,* **458,** 890-895.
94. Sharma, A. K. & Pangburn, M. K. (1997). Localization by Site-Directed Mutagenesis of the Site in Human Complement Factor H That Binds to *Streptococcus pyogenes* M Protein. *Infect. Immun.* **65,** 484-487.
95. Hallström, T., Zipfel, P. F., Blom, A. M., Lauer, N., Forsgren, A. & Riesbeck, K. (2008). *Haemophilus influenzae* Interacts with the Human Complement Inhibitor Factor H. *J. Immunol.* **181,** 537-545.
96. Alitalo, A., Meri, T., Chen, T. Lankinen, H., Cheng, Z., Jokiranta, T. S., Seppälä, I. J. T., Lahdenne, P., Hefty, P. S., Akins, D. R. & Meri, S. (2004). Lysine-Dependent Multipoint Binding of the *Borrelia burgdorferi* Virulence Factor Outer Surface Protein E to the C Terminus of Factor H. *J. Immunol.* **172,** 6195-6201.
97. Ricci, S., Janulczyk, R., Gerlini, A., Braione, V., Colomba, L., Iannelli, F., Chiavolini, D., Oggioni, M. R., Björck, L. & Pozzi, G. (2011). The factor H-binding fragment of PspC as a vaccine antigen for the induction of protective humoral immunity against experimental pneumococcal sepsis. *Vaccine,* **29,** 8241-8249.

98. Cregg, J.M., Cereghino, J.L., Shi, J. & Higgins, D. R. (2000). Recombinant Protein expression in *Pichia pastoris*. *Mol. Biotechnol.* **16,** 23-52.
99. Cereghino, J.L. & Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* **24,** 45-66.
100. Gellissen, G. (2000). Heterologous protein production in methylotrophic yeasts. *Appl. Microbiol. Biot.* **54,** 741-750.
101. Cereghino, G. P. L., Cereghino, J. L., Ilgen, C. & Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast Pichia pastoris. *Curr. Opin. Biotech.* **13,** 329-332.
102. Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenbach, G. T., Urdea, M. S., Valenzuela, P. & Barr, P. J. (1984). α-Factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proc. Nat. Acad. Sci. USA,* **81,** 4642-4646.
103. Julius D., Blair L., Brake A., Sprague G & Thorner J. (1983). Yeast α Factor Is Processed from a Larger Precursor Polypeptide: The Essential Role of a Membrane-Bound Dipeptidyl Aminopeptidase. *Cell,* **32,** 839-852.
104. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol.* **185,** 60-89.
105. Novy, R. & Morris, B. (2001). *Innovations*, **13,** 8-10.
106. Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189,** 113-130.
107. Hannig, G. & Makrides, S. C. (1998). Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends Biotechnol.* **16,** 54-60.
108. Marblestone, J. G., Edavettal, S. C., Lim, Y., Zuo, X. & Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **15,** 182-189.
109. Müller, S., Hoege, C., Pyrowolakis, G. & Jentsch, S. (2001). SUMO, Ubiquitin's Mysterious Cousin. *Nat. Rev. Mol. Cell Biol.* **2,** 202-210.
110. Mossessova, E. & Lima, C .D. (2000). Ulp1-SUMO Crystal Structure and Genetic Analysis Reveal Conserved Interactions and a Regulatory Essential for Cell Growth in Yeast. *Mol. Cell,* **20,** 2367-2377.
111. Li, S. J. & Hochstrasser, M. (1999). A new Protease required for cell-cycle progression in yeast. *Nature,* **398,** 246-251.
112. Vranken, W.F., Boucher, Steven, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. & Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins,* **59,** 687-696.
113. Bodenhausen, G. & Ruben, D.J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69,** 185.
114. Vuister, G. W. & Bax, A. (1992). Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling. *J. Magn. Reson.* **98,** 428-435.
115. Yamazaki, T., Forman-Kay, J.D. & Kay, L.E. (1993). Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *J. Am. Chem. Soc.* **115,** 11054-11055.

116. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J. Biomol. NMR,* **3,** 185-204.
117. Grzesiek, S. & Bax, A. (1992). An efficient experiment for sequential backbone assignmnet of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99,** 201.
118. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and sidechain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114,** 6291.
119. Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96,** 432.
120. Clubb, R.T., Thanabal, V. & Wagner, G. (1992). A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C chemical shifts in 15N-13C-labeled proteins. *J. Magn. Reson.* **97,** 213.
121. Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R. & Forman-Kay, J.D. (1993). A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in H2O Samples of Proteins. *J. Magn. Reson. Ser. B,* **101,** 333.
122. Montelione, G.T., Lyons, B.A., Emerson, S.D. & Tashiro, M. (1992). An efficient triple resonance experiment using carbon-13C isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.* **114,** 10974-10975.
123. Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D. & Kay, L.E. (1994). Simultaneous Acquisition of 15N- and 13C- Edited NOE Spectra of Proteins Dissolved in H2O. *J. Magn. Reson. Ser. B,* **103,** 197.
124. Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. (1993). Gradient-Tailored Water Supression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson. Ser. A,* **102,** 241.
125. Schubert, M., Labudde, D., Oschkinat H. & Schmieder, P. (2002). A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shifts stastistics. *J. Biomol. NMR,* **24,** 149-154.
126. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278,** 353-378.
127. Guntert, P., Mumenthaler, C. & Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273,** 283-298.
128. Herrmann, T., Guntert, P. & Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319,** 209-227.
129. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. & Nilges, M. (2003). Refinement of protein structures in explicit solvent. *Proteins,* **50,** 496-506.
130. Brunger, A. T., Adams, P D., Clore, G. M., DeLano, W. L., Gross, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998).

Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **54,** 905-921.
131. Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229,** 317-324.
132. Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE cross-peaks and disulfide connectivities. *J. Mol. Biol.* **245,** 645-660.
133. Kay, L. E., Nicholson, K. K., Delaglio, F., Bax, A, & Torchia, D.A. (1992). Pulse sequences for removal of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T1 and T2 values in proteins. *J. Magn. Reson.* **97,** 359-375.
134. Grzesiek, S. B. & Bax A. (1993). The importance of not saturating H2O in protein NMR-application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115,** 12593-12594.
135. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **28,** 8972-8979.
136. Koradi, R., Billeter, M. & Wuthrich, K. (1996). *J. Mol. Graph.* **14,** 51-55, 29-32.
137. DeLano, W.L.T.P.M.G.S. (2002). The PyMOL Molecular Graphics System. In., Palo Alto, CA.
138. Heinig, M., & Frishman, D. (2004). STRIDE: a Web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32,** W500-502.
139. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins,* 11, 281-296.
140. Heiden, W., Moeckel, G. & Brickmann, J. (1993). A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J. Comput. Aided Mol. Des.* **7,** 503-514.
141. Fraczkiewicz, R. and Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19,** 319-333.
142. Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D. & Wishart, D.S. (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **31,** 3316-3319.
143. Shindyalov, I. N., Bourne, P. E. (1998). Protein structural alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11,** 739-747.
144. Soares, D. C., Gerloff, D. L., Syme, N. R., Coulson, A. F., Parkinson, J. & Barlow, P. N. (2005). Large-scale modelling as a route to multiple surface comparisons of the CCP module family. *Protein Eng. Des. Sel.* **18,** 379-388.
145. Soares, D. C. & Barlow, P. N. (2005). Complement control protein modules in the regulators of complement activation. In *Structural Biology of the Complement System* (Morikis, D. and Lambris, J. D., eds), pp. 19-62, CRC Press, Taylor and Francis Group, Boca Raton, FL.

146. Laskowski. R. A., Macarthur, M. W., Moss, D.S. & Thornton, J. M. (1993). PROCHECK-a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283-291.
147. Denton, H., Smith, M., Husi, H., Uhrin, D. & Barlow, P. N. (1998). Isotopically Labeled Bovine β-Lactoglobulin for NMR Studies Expressed in *Pichia pastoris*. *Protein Expr. Purif.* **14,** 97-103.
148. Faull, P. A., Florance, H. V., Schmidt, C. Q., Tomczyk, N., Barlow, P. N., Hupp, T. R., Nikolova, P. V. & Barran, P. E. (2010). Utilising ion mobility-mass spectroscopy to interrogate macromolecules: Factor H complement control protein modules 10-15 and 19-20 and the DNA-binding core domain of tumor suppressor p53. *Int. J. Mass. Spectrom.* **298,** 99-110.
149. Makou, E., Mertens, H. D. T., Maciejewski, M., Soares, D. C., Matis, I., Schmidt, C. Q., Herbert, A. P., Svergun, D. I. & Barlow, P. N. (2012). Solution Structure of CCP Modules 10-12 Illuminates Functional Architecture of the Complement Regulator, Factor H. *J. Mol. Biol.* **424,** 295-312.
150. Henderson, C. E., Bromek, K., Mullin, N. P., Smith, B. O., Uhrín, D. & Barlow, P. N. (2001). Solution Structure and Dynamics of the Central CCP Module Pair of a Poxvirus Complement Control Protein. *J. Mol. Biol.* **307,** 232-339.
151. Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M. & Chen, X. S. (2001). Structure of Complement Receptor 2 in Complex with its C3d ligand. *Science,* **292,** 1725-1728.
152. Midler, F. J., Gomes, L., Schouten, A., Janssen, B. J. C., Huizinga, E. G., Romijn, R. A., Hemrika, W., Roos, A., Daha, M. R. & Gros, P. (2007). Factor B structure provides insights into activation of the central protease of the complement system. *Nat. Struct. Mol. Biol.* **14,** 224-228.
153. Krishnan, V., Xu, Y., Macon, K., Volanakis, J. E. & Narayana, S. V. L. (2009). The structure of C2b, a fragment of complement component C2 produced during C3 convertase formation. *Acta Crystallogr. D. Biol. Crystallogr.* **65,** 226-274.
154. Merterns, H. D. T. & Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172,** 128-141.
155. Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28,** 768-773.
156. Franke, D. & Svergun, D. I. (2009). *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Cryst.* **42,** 342-346.
157. Kozin, M. B. & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* **34**, 33-41.
158. Uhrínová, S., Smith, M. H., Jameson, G. B., Uhrín, D., Sawyer, L. & Barlow, P. N. (2000). Structural Changes Accompanying pH-Induced Dissociation of the β-Lactoglobulin Dimer. *Biochemistry* **39,** 3565-3574.
159. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992). Backbone Dynamics of Calmodulin Studied by ^{15}N Relaxation Using Inverse Detected Two-Dimensional NMR Spectroscopy: The Central Helix is Flexible. *Biochemistry* **31,** 5269-5278.

160. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *J. Am. Chem. Soc.* **129,** 5656-5664.
161. Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Cryst.* **39,** 277-286.
162. Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89,** 1237-1250.
163. Nieba, L., Nieba-Axmann, S., E., Persson, A., Hämäläinen, M., Edebratt, F., Hansson, A., Lidholm, J., Magnusson, K., Karlsson, A., F. & Plückthun, A. (1997). BIACORE Analysis of Histidine-Tagged Proteins Using a Chelating NTA Sensor Chip. *Anal. Biochem.* **252,** 217-228.
164. Perkins, S. J., Nan, R., Li, K., Khan, S. & Miller, A. (2012). Complement Factor H-ligand interactions:Self-association, multivalency and dissociation constants. *Immunobiology* **217,** 281-297.
165. Mold, C., Gewurz, H. & Du Clos, T. W. (1999). Regulation of complement activation by C-reactive protein. *Immunopharmacology* **42,** 23-30.
166. Jarva, H., Jokiranta, T. S., Hellwage, J., Zipfel, P. F. & Meri, S. (1999). Regulation of complement activation by C-reactive protein: targeting the complement inhibitory activity of factor H by an interaction with short consensus repeat domains 7 and 8-11. *J. Immunol.* **163,** 3957-62.
167. Blein, S., Ginham, R., Uhrin, D., Smith, B. O., Soares, D. C., Veltel, S., McIlhinney, R. A., White, J. H. & Barlow, P. N. (2004). Structural analysis of the complement control protein (CCP) modules of GABA(B) receptor 1a: only one of the two CCP modules is compactly folded. *J. Biol. Chem.* **279,** 48292-306.
168. Rickert, M., Wang, X., Boulanger, M. J., Goriatcheva, N. & Garcia, K. C. (2005). The Structure of Interleukin-2 Complexed with Its Alpha Receptor. *Science,* **308,** 1477-1480.
169. Andersen, C. B., Torvund-Jensen, M., Nielsen, M. J., de Oliveira, C. L., Hersleth, H. P., Andersen, N. H., Pedersen, J. S., Andersen, G. R. & Moestrup, S. K. (2012). Structure of haptoglobin-haemoglobin complex. *Nature,* **489,** 456-459.

APPENDICES

APPENDIX A

Media and buffers

MEDIA	COMPOSITION (w/v unless stated otherwise)
BMG (Buffered minimal glycerol)	100 mM potassium phosphate pH 6 1.34 % YNB $4x10^{-5}$ % biotin 1 % glycerol
BMM (Buffered minimal methanol)	100 mM potassium phosphate pH 6 1.34 % YNB $4x10^{-5}$ % biotin 0.5 % methanol
Low-salt LB (Luria-Bertani) (low salt is needed for use with zeocin antibiotic)	0.5 % yeast extract 1 % tryptone 0.5 % sodium chloride +/- 1.5 % agar
LB (Luria-Bertani)	0.5 % yeast extract 1 % tryptone 1 % sodium chloride +/- 1.5 % agar
SOC (Super Optimal broth with Catabolite repression)	0.5 % yeast extract 2 % tryptone 10 mM sodium chloride 2.5 mM potassium chloride 10 mM magnesium chloride 10 mM magnesium sulfate 20 mM D-glucose
YNB (Yeast Nitrogen Base)	10x YNB stock (Sigma) with ammunium sulfate without amino acid
YPD (Yeast extract, Peptone, Dextrose)	1 % yeast extract 2 % peptone 2 % dextrose (D-glucose) +/- 1.5 % agar
EDTA stock solution (ethylenediaminetetraacetic	0.5 M stock adjusted to pH with sodium

