Structure and function of the central part of complement factor H

Christoph Schmidt

PhD

The University of Edinburgh

2008

Abstract

Factor H (FH) is a key regulator of the major humoral component of the innate immune system, the complement system. Activation of complement leads to opsonisation and immune clearance, cytolysis and release of mediators of inflammation as well as augmentation of the adaptive immune response. Strict control of the alternative pathway of the complement system is crucial to discrimination between healthy host tissues and potentially dangerous cells and particles. Most regulators, including FH, act by selectively protecting self-surfaces. Imbalance or failure in this tight regulation causes several human diseases. Central to the function of FH is its ability to bind to the C3/C5 convertase of the alternative pathway, to C3 fragments deposited on cells and particles and to specific polyanions on self surfaces. Detailed understanding of these binding events will not only shed light on the mechanism of action of this regulator, but also help to understand the mechanisms of diseases associated with an under-regulated alternative pathway.

This study investigates the binding regions of FH for its major targets - the C3 fragments (C3b, C3c and C3d) and polyanionic carbohydrates such as glycosaminoglycans and sialic acid - focusing on the hypothesis that multiple sites on FH are occupied simultaneously by ligands. Factor H consists of 20 CCP modules; truncation fragments containing different numbers of CCP modules were recombinantly expressed, characterised and submitted to heparin-affinity chromatography, gel-mobility-shift assays (GMSA) and surface plasmon resonance. This approach allowed detailed characterisation of the N- and C-terminal binding sites. However, contrary to some reports in the literature, fragments encompassing CCP modules 9 to 15 of FH do not show significant affinity for any of the primary ligands. An atomic resolution NMR structure of the two CCP modules 12 and 13 of FH, i.e. FH-12-13, and architectural information from biophysical analysis of the six-module construct FH-10-15 show that the primary role of this region is not to bind the primary ligands but to form a hinge allowing N- and C-termini to cooperate in engagement with their targets.

Thus this report allows the formulation of a new model for FH action, where some of the central modules orient the key binding sites, at either end of FH, to act in concert facilitating selective protection of non-activating self-surfaces.

Acknowledgments

Primarily, I would like to thank my supervisor Prof. Paul N. Barlow for his help, support, and encouragement over the past four years. I enjoyed both, our discussions about science as well as our debates about a lot of different other unrelated subjects. His encouragement and help writing up this thesis were extraordinary!

Many thanks to the Darwin Trust which made it possible for me to come to Edinburgh.

For his continuous support, invaluable help from the very beginning and for his fun company on different occasions, I would like to give many thanks to Dr. Andy Herbert.

I extend my thanks to Prof. Arthur Rowe for a smooth and fruitful collaboration.

My grateful acknowledgements go to all the people who made it possible for me to experience a fantastic time in Edinburgh. I individually would like to thank Janice, Dušan, Juraj, Henry, Claire, Dave, Julie, Dinesh, Carina, Marie, Isabell, Patience, Carla, Maria, Yoshi, John, Toni, Stine, Huw, Lena, Eve, Bärbel, Conny, Elin, Damia and Marcel.

Special thanks also to everyone who joined me in exploring the fantastic beauty of the Scottish Highlands.

Finally, I would like to thank all my family, my Mum and my Dad and my most special one, Mara.

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. Work described in this thesis has been published in (1-3).

Christoph Schmidt

Contents

Abstract i
Acknowledgments ii
Contents iv
Table of Figures viii
List of Tables xii
Abbreviations xiii
Abbreviations xiii
CHAPTER 1 1
1.1 Innate Immunity 2
1.2 The complement system 2
  1.2.1 Introduction and overview 3
  1.2.2 The alternative pathway 6
  1.2.3 The classical pathway 12
  1.2.4 The lectin pathway 14
  1.2.5 Terminal events 15
  1.2.6 Other roles of complement 16
  1.2.7 Note about complement regulators 17
  1.2.8 The role of complement and disease 19
1.3 Regulators of complement activation (RCA) 21
  1.3.1 Compendium of regulators of complement activation 22
  1.3.2 The complement control protein module 24
1.4 Factor H 26
  1.4.1 Structure of Factor H 26
  1.4.2 Binding sites for C3b 28
  1.4.3 Binding sites for polyanions 32
  1.4.4 Dominant role of the C-terminus of Factor H in host recognition 34
  1.4.5 Other binding partners 35
  1.4.6 Factor H protein family 36
  1.4.7 Atypical haemolytic uraemic syndrome 37
  1.4.8 Dense deposit disease 41
  1.4.9 Age-related macular degeneration 42
  1.4.10 Summary of literature survey 45
1.5 Project aims 46
CHAPTER 2 48

2.1 DNA amplification, cloning, protein production and purification 49
  2.1.1 Polymerase chain reaction (PCR) 49
  2.1.2 Horizontal Agarose Gel Electrophoresis for Nucleic Acids 54
  2.1.3 TOPO® cloning reaction 54
  2.1.4 Restriction enzyme single and double digests 54
  2.1.5 DNA purification and Plasmid DNA extraction 55
  2.1.6 Ethanol precipitation 55
  2.1.7 Phenol-Chloroform extraction 56
  2.1.8 DNA quantification 56
  2.1.9 Ligation reactions 56
  2.1.10 Transformations of plasmids into *E. coli* competent cells and culture growth . 57
  2.1.11 Transformation of *P. pastoris* 57
  2.1.12 *P. pastoris* recombinant protein expression in shaker flasks 60
  2.1.13 *P. pastoris* recombinant protein expression in fermentors 61
  2.1.14 Estimation of protein concentrations 63
  2.1.15 Trichloroacetic acid (TCA) precipitation to concentrate protein samples 64
  2.1.16 Protein concentration in membrane spin concentrators 64
  2.1.17 Sodium dodecyl-sulphate polyacrylamide gel elctrophoresis (SDS PAGE) 64
  2.1.18 Glycoprotein detection 65
  2.1.19 Enzymatic deglycosylation of N-linked glycans 65
  2.1.20 Western blot 65
  2.1.21 Protein chromatography 66
  2.1.22 Mass spectroscopy 68
  2.1.23 N-terminal sequencing 69
2.2 Functional characterisation 69
  2.2.1 Heparin binding studies 69
  2.2.2 Co-factor activity 70
  2.2.3 Binding to C3 fragments C3b, C3c and C3d 70
2.3 NMR structural studies 71
  2.3.1 Sample preparation 71
  2.3.2 Data collection 72
  2.3.3 NMR software 73
  2.3.4 Data processing 73
  2.3.5 General resonance assignment strategy 73
  2.3.6 Assignment of backbone resonances 75

2.3.7 Assignment of side-chain resonances 77
2.3.8 Assignment of aromatic resonances 79
2.3.9 Atom assignment – summary 79
2.3.10 Cis-trans- Proline assignment 79
2.3.11 Assignment of NOE cross-peaks 80
2.3.12 Structure calculation in CYANA 82
2.3.13 Structure calculation in CNS 85
2.3.14 Molecule visualisation programs 88
2.4 Analytical ultracentrifugation 88
CHAPTER 3 89
3.1 Introduction and overview of construct 90
3.1.1 Construct FH-1-4 96
3.1.2 Constructs FH-6-7$_{Y402}$ and FH-6-7$_{H402}$ 101
3.1.3 Constructs FH-6-8$_{Y402}$ and FH-6-8$_{H402}$ 103
3.1.4 Constructs FH-7$_{Y402}$ and FH-7$_{H402}$ 105
3.1.5 Constructs FH-7-8$_{Y402}$ and FH-7-8$_{H402}$ 107
3.1.6 Constructs FH-8-9 and FH8-9$_{K446}$ 109
3.1.7 Construct FH-$_{RR}$8-9 114
3.1.8 Construct FH-8-15 116
3.1.9 Construct FH-10-12 119
3.1.10 Construct FH-10-15 122
3.1.11 Construct FH-11 126
3.1.12 Construct FH-11-14 127
3.1.13 Construct FH-12 130
3.1.14 Construct FH-12-13 131
3.1.15 Construct FH-13-14 and FH-13-14$_{Q802, Q822}$ 135
3.1.16 Construct FH-13 138
3.1.17 Construct FH-13-15 139
3.1.18 Construct FH-14 142
3.2 Conclusions – protein production, purification and characterisation 144
CHAPTER 4 146
4.1 Perspective 147
4.1.1 Analysis of Heparin binding sites in FH 148
4.1.2 Comparison of heparin binding between the AMD-protective 402Tyr and the at-risk 402H allotypic variants in the contexts of CCP 7 and CCPs 6-8 152

4.2 Analysis of C3b-binding sites 154
4.3 Analysis of the binding sites for C3b-cleavage products, C3c and C3d 160
4.4 Conclusions 164
CHAPTER 5 168
5.1 Overview 169
5.2 NMR structure of FH-12-13 170
5.2.1 Structure calculation 171
5.2.2 Description of structure and analysis 175
5.2.3 Analysis of the intermodular interface 181
5.2.4 Analysis of FH-12-13 surface 183
5.3 Analytical ultracentrifugation 185
5.4 Conclusions 190
CHAPTER 6 192
6.1 Factor H has just two principal binding sites each for GAGs and C3b 193
6.2 A model for engagement of FH with C3b on the cell surface (Figure 6.1) 196
6.3 Implications of the working model for the link between mutations/SNPS and disease 199
BIBLIOGRAPHY 201
APPENDIX A 213
APPENDIX B 215
APPENDIX C 218
APPENDIX D 222
APPENDIX D 227

Table of Figures

Figure 1.1 Complement cascade 5
Figure 1.2 High resolution crystal structure of human C3 8
Figure 1.3 Activation of C3 9
Figure 1.4 Formation of alternative pathway C3-convertase 11
Figure 1.5 Regulation of the alternative pathway C3-convertase 12
Figure 1.6 Structure of C1q
Figure 1.7 Sketch representation of RCAs proteins 22
Figure 1.8 Sequence alignment. 24
Figure 1.9 High resolution structure of CCP module 25
Figure 1.10 Multiple sequence alignment of 20 short consensus repeats of complement factor H (FH). 27
Figure 1.11 Summary of module-deletion/truncations of FH. 30
Figure 1.12 CCP module and linker size in FH 45
Figure 2.1 15N-HSQC spectrum. 74
Figure 2.2 Nomenclature of sequential amino acids in a polypeptide chain 75
Figure 2.3 Magnetisation transfer and 3-D spectra representation of CBCA(CO)NH and CBCANH experiments.. 76
Figure 2.4 Triple resonance NMR CBCA(CO)NH (purple) and CBCANH (green) spectra of FH-12-13. 77
Figure 2.5 Aliphatic proton assignment using 15N-TOCSY-HSQC and the HCCH-TOCSY. 78
Figure 2.6 Symmetry related NOE cross-peaks. 84
Figure 3.1 Summary of FH segments employed in this study. 90
Figure 3.2 Flowchart showing steps involved in the generation of genetically modified *P. pastoris* clones for recombinant protein expression. 92
Figure 3.3 Gradient SDS-PAGE of FH-1-4 96
Figure 3.4 Purification of FH-1-4.. 98
Figure 3.5 Mass spectrometric analysis of FH-1-4 100
Figure 3.6 Cofactor activity assay of FH-1-4 for factor I-mediated cleavage of C3b 100
Figure 3.7 PCR screen of *E. coli* colonies 102
Figure 3.8 Mini-scale expression study of FH-6-7. 102
Figure 3.9 Plasmid DNA preparation of construct FH-6-8$_{Y402}$ and mini-scale expression study of *P. pastoris* clones FH-6-8$_{H402}$ and FH-6-8$_{Y402}$; 103