acid)	hydroxide
SDS-PAGE sample loading buffer	50 mM Tris-HCl 100 mM %-mercaptoethanol 2 % sodium-dodecyl-sulfate 0.1 % bromophenol blue 10 % glycerol
TCA (trichloroacetic acid)	30 % stock (Sigma) made up in H_2O
TAE buffer 50x in distilled H_2O	Tris-acetate (2 M) EDTA (100 mM)
TAE Agarose	Agarose 1 % in TAE buffer

APPENDIX B

Vector maps

1) pCR®-4Blunt-TOPO®

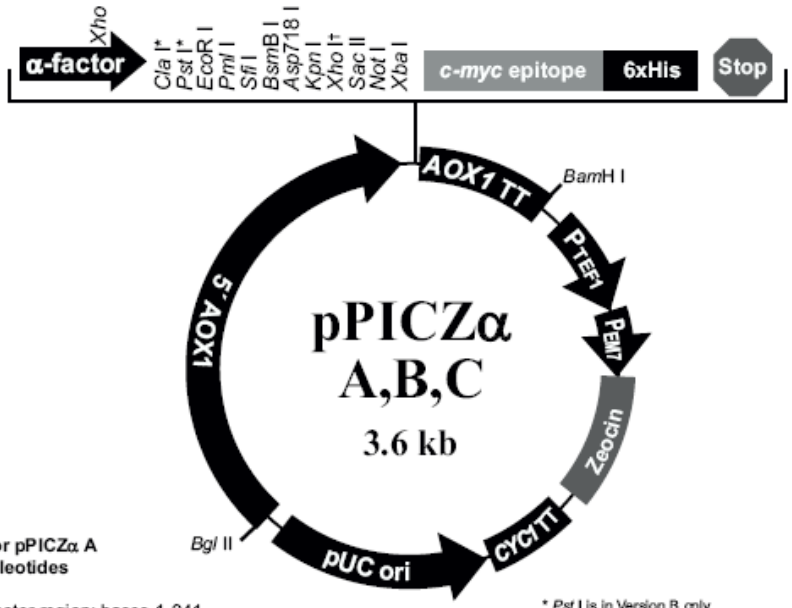
pCR®4Blunt-TOPO® Map

The map below shows the features of pCR®4Blunt-TOPO® and the sequence surrounding the TOPO® Cloning site. Restriction sites are labeled to indicate the actual cleavage site. **The complete sequence of pCR®4Blunt-TOPO® is available for downloading from our Web site (www.invitrogen.com) or by contacting Technical Service (page 27).**

Comments for pCR®4Blunt-TOPO®
3956 nucleotides

lac promoter region: bases 2-216
  CAP binding site: bases 95-132
  RNA polymerase binding site: bases 133-178
  Lac repressor binding site: bases 179-199
Start of transcription: base 179
M13 Reverse priming site: bases 205-221
LacZα-*ccdB* gene fusion: bases 217-810
  LacZα portion of fusion: bases 217-497
  *ccd*B portion of fusion: bases 508-810
T3 priming site: bases 243-262
TOPO® Cloning site: bases 294-295
T7 priming site: bases 328-347
M13 Forward (-20) priming site: bases 355-370
Kanamycin promoter: bases 1021-1070
Kanamycin resistance gene: bases 1159-1953
Ampicillin (*bla*) resistance gene: bases 2203-3063 (c)
Ampicillin (*bla*) promoter: bases 3064-3160 (c)
pUC origin: bases 3161-3834
(c) = complementary strand

2) pPICZαB



Comments for pPICZα A
3593 nucleotides

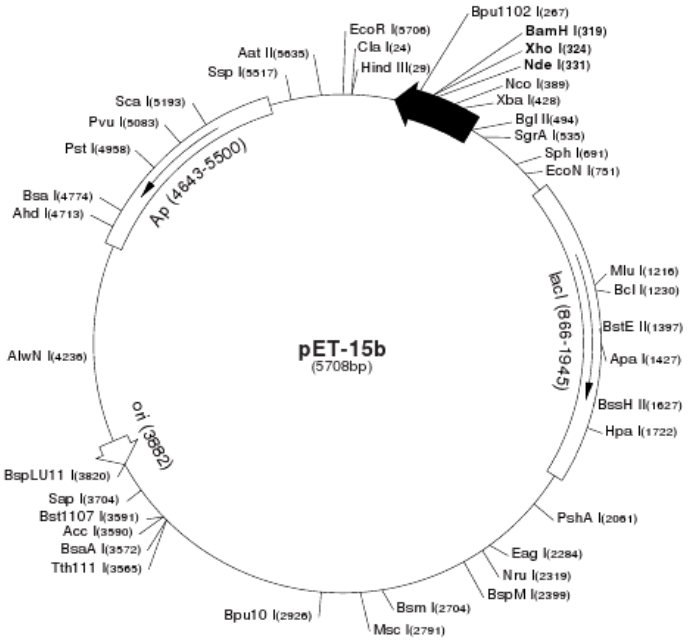
5′ *AOX1* promoter region: bases 1-941
5′ *AOX1* priming site: bases 855-875
α-factor signal sequence: bases 941-1207
α-factor priming site: bases 1144-1164
Multiple cloning site: bases 1208-1276
c-myc epitope: bases 1275-1304
Polyhistidine (6xHis) tag: bases 1320-1337
3′ *AOX1* priming site: bases 1423-1443
AOX1 transcription termination region: bases 1341-1682
TEF1 promoter: bases 1683-2093
EM7 promoter: bases 2095-2162
Sh ble ORF: bases 2163-2537
CYC1 transcription termination region: bases 2538-2855
pUC origin: bases 2866-3539 (complementary strand)

* *Pst* I is in Version B only
Cla I is in Version C only

†The two *Xho* I sites in the vector allow the user to clone their gene in frame with the Kex2 cleavage site, resulting in expression of their native gene without additional amino acids at the N-terminus.

3) pET-15b

pET-15b sequence landmarks	
T7 promoter	453-469
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde* I - *Bam* H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500

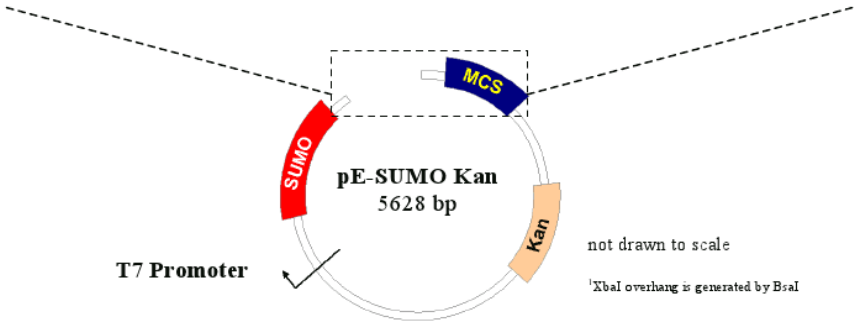


4) pE-SUMOPro Kan

Multiple Cloning Site (MCS)

pE-SUMO vector has to be linearized with BsaI or Eco31I restriction enzyme

Gly	Gly	XbaI[1]	BamHI		SacI	SalI	HindIII	NotI EagI	XhoI
GG		CTAGA	GGATCC	GAATTC	GAGCTCC	GTCGAC	AAGCTT	GCGGCCGC	ACTCGAG CACCA
CCTCCA		T	CCTAGG	CTTAAG	CTCGAGG	CAGCTG	TTCGAA	CGCCGGCG	TGAGCTC GTGGT



APPENDIX C

1) CYANA quality control criteria for the NMR structures of FH-10-11 and FH-11-12

Quality control criteria based on the output form the CYANA calculations
1. The average target function value of cycle should be below 250 $Å^2$
2. The average target function value of the final cycle should be below 10 $Å^2$
3. There should be less than 20 % unassigned NOEs
4. There should be less than 20 % discarded long-range NOEs
5. The RMSD value in cycle 1 should be below 3 Å
6. The RMSD between the mean structures of the first and the last cycle should be below 3 Å.

Table C1 Summary of the CYANA quality control criteria.