Figure 3.10 Expression and purification of FH-6-8_{H402};.. 104
Figure 3.11 Mass spectrum of tryptic-peptide fingerprint of FH-6-8_{H402}........................... 105
Figure 3.12 Expression and purification of FH-7_{Y402}.. 106
Figure 3.13 *M*ini-scale expression study of FH-7-8;... 107
Figure 3.14 Expression and purification of FH-7-8_{Y402};.. 108
Figure 3.15 Mass spectrum of peptide fingerprint of construct FH-7-8_{Y402}........................ 109
Figure 3.16 Gradient SDS-PAGE of mini-scale expression study of two *P. pastoris* clones expressing FH-8-9 110
Figure 3.17 Log of ^{15}N-FH-8-9 fermentation.. 110
Figure 3.18 Initial purification of FH-8-9. 111
Figure 3.19 Second purification step for FH-8-9; 112
Figure 3.20 Validation of FH-8-9.. 113
Figure 3.21 Gradient SDS-PAGE of mini-scale expression trial of four *P. pastoris* clones expressing FH-8-9_{K446}.. 114
Figure 3.22 Expression and purification of FH-$_{RR}$8-9;.. 115
Figure 3.23 Sequence coverage in peptide mass fingerprinting for FH-$_{RR}$8-9.. 116
Figure 3.24 First purification step of FH-8-15. 117
Figure 3.25 Second purification step of FH-8-15.. 117
Figure 3.26 Validation of FH-8-15; 118
Figure 3.27 Gradient SDS-PAGE of recombinantely expressed FH-10-12. 119
Figure 3.28 First purification step of FH-10-12.. 120
Figure 3.29 Second purification step of FH-10-12.. 121
Figure 3.30 Validation of FH-10-12... 122
Figure 3.31 Gradient SDS-PAGE of *P. pastoris* FH-10-15 expression. 123
Figure 3.32 First purification step of FH-10-15 124
Figure 3.33 Second purification step of FH-10-15.. 124
Figure 3.34 Validation of FH-10-15... 125
Figure 3.35 Gradient SDS-PAGE of recombinant FH-11.. 127
Figure 3.36 Gradient SDS-PAGE of *P. pastoris* FH-11-14 expression. 128
Figure 3.37 First purification step of FH-11-14 128
Figure 3.38 Second purification step of FH-11-14.. 129
Figure 3.39 Validation of FH-11-14... 130
Figure 3.40 Gradient SDS-PAGE of *m*ini-scale expression trial of five *P. pastoris* clones expressing FH-12.. 131
Figure 3.41 Gradient SDS-PAGE of *P. pastoris* FH-12-13 expression. 131

Figure 3.42 Purification of FH-12-13 132
Figure 3.43 Validation of FH-12-13.......... 134
Figure 3.44 Gradient SDS-PAGE of *P. pastoris* FH-13-14$_{Q802\_Q822}$ expression.......... 135
Figure 3.45 Gradient SDS-PAGE of *P. pastoris* FH-13-14 expression. 136
Figure 3.46 Size-exclusion chromatogram of FH-13-14.......... 137
Figure 3.47 Partial validation of FH-13-14.......... 138
Figure 3.48 Expression and purification of FH-13.......... 139
Figure 3.49 Gradient SDS-PAGE of *P. pastoris* FH-13-15 expression trials. 140
Figure 3.50 Purification of FH-13-15.......... 141
Figure 3.51 Validation of FH-13-15.......... 142
Figure 3.52 Expression-level comparison on gradient SDS-PAGE. 143
Figure 4.1 Factor H segments employed in this study and results of heparin-affinity chromatography. 148
Figure 4.2 Gel-mobility shift assays.......... 149
Figure 4.3 A mixture of His402 and Tyr402 CCP 6-8 (triple modules) applied on a Poros heparin-affinity column. 152
Figure 4.4 Affinity of FH-7 and FH-6-8 for GAGs 156
Figure 4.6 Surface plasmon resonance experiments to identify C3b-binding segments of FH.. 157
Figure 4.7 Use of SPR to measure dissociation constants for major C3b-binding sites in FH. 159
Figure 4.8 Surface plasmon resonance experiments to identify C3b-, C3c- and C3d-binding segments of FH.. 162
Figure 4.9 Surface plasmon resonance experiments with FH and FH-19-20 to identify affinities for C3b, C3c, and C3d. 163
Figure 4.10 Use of SPR to measure FH-19-20 dissociation constant for C3d.. 164
Figure 4.11 Bar-chart diagram summarising K_d values determined for C3b binding of FH 166
Figure 5.1 Assigned ^{15}N-HSQC spectrum of FH-12-13.......... 171
Figure 5.2 Report from CYANA structure calculation cycles for FH-12-13.. 173
Figure 5.3 Energy plot of 100 structures from the final round of CNS structure calculations 174
Figure 5.4: Ensemble of NMR-derived structures.. 175
Figure 5.5 Characteristics of CCP in the structure of FH-12-13 176
Figure 5.6 Ramachandran statistics.......... 178
Figure 5.7 ^{15}N-HSQC spectrum of FH-12-13 30 minutes after resuspension in D_2O.......... 179

Figure 5.8 Amino acids corresponding to resonances unaffected by the hydrogen/deuterium exchange 180

Figure 5.9 Type IV β-turn 181

Figure 5.10 Sketch representation of the eight linker residues between module CCPs 12 and 13 182

Figure 5.11 Surface and cartoon representations of FH-12-13. 183

Figure 5.12 GRASP electrostatic surface representations of FH-12-13........ 185

Figure 5.13 A screenshot showing the outcome of processing the data from a sedimentation velocity AUC run........ 187

Figure 5.14 Schematic to illustrate the thinking behind the AUC-based experiments and their interpretation........ 189

Figure 6.1 In this model FH binds to surface-bound C3b. 198

List of Tables

Table 1.1 Effects of complement activation....4

Table 1.2 Summary of alternative pathway proteins....7

Table 1.3 Complement regulators....19

Table 1.4 Mutations in complement control proteins (CCPs) 19 and 20 of FH....39

Table 2.1 Primer sequences....50

Table 2.2 PCR cycling parameters....51

Table 2.3 Mastermix PCR cycling parametres....52

Table 2.4 Sequencing cycle programme....53

Table 2.5 PCR cycling parameters for QuickChange® site-directed mutagenesis....53

Table 2.6 NMR experiments used in this study....73

Table 2.7 List of ΔC^{β}-C^{γ} values and conformation assignment for all proline residues in FH-12-13....80

Table 2.8 Relative NOE intensity rating and translation into distance restraints....86

Table 3.1 Sequences of recombinant proteins used in this study....92

Table 3.2 Overview of recombinantly expressed proteins....95

Table 4.1 Chip-surface summary....155

Table 4.2 Derived K_d values for interaction of C3b with FH, and with FH fragments that correspond to its two major binding sites....160

Table 5.1 Summary of results from sedimentation velocity AUC....188

Abbreviations

2D	two-dimensional
3D	three-dimensional
AMD	Age related macular degeneration
aHUS	Atypical haemolytic uraemic syndrome
C4BP	C4b-binding protein
CA	Cofactor activity
CCP	complement control protein
CNS	Crystallography and NMR system
CR1	complement receptor type 1
CR2	complement receptor type 2
CRP	C-reactive protein
DAA	Decay accelerating activity
DAF	decay accelerating factor
FH	factor H
FI	factor I
GMB	Glomerular basement membrane
HSQC	heteronuclear single quantum coherence
K_d	equilibrium dissociation constant
M_r	relative molecular mass
MAC	membrane attack complex
MASP	mannan-binding lectin associated serine protease
MBL/MBP	mannan-binding lectin/mannan-binding protein
MCP	membrane cofactor protein
MPGN	Membrano-proliferative glomerulonephritis
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PDB	protein data bank
RCA	regulator of complement activation
RMSD	root mean square deviation
SCR	short consensus repeat
SPR	surface plasmon resonance
TOCSY	total correlation spectroscopy
VCP	*Vaccinia* virus complement-binding protein

CHAPTER 1

INTRODUCTION

1.1 Innate Immunity

The immune system protects the host organism by recognizing and disposing of invading microorganisms and potentially pathogenic modified host constituents. In vertebrates adaptive, or acquired (specific), immunity evolved alongside the more phylogenetically ancient (4) innate (or non-specific) immune system. The effectors of the innate immune system have the ability to recognise danger in a non-specific fashion and in a way that usually precedes responses mediated by the adaptive immune system. Thus innate immunity delivers prompt action against microbial threats and self-generated hazards (5). The innate immune response is also activated by antibody-antigen complexes. Moreover, the activation products of the innate immune system induce an enhanced state of awareness amongst some B- and T-lymphocytes. Thus there is crosstalk between innate and adaptive immunity. Both of these defence systems have humoral and cellular effector arms.

1.2 The complement system

The complement system is a collection of proteins that comprise the major humoral defence mechanism of the innate immune system. When it was first discovered it was described as a heat-sensitive factor in serum that "complemented" the effects of specific antibodies in bacterial lysis (6). The importance of complement is reflected in the fact that soluble complement components make up to 5% of the total protein content in human plasma (4).

1.2.1 Introduction and overview

The complement system comprises a network of over 30 serum and cell-surface proteins (7). It has three, interconnected, activation routes: the 'classical pathway', the 'lectin pathway' and the 'alternative pathway'. Each of these has the potential to proceed to a common 'terminal pathway' that generates a catalytic membrane attack complex. All three activation pathways depend upon a proteolytic cascade whereby the activation of pro-enzymes creates functional enzymes in a series of steps that amplify an initial response. Non-enzymatic products of proteolysis include mediators of inflammation and enhancers of adaptive immunity. Different stimuli trigger each of the three pathways. The classical pathway is generally activated by antigen-antibody complexes, while the closely related lectin pathway recognises foreign carbohydrate motifs. Spontaneous, low-level and continuous activation is the distinctive feature of the alternative pathway. This is referred to as 'tick over'. Rapid amplification of the alternative pathway ensues in the absence of action of specific complement regulators.