Cycle	1	2	3	4	5	6	7	Final
Peaks								
selected	5045	5045	5045	5045	5045	5045	5045	
assigned	4653	4681	4582	4593	4536	4479	4455	
unassigned	392	364	463	452	509	566	590	
with diagonal assignment	0	0	0	0	0	0	0	
Cross peaks								
with off-diagonal assignment	4653	4681	4582	4593	4536	4479	4455	
with unique assignment	1019	2793	3258	3340	3541	3701	3678	
with short-range assignment |i-j|<=1	2868	2835	2747	2724	2672	2628	2638	
with medium-range assignment 1<|i-j|<5	635	502	457	466	455	436	432	
with long-range assignment |i-j|>=5	1150	1344	1378	1403	1409	1415	1385	
Upper distance limits								
total	3633	3424	3218	3179	3071	2964	3099	3207
short-range, |i-j|<=1	1853	1727	1597	1543	1470	1406	1369	1405
medium-range, 1<|i-j|<5	1036	836	423	430	408	386	399	411
long-range, |i-j|>=5	744	861	1198	1206	1193	1172	1331	1391
Assignment/constraint	5.77	2.35	1.4	1.38	1.29	1.22	1	1
Average target function value	162.46	85.7	160.81	30.73	15.41	9.46	12.32	4.38
RMSD (residues 566..687)								
Average backbone RMSD to mean	6.18	334	1.92	1.15	0.99	1.05	1.13	0.87
Average heavy atom RMSD to mean	6.8	3.7	2.22	1.51	1.37	1.44	1.47	1.26

Table C2 Report from all cycles of the CYANA-based structure calculations for FH-10-11.

Cycle	1	2	3	4	5	6	7	Final
Peaks								
selected	5948	5948	5948	5948	5948	5948	5948	
assigned	5319	5246	5072	5055	4983	4905	4873	
unassigned	629	702	876	893	965	1043	1075	
with diagonal assignment	0	0	0	0	0	0	0	
Cross peaks								
with off-diagonal assignment	5319	5246	5072	5055	4983	4905	4873	
with unique assignment	1236	3481	3866	3902	4147	4235	4230	
with short-range assignment |i-j|<=1	3395	3324	3240	3189	3114	3071	3069	
with medium-range assignment 1<|i-j|<5	589	424	361	368	358	342	337	
with long-range assignment |i-j|>=5	1335	1498	1471	1498	1511	1492	1467	
Upper distance limits								
total	4183	3668	3365	3300	3139	3021	3030	3105
short-range, |i-j|<=1	2280	1930	1763	1681	1564	1499	1438	1482
medium-range, 1<|i-j|<5	987	697	325	329	317	298	300	301
long-range, |i-j|>=5	916	1041	1277	1290	1258	1224	1292	1322
Assignments/constraint	5.17	2.18	1.33	1.32	1.22	1.18	1	1
Average target function value	385.54	187.1	288.64	41.9	19.19	10.6	6.88	3.09
RMSD (residues 630..744)								
Average backbone RMSD to mean	5.67	3.22	1.23	0.64	0.54	0.56	0.65	0.46
Average heavy atom RMSD to mean	6.11	3.5	1.45	0.95	0.89	0.9	0.96	0.77

Table C3 Report from all cycles of the CYANA structure calculations for FH-11-12.

2) Intermodular angles for FH 10-11 and FH 11-12

FH-10-11	Tilt (°)	Twist (°)	Skew (°)
Model 1	128.08	40.08	80.11
Model 2	119.35	41.60	85.69
Model 3	114.93	62.94	85.80
Model 4	125.51	59.97	96.55
Model 5	122.27	74.60	88.89
Model 6	118.36	57.97	80.23
Model 7	114.64	46.67	74.89
Model 8	123.29	54.55	93.91
Model 9	119.58	53.07	86.61
Model 10	110.76	77.77	97.55
Model 11	**118.15**	**64.45**	**96.60**
Model 12	110.65	57.72	78.09
Model 13	119.51	69.68	106.94
Model 14	120.02	47.82	90.15
Model 15	113.00	54.19	87.52
Model 16	123.42	54.46	84.76
Model 17	118.46	61.30	89.10
Model 18	117.58	64.69	90.16
Model 19	115.45	47.00	86.27
Model 20	115.91	59.45	92.44
Minimum	110.65	40.08	74.89
Maximum	128.08	77.77	106.94
Mean	118.45	57.25	88.61
Standard deviation	4.61	10.14	7.52

Table C4 Tilt, twist and skew angles for FH-10-11.

FH-11-12	Tilt (°)	Twist (°)	Skew (°)
Model 1	90.73	87.00	169.96
Model 2	92.46	66.66	156.13
Model 3	84.79	72.35	157.81
Model 4	82.23	68.74	154.61
Model 5	85.77	72.08	158.32
Model 6	98.70	71.43	158.51
Model 7	87.74	70.57	163.51
Model 8	91.47	71.11	157.02
Model 9	95.56	63.06	152.28
Model 10	82.26	71.81	154.79
Model 11	96.94	74.93	153.24
Model 12	**88.47**	**79.37**	**158.98**
Model 13	84.21	74.76	157.04
Model 14	85.27	78.24	160.70
Model 15	90.12	66.66	154.48
Model 16	83.95	83.15	159.62
Model 17	88.63	67.70	159.23
Model 18	83.71	77.06	158.30
Model 19	75.94	63.50	157.65
Model 20	95.19	58.70	148.82
Minimum	75.94	58.70	148.82
Maximum	98.70	87.00	169.96
Mean	88.26	71.97	157.55
Standard deviation	5.83	6.85	4.36

Table C5 Tilt, twist and skew angles for FH-11-12.