The central components of the complement system, like C3 and factor B[1], have their evolutionary origins more than 1300 million years ago (8). Paralogues are found in invertebrates like the sea urchin and horseshoe crab. In a recent study, C3 was also found in corals and sea anemones. For comparison, the evolutionary origin of the adaptive immune system (major histocompatibility complex – MHC) dates back to around 600 million years and is specific to jawed vertebrates. The oldest activation pathway of the complement system is the alternative pathway, based on proteins C3 and factor B. It is speculated that the lectin pathway became established around 900 million years ago with the appearance of mannan binding protein (MBP), MBP-associated serine protease (MASP) and ficolins. The classical pathway is thought to have emerged around 600 million years ago by vertebrate-specific complement gene duplications between C3/C4/C5, factor B/C2 and MASP/C1r/s.

---

[1] Note on Nomenclature of the complement system:
Proteins of the classical pathway and the membrane attack system are allocated a certain number (they are listed here in their sequential reaction order): C1q, C1r, C1s, (C4, C2, C3), C5, C6,C7,C8,C9
Suffix letters distinguish proteolytically produced fragments from their parent molecule. Usually the suffix letter "a" characterises the smaller product of a cleaving reaction, whereas the letter "b" designates the bigger cleavage product. Proteins of the alternative pathway are called "Factors".
There are three different methods of characterizing complement receptors. The receptors can be named either according to their ligand (*e.g.* C5a receptor), or using the cluster of differentiation (CD) system. Additionally there is a numbering system for receptors for the major fragments of C3. This distinguishes four different types of complement receptors (*e.g.* CR1 to CR4). Thus there can be three different notations for one protein (*e.g.* the receptor for C3b: C3b receptor, CD35 and CR1).

Activation of these pathways may results in different responses (5, 6), (9) (summarised in Table 1.1).

Effect	Mechanism
elimination of pathogens	opsonisation and clearance by phagocytosis
	direct lysis through activation of terminal pathway membrane attack complex (MAC)
clearance of apoptotic bodies	opsonisation and clearance by phagocytosis
clearance of immune complexes	recognition of immune complexes by C1q promotes clearance by macrophages and dentirtic cells
mediation of inflammation	release of anaphylatoxic and chemotactic proteins
regulation of adaptive immune responses	interaction of complement activation products with cell surface receptors of myeloid, lymphoid and stromal cells

Table 1.1 Effects of complement activation

The high levels of complement proteins circulating in serum (and draining into adjacent tissues) are constitutively expressed by hepatocytes (5). During acute infections, production of some complement proteins (*e.g.* C3 and MBL) is up-regulated and thus they may be classified as acute phase proteins (10). Local generation of complement components expressed by different cells (like macrophages, gut epithelial cells and fibroblasts) is critical for supplying a source of complement to peripheral lymph nodes and difficult-to-access tissues like skin and brain.

All three activation pathways converge at the pivotal event of the complement cascade, the formation of C3-convertases. Subsequent proteolytic activation of the glycoprotein C3 (that circulates permanently in the bloodstream at levels of about 1-2 mg/ml) produces the anaphylatoxin C3a and the opsonin C3b. C3b is able to participate in further reactions that culminate in the formation of the membrane attack complex (MAC). A set of specific regulators of the complement cascade are necessary to maintain a delicate steady state between activation and repression of complement activity.

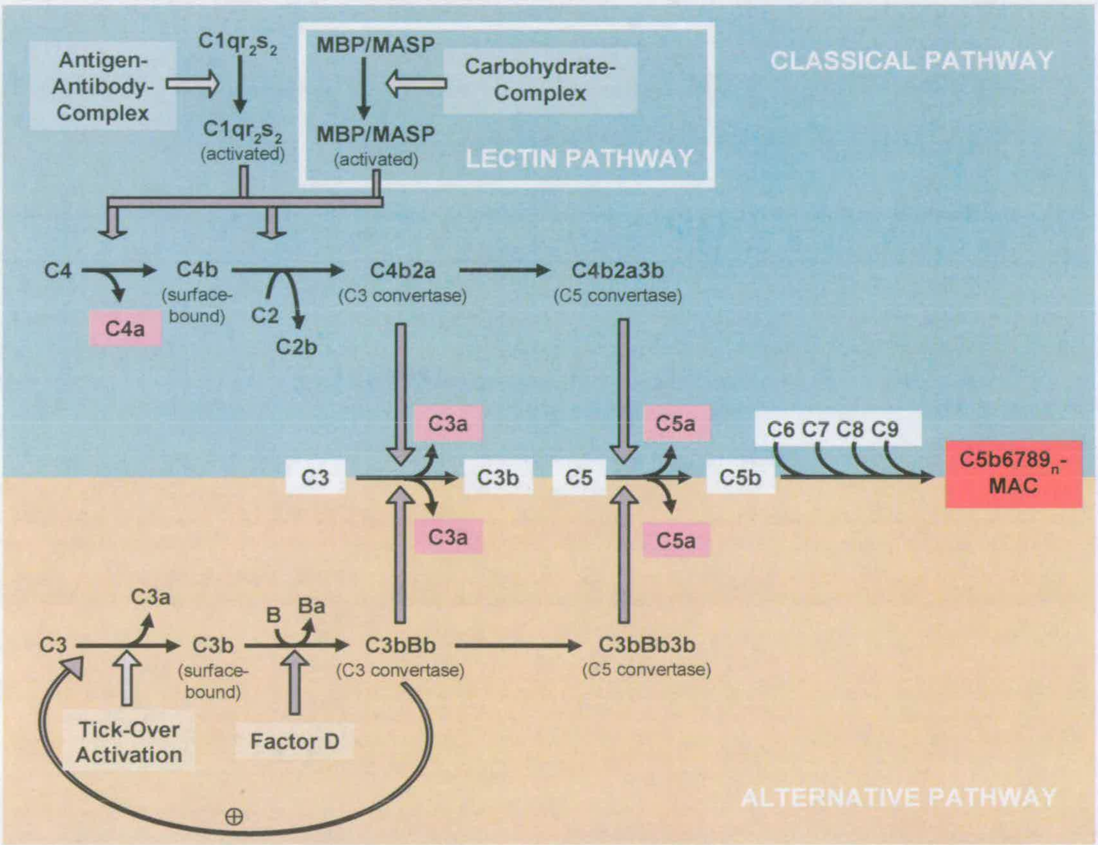

Antigen-Antibody-Complex
C1qr2s2
C1qr2s2 (activated)
MBP/MASP
MBP/MASP (activated)
Carbohydrate-Complex
LECTIN PATHWAY
CLASSICAL PATHWAY
C4
C4a
C4b (surface-bound)
C2
C2b
C4b2a (C3 convertase)
C4b2a3b (C5 convertase)
C3a
C5a
C6 C7 C8 C9
C3
C3b
C5
C5b
C5b6789n-MAC
C3a
C5a
C3a
B
Ba
C3
C3b (surface-bound)
C3bBb (C3 convertase)
C3bBb3b (C5 convertase)
Tick-Over Activation
Factor D
⊕
ALTERNATIVE PATHWAY

Figure 1.1 Complement cascade. Each of the three activation routes (classical, lectin and alternative pathway) can induce the formation of C3 and C5 convertases resulting in the generation of many central complement activation products (C3a, C3b, C5a, C5b). Nascent C5b recruits the components of the terminal events necessary for the formation of the MAC. Anaphylatoxic protein fragments are coloured purple. For more details please see the relevant sections below.

1.2.2 The alternative pathway

Overview

The alternative pathway is characterised by a constant, low-level, spontaneous cleavage of C3 into C3a and C3b (the "tick-over" reaction – this process involves a form of C3 known as C3(H2O) which is discussed below), exposing and activating the reactive thioester group that is deployed by C3b for attachment to nucleophilic groups. Under normal circumstances most newly generated C3b molecules are quickly cleaved by factor I in association with one of the complement regulators, factor H, complement receptor type I (CR1, CD35) or membrane cofactor protein (MCP, CD46) and the cleavage products are no longer able to participate in the complement cascade. Some C3b molecules interact with factor B forming C3b,B. In the context of such a complex, factor B is susceptible to the cleavage by the serine protease factor D. The fragment Bb remains associated with C3b, forming the C3 convertase, C3b,Bb, of the alternative pathway. In the presence of regulators (*e.g.* on a the membrane of a host cell that typically contains multiple complement regulators) the C3 convertase has an extremely short half-life and no amplification ensues. In the absence of regulators (*e.g.* on some bacterial surfaces) fresh molecules of C3 are cleaved by C3b,Bb to C3b and C3a thus stoking a positive-feedback loop (see Figure 1.1 and below for a detailed mechanism). Moreover, nascent C3b molecules bind to existing C3b,Bb complexes to form $C3b_2$,Bb; the presence of the extra C3b molecule provides a binding site for C5 (11) (12), and C3b2Bb is the C5 convertase of the alternative pathway that generates C5a and C5b. This is the first step leading to the terminal membrane attack complex.

protein	function	activated by	activated into	gained function
C3	carrier of reactive thioester	tick-over ($C3(H_2O)$,Bb) or C3b,Bb	C3a	anaphylatoxin
			C3b	reactive thioester
factor D	serine protease	n/a	n/a	n/a
factor B	pro-enzyme	factor D	Ba	not definitely established
			Bb	serine protease

Table 1.2 Summary of alternative pathway proteins

Central complement component C3

Recent crystallographic determination of high-resolution structures of C3 (13), C3b (14) and factor B (15) provide detailed insight into the mechanism of alternative pathway activation. Proteolysis, complex formation and conformational changes are the underlying molecular principles of this pathway. The events of C3 cleavage, C3b attachment to surfaces and the steps leading to the formation of the alternative pathway C3-convertase will be discussed here on the basis of graphical presentations from (16). Complement component C3 is a member of the C3/α2-macroglobulin family of host-defence molecules. A characteristic feature of this family of molecules, which are also found in groups as diverse as insects and nematodes, is the all-alpha helical thioester domain (TED). C4 and C5 are also member of this family and each has 26-30% sequence identity to C3. However, C5 lacks the reactive thioester that, once activated, covalently attaches the parent protein to molecular and cellular targets. Human C3 (Figure 1.2) is synthesised as a 1641-amino acid residue precursor polypeptide chain. Prior to secretion the C3 precursor undergoes posttranscriptional cleavage at a tetra-arginine sequence (Arg646-649). Mature C3 contains 13 domains within its two polypeptide chains, the α–chain and β–chain. Eight homologous macroglobulin (MG) domains form the core of the molecule.