Tables C4 and C5: Inter-modular angles (°) for FH-10-11 and FH-11-12. The values for the closest-to-mean model for the two CCP pairs is shown in bold.

3) Overlay of FH-10-11 and FH-11-12 with known CCP structures

Protein; module; PDB ID Reference	RMSD in Å (alignment length, gaps included)			
	FH-10	FH-11[a]	FH-11[b]	FH-12
C1r; CCP 1; (1GPZ) (*1*)	1.89 (59)	2.16 (58)	2.18 (58)	2.28 (58)
C1r; CCP 2; (1GPZ) (*1*)	2.52 (34)	2.68 (34)	2.41 (35)	1.98 (38)
C1s; CCP 2; (1ELV) (*2*)	2.20 (58)	2.31 (50)	2.45 (50)	2.05 (50)
C2; CCP1; (3ERB) (*3*)	3.17 (51)	3.07 (50)	3.25 (50)	3.01 (50)
C2; CCP2; (3ERB) (*3*)	2.13 (59)	2.45 (58)	2.68 (58)	2.88 (58)
C2; CCP3; (3ERB) (*3*)	1.94 (57)	2.13 (56)	2.20 (56)	2.17 (56)
C4BPα; CCP 1; (2A55) (*4*)	3.51 (59)	2.27 (58)	2.41 (58)	2.47 (58)
C4BPα; CCP 2; (2A55) (*4*)	2.05 (58)	2.07 (52)	1.99 (52)	2.31 (57)
C7; CCP 1; (Clark *et al*., unpublished)	1.95 (58)	1.97 (58)	1.94 (57)	2.10 (58)
C7; CCP2; (Clark *et al*., unpublished)	1.77 (59)	2.47 (57)	2.44 (57)	2.86 (57)
CR1; CCP 15; (1GKN) (*5*)	2.81 (57)	2.26 (58)	2.24 (57)	2.37 (56)
CR1; CCP 16; (1GKN) (*5*)	2.06 (58)	2.03 (51)	2.06 (51)	2.31 (57)
CR1; CCP 17; (1GKG) (*5*)	2.80 (61)	2.14 (58)	1.99 (57)	2.09 (57)
CR2; CCP 1; (1LY2) (*6*)	2.05 (59)	1.71 (58)	1.78 (58)	2.10 (58)
CR2; CCP 2; (1LY2) (*6*)	1.94 (58)	1.70 (57)	1.82 (57)	1.41 (54)*
CRRY; CCP 1; (2XRB) (*7*)	3.36 (58)	2.42 (57)	2.48 (58)	2.47 (56)
CRRY; CCP 2; (2XRB) (*7*)	2.28 (58)	2.35 (53)	2.08 (53)	2.28 (57)
CRRY; CCP 3; (2XRB) (*7*)	1.71 (59)	1.72 (59)	1.74 (59)	2.03 (58)
CRRY; CCP 4; (2XRB) (*7*)	1.95 (57)	1.53 (54)	1.59 (54)	1.96 (56)
DAF; CCP 1; (1OK3) (*8*)	2.39 (59)	2.20 (59)	1.89 (59)	2.25 (58)
DAF; CCP 2 (1OK3) (*8*)	2.58 (56)	2.24 (57)	1.94 (56)	1.94 (58)
DAF; CCP 3; (1H03) (*9*)	2.28 (58)	2.38 (57)	2.36 (57)	2.31 (57)
DAF; CCP4; (1H03) (*9*)	1.75 (58)	1.59 (57)	1.62 (57)	1.79 (57)
FB; CCP 1; (2OK5) (*10*)	2.94 (62)	2.61 (46)	3.86 (65)	3.10 (63)
FB; CCP 2; (2OK5) (*10*)	2.01 (59)	2.21 (58)	2.43 (58)	2.72 (58)
FB; CCP 3; (2OK5) (*10*)	2.04 (57)	2.16 (56)	2.24 (56)	2.47 (56)
FH; CCP 1; (2RLP) (*11*)	2.73 (51)	2.04 (57)	2.09 (57)	2.39 (56)
FH; CCP 2; (2RLQ) (*11*)	2.41 (57)	2.29 (54)	2.34 (55)	2.38 (56)
FH; CCP 3; (2RLQ) (*11*)	1.97 (57)	2.08 (57)	2.14 (57)	2.47 (56)
FH; CCP 4; (2WII) (*12*)	2.16 (57)	1.72 (57)	1.82 (57)	2.14 (55)
FH; CCP 5; (not deposited) (*13*)	2.24 (56)	2.42 (53)	2.45 (53)	2.37 (52)
FH; CCP 6; (2UWN) (*14*)	3.16 (59)	2.51 (57)	2.19 (56)	2.53 (56)
FH; CCP 7; (2UWN) (*14*)	2.89 (56)	3.05 (57)	2.85 (55)	3.56 (57)
FH; CCP 8; (2UWN) (*14*)	2.00 (54)	1.80 (51)	1.95 (51)	2.25 (56)
FH; CCP 10; (this study)	-	2.04 (57)	2.15 (58)	2.37 (58)
FH; CCP 11[a]; (this study)	2.04 (57)	-	0.87 (59)	1.61 (58)
FH; CCP 11[b]; (this study)	2.15 (58)	0.87 (59)	-	1.57 (58)
FH; CCP 12; (this study)	2.37 (58)	1.61 (58)	1.57 (58)	-