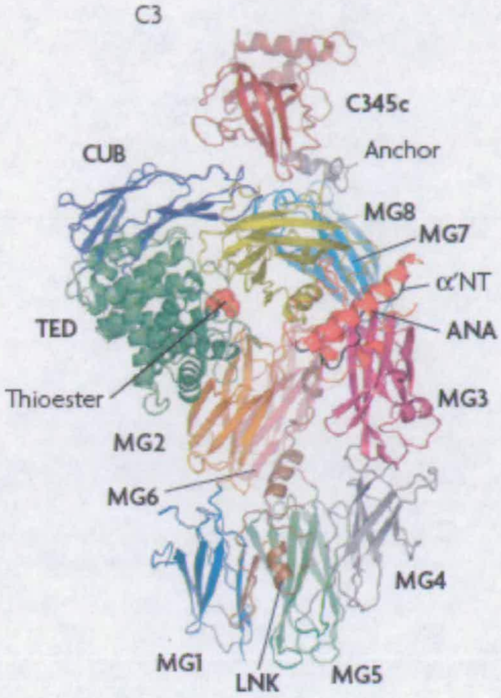


Figure 1.2 High resolution crystal structure of human C3 (from (13)). α'NT, N-terminus of the α-chain; ANA, anaphylatoxin domain; CUB, complement C1r/C1s, UEGF, BMP1; LNK, linker; MG, macroglobulin; TED, thioester-containing domain.

The first of two insertions containing the linker domain (LNK) and the anaphylatoxin domain (ANA) (and including the cleaved tetra-arginine segment) separates the primary sequence of MG6 into MG6β and MG6α components (Figure 1.2). The second insertion (between MG7 and MG8) adds the CUB domain (complement C1r/C1s, UEGF, BMP1) and the TED. The C-terminal extension forms the C345c domain.

Activation of C3

The thioester moiety of C3 is buried between the TED and MG8 domain. Thus shielded from reactions with nucleophiles it takes from hours to days for native C3 to react with amino or hydroxyl nucleophiles (see below). Activation by C3 convertases cleaves the small anaphylatoxin C3a from C3 triggering conformational rearrangement in the remaining C3b (Figure 1.3).

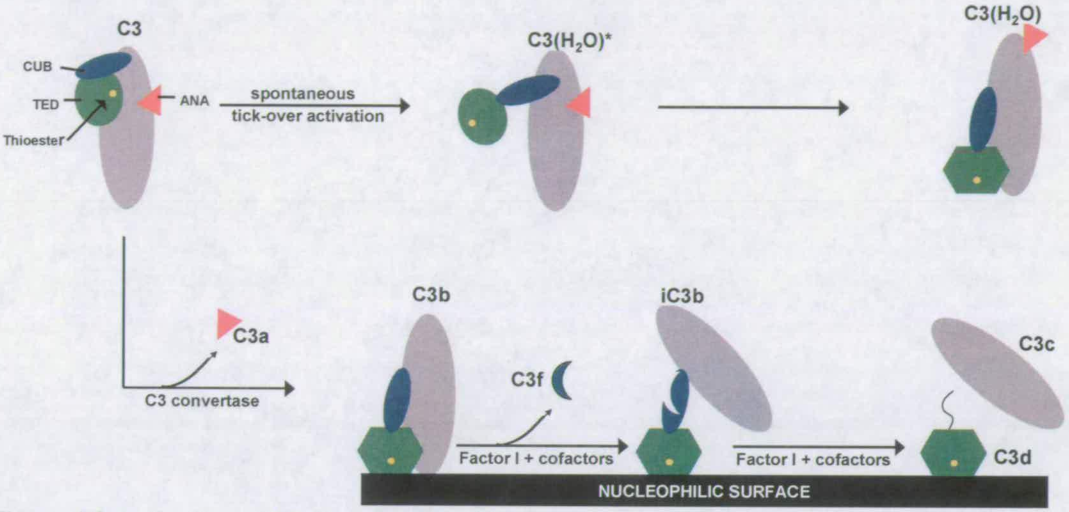

C3
CUB
TED
Thioester
ANA
spontaneous
tick-over activation
C3(H2O)*
C3(H2O)
C3a
C3 convertase
C3b
C3f
iC3b
C3c
Factor I + cofactors
Factor I + cofactors
C3d
NUCLEOPHILIC SURFACE

Figure 1.3 Activation of C3. Spontaneous (tick-over) activation is shown (top). Activation via C3 convertase followed by inactivation of factor I mediated cleavage in the presence of cofactors is shown (below). Figure adapted from (16).

Removal of the anaphylatoxin domain, which stabilises the TED-MG8 interface, allows the CUB and TED domain to swing out. Following solvent exposure and relocation by 85 Å of the thioester group, the TED undergoes conformational changes that promote formation of the highly reactive thiolate and acyl-imidazole intermediate. This short-lived, meta-stable intermediate reacts very quickly (half-life time less than 100 µs for reaction with hydroxyl groups) with available nucleophiles.

Molecules of C3b generated close to a nucleophile-bearing surface are likely to become attached to the surface via their thiol esters. If on the other hand formation of C3b occurs far from an activating surface the nucleophile is likely to be the oxygen of a water molecule (16). Indeed, intact C3 itself is able to undergo tick-over activation due to spontaneous hydrolysis of its thioester. This reaction proceeds at a very low rate as is obvious from the 230-hour half-life of C3 in water. Electron microscopic data is consistent with a C3 conformational intermediate, C3(H_2O)*, in which the TED is swung out and away from the main body of C3. This intermediate has a half-life of about one hour, and it either flips back into its native C3 conformation or undergoes hydrolysis and an associated irreversible conformational change yielding C3(H_2O). It is suggested that in C3(H_2O) the anaphylatoxin domain remains attached but is translocated away from its position in native C3. As a consequence of such conformational changes C3(H_2O) is able to interact with factor B, promote B cleavage and form part of a C3 convertase, C3(H_2O),Bb. In terms of comparing the two processes whereby C3 can become activated (*i.e.* via formation of C3(H_2O) versus cleavage to C3b) it has been suggested that the relative chances of the structural transition from native C3 to an activated conformer (C3(H_2O) or C3b) depend on the likelihood that the anaphylatoxin domain is translocated from its native position. According to this viewpoint, proteolytic activation to C3b results in highly probable and therefore relatively fast conversion to the highly reactive C3b; on the other hand the unassisted conformational transition to C3(H_2O) is a rare, and therefore slow event.

Formation of the alternative pathway C3 convertase

Both C3b and C3(H_2O) are able to nucleate assembly of alternative pathway C3 convertases by interacting with factor B, resulting either in surface-bound C3b,Bb or in fluid phase C3b,Bb or C3(H_2O) (17). The formation of fluid-phase and surface-bound C3 convertases require very similar or identical sequences of binding events; these are discussed below in the context of surface-bound C3 convertase formation.

In the local absence of specific regulators of complement activation, a C3b-coated surface provides a starting point for the formation of alternative pathway C3 convertases and the positive-feedback loop that generates additional C3b (and C3a). Factor B is an inactive pro-enzyme that adheres to its cofactor C3b in an Mg^{2+}-dependent manner and undergoes proteolytical activation (to Bb and Ba) by Factor D to form the active C3 convertase C3b,Bb

(Figure 1.4). In contrast to other complement serine proteases, Factor D circulates in its active form and does not require proteolytic activation.

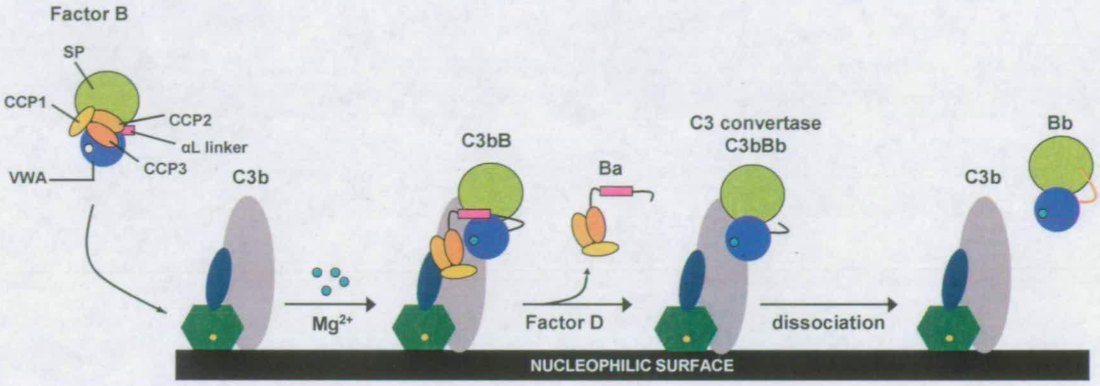


Figure 1.4 Formation of alternative pathway C3-convertase. Factor B comprises five domains; Three complement control protein (CCP) domains, a long linker domain (αL linker) containing the scissile bond for Factor D cleavage, Von Willebrand Factor A (VWA), and the C-terminal serine protease (SP) domain. Figure adapted from (16).

Typically, the alternative C3 convertase has an *in-vivo* half-life of about 90 seconds. However, binding of human glycoprotein properdin to C3b,Bb complexes can prolong its lifetime. Once dissociated, Bb is unable to re-associate with C3b. It is thought that a part of the factor D-cleaved linker domain in Bb, which is attached to Von Willebrand factor A domain (VWA), prevents re-association of the two convertase components. However, C3b, if not cleaved by factor I/factor H, can again serve as a starting point for C3 convertase formation.