FH; CCP 12; (2KMS) (*15*)	2.19 (57)	1.21 (58)	1.26 (58)	1.25 (58)
FH; CCP 13; (2KMS) (*15*)	2.83 (48)	3.28 (56)	3.61 (43)	3.40 (50)
FH; CCP 15; (1HFH) (*16*)	2.53 (58)	1.90 (59)	1.89 (59)	1.84 (57)
FH; CCP 16; (1HFH) (*16*)	2.89 (51)	1.79 (54)	1.89 (56)	1.81 (56)
FH; CCP 18; (3SW0) (*17*)	1.96 (58)	1.25 (59)	1.38 (59)	1.80 (58)
FH; CCP 19; (2G7I) (*18*)	1.93 (57)	1.18 (59)*	1.13 (58)*	1.70 (57)
FH; CCP 20; (2G7I) (*18*)	2.08 (56)	2.10 (56)	2.49 (57)	1.69 (48)
MASP1; CCP 1; (3GOV) (*19*)	1.92 (59)	2.11 (58)	2.17 (58)	2.42 (58)
MASP1; CCP 2; (3GOV) (*19*)	2.43 (59)	2.50 (59)	2.42 (59)	1.97 (58)
MASP2; CCP 1; (1ZJK) (*20*)	2.02 (59)	2.09 (58)	2.14 (58)	2.36 (58)
MASP2; CCP 2; (1ZJK) (*20*)	2.25 (59)	2.24 (59)	2.26 (59)	1.98 (58)
MCP; CCP 1; (1CKL) (*21*)	2.40 (55)	2.09 (57)	1.98 (57)	2.14 (57)
MCP; CCP 2; (1CKL) (*21*)	2.60 (59)	2.88 (58)	3.00 (58)	2.79 (58)
MCP; CCP 3; (3O8E) (*22*)	1.69 (59)*	1.77 (59)	1.77 (59)	2.27 (58)
MCP; CCP 4; (3O8E) (*22*)	2.39 (59)	1.89 (55)	1.92 (55)	1.99 (54)
VCP; CCP 1; (1G40) (*23*)	2.52 (58)	2.70 (59)	2.60 (59)	2.87 (58)
VCP; CCP 2; (1G40) (*23*)	2.34 (57)	2.23 (51)	2.34 (51)	2.27 (50)
VCP; CCP 3; (1G40) (*23*)	2.47 (58)	1.89 (58)	2.04 (58)	1.99 (57)
VCP; CCP 4; (1G40) (*23*)	1.89 (56)	2.00 (56)	2.04 (56)	2.21 (55)

Table C6 Pair-wise CCP module structural comparisons of FH 10, FH 11[a] (where [a] = from FH 10-11 bi-module pair), FH 11[b] (where [b] = from FH 11-12 bi-module pair) and FH 12. Colour key used in table: Blue: 0 - 1.99 Å; Green: 2.00 – 2.99 Å; Red: 3.00 – 3.99 Å; Brown: Alignment lengths < 40 amino acids. Abbreviations used in Table: C4BPα = C4b-binding protein α-chain; CR = complement receptor; CRRY = rat CR1-related protein Y; DAF = decay-accelerating factor; FB = factor B; FH = factor H; MASP1 / 2 = mannan-binding lectin-associated serine proteases 1 / 2; MCP = membrane cofactor protein; VCP = Vaccinia virus complement control protein. Some residues were not present (solved) in the electron density map for the C1r CCP 2 module crystal structure, and this explains the short structural alignment length (shown in brown). The best overlaying structures in each case are indicated by '*' (149).