Note on regulation of C3 convertases

Generation of opsonising C3b and anaphylatoxic C3a molecules is potentially harmful to host tissue and therefore requires tight regulation. Several surface-bound proteins along with the soluble plasma protein factor H regulate alternative pathway C3 convertase activity (Figure 1.5). These homologous 'regulators of complement activation' (RCA) act in multiple ways. For example, factor H competes with factor B for binding to C3b. Factor H, CR1 and decay accelerating factor (DAF, CD55) accelerate the decay of the active convertase complexes (*i.e.* they exhibit decay accelerating activity (DAA))

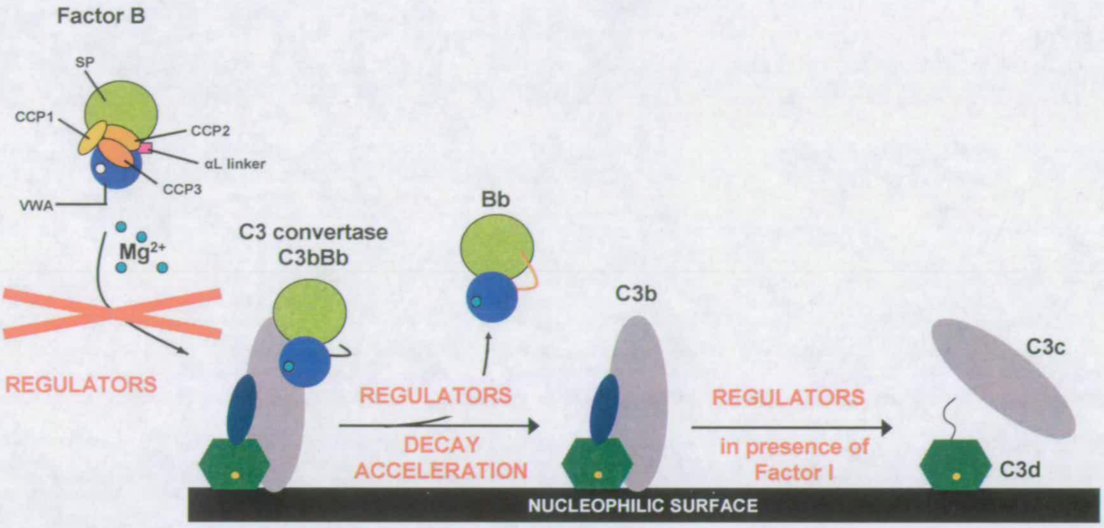

Factor B
SP
CCP1
CCP2
αL linker
CCP3
VWA
Mg2+
REGULATORS
C3 convertase
C3bBb
Bb
C3b
REGULATORS
DECAY
ACCELERATION
REGULATORS
in presence of
Factor I
C3c
C3d
NUCLEOPHILIC SURFACE

Figure 1.5 Regulation of the alternative pathway C3-convertase. Regulators of complement activation have decay accelerating (DAA) and/or cofactor (for factor I mediated cleavage) activity. Figure adapted from (16).

into their components, C3b and Bb, resulting in a reduction of convertase *in-vivo* half-life from ~90 to ~10 seconds. Factor H, MCP and CR1 act as cofactors for the proteolytic inactivation of C3b (to iC3b, or ultimately to C3c and C3d and C3g in the case of CR1) by factor I – this is termed cofactor activity (CA).

1.2.3 The classical pathway

The binding of immune complexes to complement component C1q (complexed with C1r and C1s to form C1) triggers the classical pathway. The F_c portions of immunoglobulin (Ig) G and IgM bound to antigen participate in a multivalent interaction with C1q, thus linking the adaptive antibody-derived immune response with the innate complement-derived response. It has been reported that direct binding to C1q of other substances such as bacterial lipopolysaccharide, viral glycoproteins, nucleic acids, prion infectious agent and fragments of damaged cells can also initiate the classical pathway in an antibody-independent manner (7).(12).

C1q is a member of the collectin protein family and it contains both lectin domains and collagen-like domain. Six globular heads linked together by collagen-like tails form the structure of C1q, which is non-covalently (Ca^{2+}-dependent) associated with two copies each of each the serine proteases C1r and C1s (*i.e.* $C1qr_2s_2$). Binding of two or more of the C1q heads to an immune complex (or other activating surface) is required to trigger a conformational change in the C1-complex necessary for the activation of C1r (Figure 1.6) (5). Active C1r then proteolytically activates C1s, which cleaves C4 into the two fragments C4a and C4b. A reactive thioester group on C4b attaches on the surface next to the C1 complex. The pro-enzyme C2 (homologous to the alternative pathway pro-enzyme factor B) then binds, in a Mg^{2+}-dependent step, to C4b and in this context is cleaved by the activated C1s into C2a and C2b.

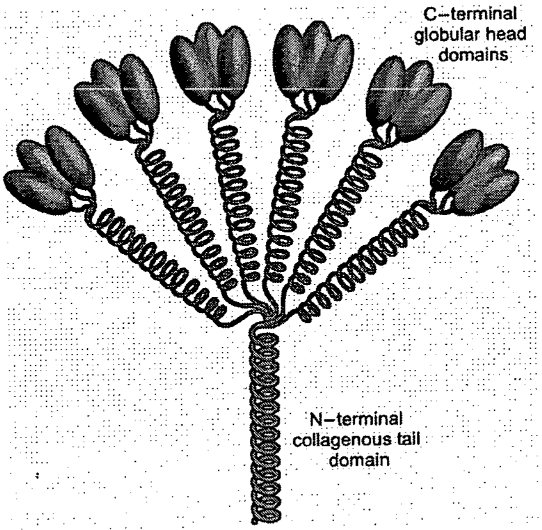

C–terminal globular head domains
N–terminal collagenous tail domain

Figure 1.6 Structure of C1q (derived from(4)). The N-terminal collagenous tail-domain connects the C-terminal globular head domains.

C2b remains in complex with the surface-attached C4b. The resultant C4b,2a complex is the classical pathway C3 convertase and exhibits a strong protease activity specific for C3 thereby generating C3b with its activated thioester group and ability to nucleate the alternative pathway convertase, as discussed earlier. Nascent C3b may also complex with C4b,2a forming the C4b,2a,3b complex, which is the C5 convertase of the complement pathway. The C5b generated by C4b,2a,3b thus represents the point of convergence between classical and alternative pathways along the route to the terminal pathway. The smaller cleavage fragments C4a, C3a and C5a are potent anaphylatoxins and trigger inflammation.

1.2.4 The lectin pathway

The lectin pathway activates complement on microbial surfaces and is similar to the classical pathway. The pathway is triggered by binding of MBL or ficolins (L-ficolin or H-ficolin) to various carbohydrate motifs that are abundant on pathogens (10). MBL is – like C1q of the classical pathway - a six-headed collectin belonging to the C-type or Ca^{2+}-dependent lectin super-family. Binding of sugar moieties to the MBL head domain (carbohydrate recognition domain) is facilitated by the formation of hydrogen bonds between four amino acid residues of the lectin domain and the pathogen-derived carbohydrate structures. Specific patterns of polar groups within the carbohydrates determine the selectivity of MBL binding. For example, ligands of MBL may be listed in order of decreasing affinity: N-acetly glucosamine > L-fucose, mannose, N-acetly mannosamine > maltose > glucose >> galactose, N-acetyl galactosamine. The low affinity of the MBL carbohydrate recognition site for single monosaccharides means that multiple engagement with repeating microbial carbohydrates is a necessity. Discrimination between self and non-self results from the galactose and sialic acid that form the termini of most host carbohydrate structures; these inhibit MBL binding.

Once activating carbohydrates are recognised by MBL, the MBL-associated serine proteases (MASP1 and 2; closely related (5) to the classical pathway proteases C1r and C1s) become activated. The liganded MBL/MASP complex triggers the activation of C4 to C4b in an analogous process to the equivalent steps of the classical pathway; C4 is cleaved to C4b which binds to C2 forming the C3 convertase. Subsequently C2 is cleaved generating the bimolecular complex C4b,C2a, which corresponds to the point of conversion between lectin and classical pathways. The acute-phase protein, C-reactive protein, is also capable of activating the lectin pathway.

1.2.5 Terminal events

All three pathways of complement activation converge at the point of proteolytic generation from C5 of the complement component, C5b. This process initiates the so-called terminal pathway wherein five soluble plasma proteins (C5-C9) self-assemble to create a pore that inserts into phospholipids bilayers - the membrane attack complex (MAC) (18). The consequent disruption of a bacterial membrane (for example) leads to osmotic lysis and neutralisation of the infective species.

The terminal pathway is thus instigated by cleavage of C5 by C4b,2a,3b or $C3b_2$,Bb. The smaller, C5a, product is the most potent of the anaphylatoxins. The C5b fragment acts as a nucleation site for assembly of a C5b,6,7 complex. This trimolecular complex is lipophilic and anchors itself in the lipid bilayer. This process is thought to convert the C5b,6,7 complex into a membrane-integrated high-affinity receptor for C8. The nascent C5b,6,7,8 (i.*e.* C5b-8) complex thus has ability to lyse certain microorganisms and eukaryotic cells. More importantly it serves as a landing pad for multiple copies of C9. Only one or two C9 molecules incorporated into C5b-8 are sufficient to elicit lysis. More typically, incorporation of 12 to 18 C9 molecules results in formation of an annular structure enclosing a 10 nm open pore through the membrane (there are parallels here with perforin - a pore-forming protein of cytotoxic killer cells and natural killer cells) (19).

1.2.6 Other roles of complement

Complement is best known for its ability to recognise and dispose of pathogens (7). Two other important functions are modulation of adaptive immune responses and maintenance of tissue homeostasis by means of mediating waste disposal (20).

Normal turnover in the body generates billions of dying cells daily and the resultant cell debris must be cleared in a non-inflammatory way. It is emerging that complement is a key component in performing endogenous recycling functions by recognition and opsonisation of cellular waste products – the initial step towards clearance by phagocytosis. For example, it has been shown that CRP, IgM, C1q and MBL are involved in binding of apoptotic bodies and triggering of the enhanced uptake of apoptotic particles by the mononuclear phagocyte system. DNA exposed on apoptotic material (10) is hypothesised to be the ligand for MBL. MBL facilitates phagocytosis by cells that bear receptors for the collagenous tail domain of MBL. Similarly, direct binding of C1q to apoptotic cells and immune complexes promotes clearance via recognition of the C1q collagenous tail domain by macrophages and dentritic cells (7). C3b deposition in response to alternative pathway activation triggers opsonisation not only of microorganisms, but also of altered self-cells (e.g. apoptotic or cancer cells (21-23)) enabling complement mediated clearance. It has been suggested that this process arises from the loss of membrane-associated complement regulators by damaged cells (due to apoptosis, injury, infection or tumor transformation) (7). Indeed, complement dysfunction may lead to impaired waste disposal, accumulation of debris, development of inflammation and autoreactivity resulting in damage to self-tissues; a causative or worsening factor for many inflammation-triggered diseases (for more information see 1.2.8 The role of complement and disease).

Complement activation also enhances, both indirectly and directly, the adaptive immune responses (6) (10). Anaphylatoxins promote inflammation and recruit cellular effectors of the innate immune system (monocytes, macrophage, neutrophils, basophiles and eosinophils) responsible for phagocytosis. Coating with opsonins promotes rapid ingestion of recognised danger particles and cells by the recruited phagocytic cells. Recognition, uptake and digestions of antigens in these cells trigger the release of cytokines and enable the presentation of these antigens to B-cells and T-cells. Both, cytokine release and antigen presentation are crucial mechanisms for improved adaptive immune responses.

In terms of direct cross-talk between innate and adaptive immunity, cleavage products of complement activation and regulation interact with many different cell-surface receptors on myeloid, lymphoid and stromal cells facilitating "communication" between the complex networks of native and adaptive immune responses. The following specific interactions are illustrative. Stimulation of B-cell complement receptors CD21/complement receptor 2 (CR2), which binds iC3b, C3dg and C3d, and CD35/CR1, which binds C4b, C3b, iC3b, C3dg and C3d, significantly enhances B-cell activation (6); co-ligation of CR2 and B-cell-antigen-receptor with C3d-coated antigenic particles (in combination with antigen presentation of CR2-postive dentritic cells towards B-cells) lowers the threshold of B-cell antigen response by a factor of 1000 (24). Absence of CR1 and CR2 (in knock-out mice) resulted in abnormal B-cell responses and impaired generation of memory B-cells. Recent studies explored the role of complement in regulating T-cell activity (25). Although direct interactions of T-cell complement receptors (*e.g.* the receptor for C3a) with complement activation products are discussed, concrete mechanisms remain elusive. A major role for T-cell priming is ascribed to antigen presenting cells in lymphoid tissues, which mature through stimulation of their complement receptors: complement receptor 3 (CR3), complement receptor 4 (CR4), C3a-receptor and C5a receptor.

1.2.7 Note about complement regulators

As already discussed, activation of complement is potentially harmful to host tissue, and its activation must be kept in check (9). Many of the complement regulators are members of the

RCA family (5). These proteins are encoded by the RCA gene cluster and will be discussed in more detail below (section 1.3 Regulators of complement activation). Table 1.3 gives a brief overview of important complement inhibitors. Although complement regulation is found at nearly all steps throughout the complement cascade, most regulators control the central amplification process, the conversion of C3 and C5 to C3b and C5b.

Name	Abbr.	Membrane bound	main family	Function
C1-inhibitor	C1-Inh	no	serpins	inhibits spontaneous activation of C1
Factor I	-	no	serine protease	proteolytical inactivation of C3b and C4b
Complement Receptor 1	CR1, CD35	yes	CCP	CA for C3b and C4b DAA of C3 & C5 convertases of both pathways
Membrane Cofactor protein	MCP, CD46	yes	CCP	CA for C3b and C4b
Decay accelerating protein	DAF, CD55	yes	CCP	DA for both pathways
C4-binding protein	C4BP	no	CCP	CA for C4b DA of classical pathway convertases of classical pathway
Factor H	FH	no	CCP	CA for C3b DA of C3 & C5 convertases of alternative pathway
Vitronectin (S protein)	-	no	Hemopexin	binds to C5b67 - thus inhibiting binding of C5b67 to membranes
Protectin	CD59 MAC-IP	yes	extracellular domain of	inhibits MAC formation

			cell surface receptors	

Table 1.3 Complement regulators

1.2.8 The role of complement and disease

The complement system is a powerful, indispensable, system for the maintenance of host protection and tissue homeostasis. Deficiencies in complement components or regulators of the complement cascade are associated with increased susceptibility to infection (9) and several diseases. For example, hereditary angioedema results from deficiencies of C1-inhibitor (the key inhibitor of classical pathway C1r/C1s proteases and an inhibitor of enzymes in the coagulation, fibrinolytic and kinin cascades). Autosomal mutations or auto-antibodies against C1-inhibitor result in impaired regulation that causes periodic swellings in multiple organs (20).

Conditions associated with mutations or polymorphisms in the key regulator of the alternative pathway, FH, include dense deposit disease (DDD), atypical haemolytic uraemic syndrome (aHUS) and age related macular degeneration (AMD). These conditions will be discussed in sections 1.4.4 to 1.4.6.

Complement also plays a secondary, or non-causative role, in many clinically adverse circumstances (9) that are characterised by prolonged or chronic inflammation. Examples include Alzheimer's disease, allo- and xenotransplantation, asthma, burn injuries, Crohn's disease, glomerulonephritis, haemolytic anaemia, ischemia/reperfusion injuries, immune-complex induced vasculitis, multiple-system organ failure, multiple sclerosis, myasthenia gravis, psoriasis, rheumatoid arthritis, septic shock, systemic lupus erythematosus and stroke. Synthetic, natural or modified complement inhibitors are being developed and tested in animal and clinical studies as a means to reduce complement-triggered tissue damage in various clinical conditions.

INTRODUCTION

A wide range of pathogens – including some viruses, bacteria, fungi and parasites - express proteins that interfere with complement activation either by recruiting host complement regulators (*e.g.* C4BP of factor H), or by directly controlling various steps of the complement cascade. Some of the latter provide pointers for the development of new anti-inflammatory drugs. One example is the *Vaccina* virus complement control protein (VCP) and its homologues amongst the other pox viridae. In this case, the virus appears to have 'hijacked' a mammalian RCA gene; the sequences of the four CCP modules of VCP is highly similar to that of the N-terminal four CCP modules of C4BP. VCP has been shown to inhibit both classical and alternative pathways (CA and DAA) (26), to enhance the virulence of the virus through these complement regulatory activities.

Another two examples of complement regulators from bacterial origin are the C5a peptidase of group A and group B *streptococci*, which is an established virulence factor that inactivates the anaphylatoxin C5a (17), and the *S. aureus* protein SCIN. The latter is expressed by 90 % of all *S. aureus* strains and was shown to strongly down-regulate C5a production. SCIN is also associated with reduced levels of both, membrane attack complex formation and alternative-pathway mediated opsonisation (C3b-deposition) (27). Unlike other host or pathogen derived complement regulators SCIN binds exclusively to the activated C3 convertases (alternative- and classical pathway convertases) and increases the stability of the complexes. It is thought that this unique regulatory feature either locks the convertases' catalytic centre and/or block substrate binding.

1.3 Regulators of complement activation (RCA)

To avoid damage to self-tissue it is obviously important that the activity of the complement system is restricted to its target cells. Furthermore, complement activation must be spatially and temporarily limited in order to prevent excessively widespread or prolonged activation because it is important not to exhaust the whole stock of complement proteins in one event (and thereby lessen responsiveness to later events). There are several different regulators of complement activity that differ in distribution and mobility as a consequence of their membrane-bound or soluble states. Arguably, the most important family of regulatory proteins are the RCA (*vide supra* – 1.2.7 Note about complement regulators). These proteins are encoded in a cluster of genes on chromosome 1 (1q32). The RCA are unusual examples of multiple-domain proteins in that they consist entirely (or almost entirely) of tandemly arranged examples of a single module-type – the complement control protein (CCP) module (also known as short consensus repeats or sushi domains) (28). The 60 amino acid residue consensus sequence of the CCP module includes four invariant cysteines, a nearly invariant tryptophan and highly conserved prolines, glycines and hydrophobic residues (Figure 1.7, Figure 1.8). The linkers connecting the individual modules (*i.e.* between the last cysteine of a CCP and the first cysteine of the next CCP) usually consist of four amino acids residues, but their lengths vary from three to eight residues (see section 1.3.2 The complement control protein module). CCP modules occur not only in RCA, but also in other proteins that are not thought to interact with the complement system (*e.g.* IL2-receptor, beta2-glycoprotein I, Factor XIII of the blood clotting system, and the GABA receptor type 1 (29)).

The soluble regulator of the alternative pathway of the complement system FH (and the FH-related proteins, some of which also have complement regulatory activity) consists exclusively of 20 CCP modules, whereas the other regulators - MCP (four CCP modules), DAF (four CCP modules), CR1 (most commonly 30 CCP modules) and C4BP (alpha chains

– eight CCP modules, beta chain – three CCP modules) - consist almost entirely of CCP modules.

1.3.1 Compendium of regulators of complement activation

The selective action of RCAs ensures down-regulation of complement activation on most self-surfaces, but allows amplification of complement to proceed on surfaces of microorganisms or other unwanted particles, thus ensuring 'organised' elimination. Due to the aforementioned positive feedback loop for C3b generation, the RCAs have apparently evolved to act upon the early steps of activation, specifically the formation and stability of the convertase complexes.

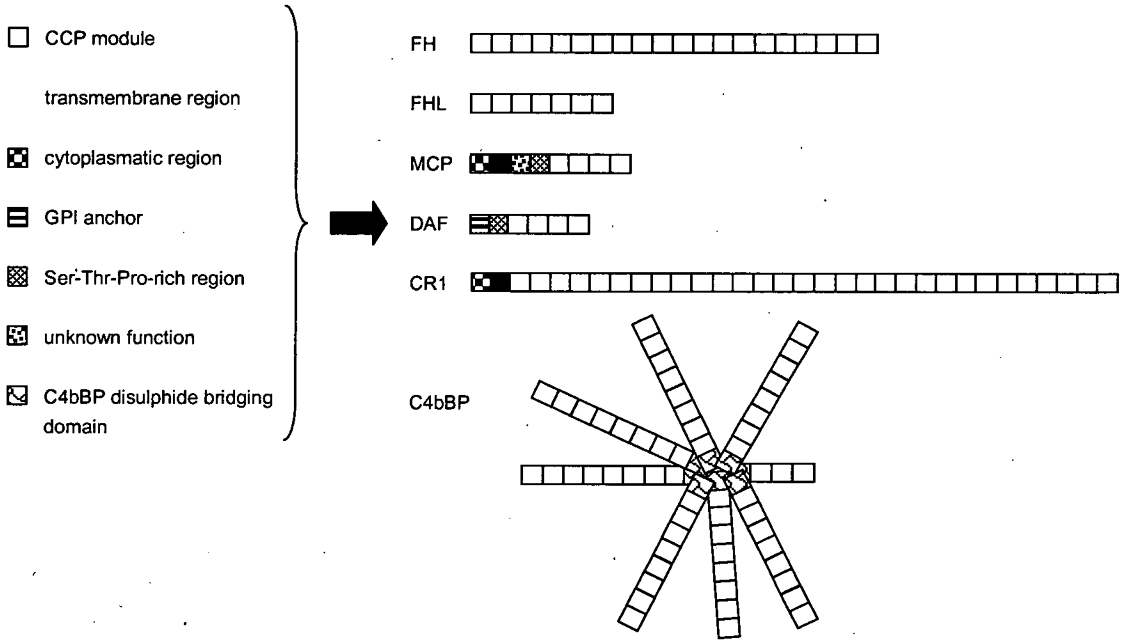

CCP module
transmembrane region
cytoplasmatic region
GPI anchor
Ser-Thr-Pro-rich region
unknown function
C4bBP disulphide bridging domain
FH
FHL
MCP
DAF
CR1
C4bBP

Figure 1.7 Sketch representation of RCAs proteins (28), (30), (12), (31).

Overview of domain composition and functional activities of RCA proteins:
CA (cofactor activity for factor I mediated cleavage)
DA (decay accelerating activity)

FH is a 20 CCP-module containing plasma glycoprotein.
CA: for C3b

DA: for alternative pathway C3 convertase (C3bBb) and C5 convertase (C3bBbC3b – short $C3b_2Bb$)

FHL is an alternative splice product of FH containing the first seven CCP-modules.
CA: for C3b
DA: for C3bBb and $C3b_2Bb$ (alternative pathway)

MCP (or CD46) is a trans-membrane protein. MCP is comprised of four CCP modules, an O-glycosylated serine/threonine/proline-rich domain, a trans-membrane region and an intracellular region.
CA: for C3b and C4b
DA: none

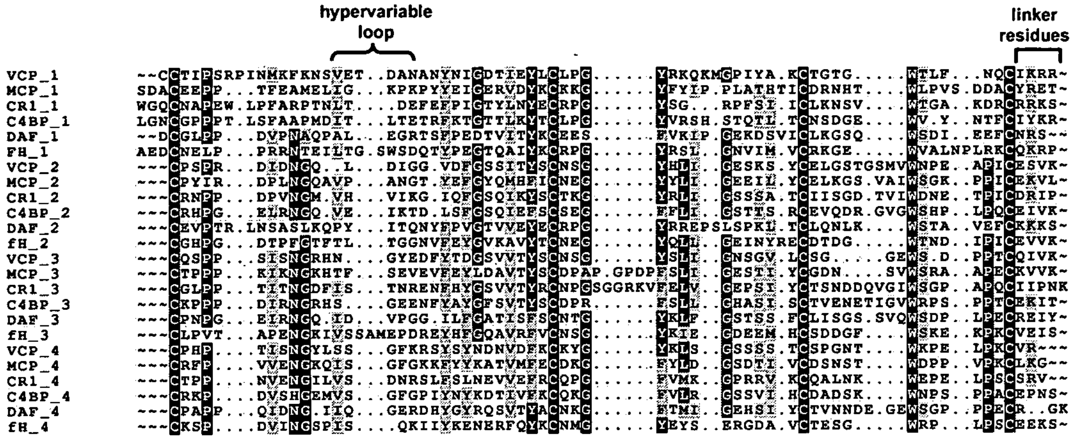
DAF (or CD55) is composed of four CCP modules followed by a serine/threonine-rich domain, which attaches the molecule to the membrane via a glycosylphosphatidylinositol anchor.
CA: none
DA: for C3bBb and $C3b_2Bb$ (alternative pathway); for C4b2a and C4b2a3b (C3 and C5 convertases of classical pathway)

CR1 (or CD35) is a cellular receptor. The most common allotype is composed of 30 CCP modules, a trans-membrane and an intracellular domain.
CA: for C3b and C4b
DA: for C3bBb and $C3b_2Bb$ (alternative pathway); for C4b2a and C4b2a3b (classical pathway)

C4BP has a spider like structure and spans a heptameric plasma protein. The major isoform consists of seven α-chains (eight CCP modules per α-chains) and one β-chain (three SCR per β-chain), adding up to 59 CCP modules in all. The C-termini of both chains have additional regions, which polymerise the single chains by disulfide formation forming the mature protein.
CA: for C4b (in vitro also observed for C3b)
DA: for C4b2a and C4b2a3b (classical pathway)

1.3.2 The complement control protein module

A sequence alignment of the first four CCP modules of six different complement regulators (human MCP, CR1, C4BP, DAF, FH and viral VCP) illustrates the consensus sequence (Figure 1.8).



```
VCP_1       --CCTIPSRPINMKFKNSVET..DANANYNIGDTIEYLCLPG......YRKQKMGPIYA.KCTGTG.....WTLF...NQCIKRR~
MCP_1       SDACEEPP...TFEAMELIG...KPKPYYEIGERVDYKCKKG......YFYIP.PLATHTICDRNHT....WLPVS.DDACYRET~
CR1_1       WGQCNAPEW.LPFARPTNLT....DEFEFPIGTYLNYECRPG......YSG...RPFSI.ICLKNSV....WTGA..KDRCRRKS~
C4BP_1      LGNCGPPPT.LSFAAPMDIT...LTETRFKTGTTLKYTCLPG......YVRSH.STQTL.TCNSDGE....WV.Y..NTFCIYKR~
DAF_1       --DCGLPP...DVPNAQPAL...EGRTSFPEDTVITYKCEES......FVKIP.GEKDSVICLKGSQ....WSDI..EEFCNRS~~
FH_1        AEDCNELP...PRRNTEILTG.SWSDQTYPEGTQAIYKCRPG......YRSL..GNVIM.VCRKGE.....WVALNPLRKCQKRP~
VCP_2       ---CPSPR...DIDNGQ..L...DIGG.VDFGSSITYSCNSG......YHLI..GESKS.YCELGSTGSMVWNPE..APICESVK~
MCP_2       ---CPYIR...DPLNGQAVP...ANGT.YEFGYQMHFICNEG......YYLI..GEEIL.YCELKGS.VAIWSGK..PPICEKVL~
CR1_2       ---CRNPP...DPVNGM.VH...VIKG.IQFGSQIKYSCTKG......YRLI..GSSSA.TCIISGD.TVIWDNE..TPICDRIP~
C4BP_2      ---CRHPG...ELRNGQ.VE...IKTD.LSFGSQIEFSCSEG......FFLI..GSTTS.RCEVQDR.GVGWSHP..LPQCEIVK~
DAF_2       ---CEVPTR.LNSASLKQPY...ITQNYFPVGTVVEYECRPG......YRREPSLSPKL.TCLQNLK....WSTA..VEFCKKKS~
fH_2        ---CGHPG...DTPFGTFTL...TGGNVFEYGVKAVYTCNEG......YQLL..GEINYRECDTDG.....WTND..IPICEVVK~
VCP_3       ---CQSPP...SISNGRHN....GYEDFYTDGSVVTYSCNSG......YSLI..GNSGV.LCSG.....GEWS.D..PPTCQIVK~
MCP_3       ---CTPPP...KIKNGKHTF...SEVEVFEYLDAVTYSCDPAP.GPDPFSLI..GESTI.YCGDN....SVWSRA..APECKVVK~
CR1_3       ---CGLPP...TITNGDFIS...TNRENFHYGSVVTYRCNPGSGGRKVFELV..GEPSI.YCTSNDDQVGIWSGP..APQCIIPNK
C4BP_3      ---CKPPP...DIRNGRHS....GEENFYAYGFSVTYSCDPR......FSLL..GHASI.SCTVENETIGVWRPS..PPTCEKIT~
DAF_3       ---CPNPG...EIRNGQ.ID...VPGG.ILFGATISFSCNTG......YKLF..GSTSS.FCLISGS.SVQWSDP..LPECREIY~
fH_3        ---CLPVT...APENGKIVSSAMEPDREYHFGQAVRFVCNSG......YKIE..GDEEM.HCSDDGF....WSKE..KPKCVEIS~
VCP_4       ---CPHP....TISNGYLSS...GFKRSYSYNDNVDFKCKYG......YKLS..GSSSS.TCSPGNT....WKPE..LPKCVR~~~
MCP_4       ---CRFP....VVENGKQIS...GFGKKFYYKATVMFECDKG......FYLD..GSDTI.VCDSNST....WDPP..VPKCLKG~~
CR1_4       ---CTPP....NVENGILVS...DNRSLFSLNEVVEFRCQPG......FVMK..GPRRV.KCQALNK....WEPE..LPSCSRV~~
C4BP_4      ---CRKP....DVSHGEMVS...GFGPIYNYKDTIVFKCQKG......FVLR..GSSVI.HCDADSK....WNPS..PPACEPNS~
DAF_4       ---CPAPP...QIDNG.IIQ...GERDHYGYRQSVTYACNKG......FTMI..GEHSI.YCTVNNDE.GEWSGP..PPECR..GK
fH_4        ---CKSP....DVINGSPIS....QKIIYKENERFQYKCNMG......YEYS..ERGDA.VCTESG.....WRP...LPSCEEKS~
```

Figure 1.8 Sequence alignment. Alignment of selected CCP modules (VCP, MCP, CR1, C4BP, DAF, FH - identifier at the beginning of each sequence). Sequences were aligned using CLUSTAL W (2.0) (32). Conserved residues are shaded according to conservation using BOXSHADE version 3.21 (33). A square bracket indicates the linker region between the last cysteine of the previous CCP and the first cysteine of the consecutive CCP. A curly brace indicates the hypervariable (hv) loop, which varies in sequence and length between different CCP modules. The hv-loop's border to the C-terminus consists of a relatively conserved hydrophobic amino acid (shaded grey), which either can be one of the two aromatic amino acids tyrosine and phenylalanine, or the aliphatic amino acids isoleucine or leucine.

The three-dimensional structures of numerous examples of CCP modules have been determined experimentally. Typically a hydrophobic core, containing the highly conserved tryptophan and other conserved residues, is sandwiched between small anti-parallel beta-sheets. The four cysteines form disulfide bridges (Cys-I-Cys-III and Cys-II-Cys-IV (see Figure 1.9). In some cases, neighbouring CCP modules within a protein are thought to stabilise one another. From a functional perspective, tandem CCP modules often form a composite ligand-binding surfaces (28). Such intermodular cooperation – influenced by linker-length and the extent and chemical nature of intermodular interfaces - presumably contributes critically towards the functional diversity of the numerous proteins (both within the complement system and elsewhere) that utilise these modules to recognise and bind to a wide range of partners. Moreover, two or more such distinct groups of neighbouring CCP modules within an RCA protein molecule may bind simultaneously to a common target molecule (*vide infra*). Therefore there is a particular interest in the determination not just of individual CCP structures within the RCAs but also in elucidating the overall architecture of the parent proteins.

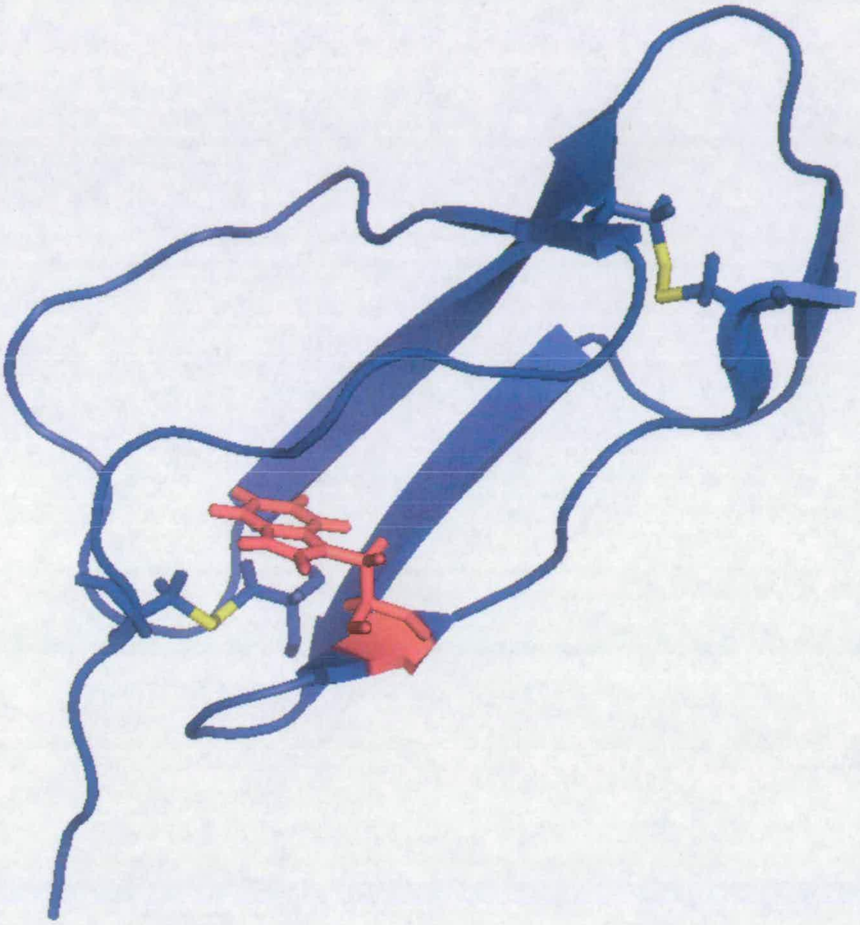


Figure 1.9 High resolution structure of CCP module. Ribbon diagram of the tertiary structure of FH CCP16 (pdb-code: 1HCC). The N-terminus is on the left, the C-terminus on the right. Broad arrows represent ß-strands that are connected by loops and areas of non-standard secondary structure (34), (35). The two disulfide bonds are presented in yellow. The conserved tryptophan is shown in red.

1.4 Factor H

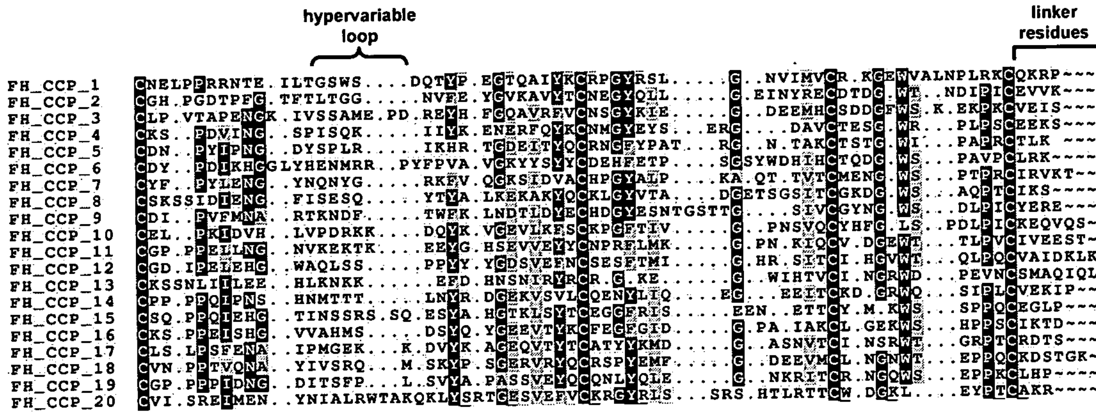
Factor H is a 155-kDa soluble glycoprotein regulator of the complement system[2]. It is abundant in plasma (300-790 μg/ml plasma concentrations (36)) and can associate with host cell membranes and other self-surfaces via recognition of polyanions such as glycosaminoglycans (GAGs) and sialic acid (37). Through intervention at the level of the alternative-pathway C3 and C5 convertase enzymes it modulates both fluid-phase and surface-associated complement amplification. Factor H works in several ways (38): it competes with factor B for binding to C3b thus impeding formation of alternative-pathway C3 convertases (C3bBb); when bimolecular convertase complexes do succeed in assembling, FH accelerates their subsequent dissociation (decay); FH also accelerates decay of the alternative-pathway C5 convertase ($C3b_2Bb$); and FH is a cofactor for factor I-mediated proteolytic cleavage of C3b to iC3b (see also sections 1.2.2 The alternative pathway and 1.3.1 Compendium of regulators of complement activation).

1.4.1 Structure of Factor H

The 1213 amino acid residues of mature FH (155 kDa) (39) (40) consist of 20 "short consensus repeats" (SCR; SCR is another term for CCP), each of ~60 residues (41) and containing the consensus sequence of a CCP module. A multiple alignment of the 20 SCRs (Figure 1.10) highlights the four invariant Cys residues and a near-invariant Trp residue between Cys(III) and (IV), as was discussed above.

---

[2] Parts of the section 1.4 Factor H have been published in (2. Schmidt, C.Q., A.P. Herbert, H.G. Hocking, D. Uhrin, and P.N. Barlow. 2008. Translational mini-review series on complement factor H: structural and functional correlations for factor H. *Clin Exp Immunol* 151:14-24.



```
FH_CCP_1   CNELPPRRNTE.ILTGSWS....DQTYP.EGTQAIYKCRPGYRSL.....G..NVIMVCR.KGEWVALNPLRKCQKRP~~~~
FH_CCP_2   CGH.PGDTPFG.TFTLTGG.....NVFE.YGVKAVYTCNEGYQLL.....G.EINYRECDTDG.WT..NDIPICEVVK~~~~
FH_CCP_3   CLP.VTAPENGK.IVSSAME.PD.REYH.FGQAVRFVCNSGYKIE.....G..DEEMHCSDDGFWS.K.EKPKCVEIS~~~~
FH_CCP_4   CKS..PDVING..SPISQK.....IIYK.ENERFQYKCNMGYEYS...ERG....DAVCTESG.WR...PLPSCEEKS~~~~
FH_CCP_5   CDN..PYIPNG..DYSPLR.....IKHR.TGDEITYQCRNGFYPAT...RG..N.TAKCTSTG.WI...PAPRCTLK ~~~
FH_CCP_6   CDY..PDIKHGGLYHENMRR..PYFPVA.VGKYYSYYCDEHFETP....SGSYWDHIHCTQDG.WS...PAVPCLRK~~~~~
FH_CCP_7   CYF..PYLENG..YNQNYG.....RKFV.QGKSIDVACHPGYALP....KA.QT.TVTCMENG.WS...PTPRCIRVKT~~~
FH_CCP_8   CSKSSIDIENG..FISESQ.....YTYA.LKEKAKYQCKLGYVTA....DGETSGSITCGKDG.WS...AQPTCIKS~~~~~
FH_CCP_9   CDI..PVFMNA..RTKNDF.....TWFK.LNDTLDYECHDGYESNTGSTTG....SIVCGYNG.WS...DLPICYERE~~~~
FH_CCP_10  CEL..PKIDVH..LVPDRKK....DQYK.VGEVLKFSCKPGFTIV.....G..PNSVQCYHFG.LS..PDLPICKEQVQS~~
FH_CCP_11  CGP.PPELLNG..NVKEKTK....EEYG.HSEVVEYYCNPRFLMK.....G.PN.KIQCV.DGEWT...TLPVCIVEEST~~
FH_CCP_12  CGD.IPELEHG..WAQLSS.....PPYY.YGDSVEFNCSESFTMI.....G.HR.SITCI.HGVWT...QLPQCVAIDKLKK
FH_CCP_13  CKSSNLIILEE..HLKNKK......EFD.HNSNIRYRCR.G.KE......G..WIHTVCI.NGRWD...PEVNCSMAQIQL~
FH_CCP_14  CPP.PPQIPNS..HNMTTT.....LNYR.DGEKVSVLCQENYLIQ....EG...EEITCKD.GRWQ...SIPLCVEKIP~~~
FH_CCP_15  CSQ.PPQIEHG..TINSSRS.SQ.ESYA.HGTKLSYTCEGGFRIS.....EEN..ETTCY.M.KWS...SPPQCEGLP~~~~
FH_CCP_16  CKS.PPEISHG..VVAHMS.....DSYQ.YGEEVTYKCFEGFGID.....G.PA.IAKCL.GEKWS...HPPSCIKTD~~~~
FH_CCP_17  CLS.LPSFENA..IPMGEK...K.DVYK.AGEQVTYTCATYYKMD.....G..ASNVTCI.NSRWT...GRPTCRDTS~~~~
FH_CCP_18  CVN.PPTVQNA..YIVSRQ...M.SKYP.SGERVRYQCRSPYEMF.....G..DEEVMCL.NGNWT...EPPQCKDSTGK~~
FH_CCP_19  CGP.PPPIDNG..DITSFP...L.SVYA.PASSVEYQCQNLYQLE.....G..NKRITCR.NGQWS...EPPKCLHP~~~~~
FH_CCP_20  CVI.SREIMEN..YNIALRWTAKQKLYSRTGESVEFVCKRGYRLS...SRS.HTLRTTCW.DGKL....EYPTCAKR~~~~~
```

Figure 1.10 Multiple sequence alignment of 20 short consensus repeats in sequence of complement factor H (FH). One-letter codes used throughout; Sequences were aligned using CLUSTAL W (2.0) (32). Conserved residues are shaded according to conservation using BOXSHADE version 3.21 (33). A square bracket indicates the linker region between the last cysteine of the previous CCP and the first cysteine of the consecutive CCP. Invariant Cys residues and almost invariant Trp residue are highlighted in black. Each short consensus repeat (SCR) probably folds into a complement control protein (CCP) module;

While circular dichroism spectra of FH were uninterpretable (due to contributions from the 40 putative disulfide bonds) (42), Fourier-transform infrared spectroscopy (43) confirmed extensive β-sheet in FH and a paucity of α-helices. Infrared spectroscopy also demonstrated rapid exchange of backbone amide protons with solvent, indicative of an elongated structure for FH in which the multiple CCPs do not, in general, contribute to a common globular arrangement. This inference of an extended structure concurs with electron microscopy, small angle-X ray scattering and analytical ultracentrifugation studies of C4BP, FH and other RCAs (44) (45) (46) (47). Currently, a "folded-back" model for FH is favoured, bringing together its N- and C-terminal regions (48).

There are 3D structures currently available for eleven out of the 20 FH CCPs that have been expressed recombinantly as single, double or triple modules (49) (35) (50) (51) (52) (1) (53) (and data presented in this thesis); and reliable homology-based models have been produced for several others. Each CCP has an ovoid structure (Figure 1.9), of dimensions approximately 40 Å by 15 Å by 10 Å. Five extended stretches of residues (that often form β-strands and small anti-parallel β-sheets), run back or forth in a direction that is approximately parallel with the long axis of the CCP. Thus the module's N and C termini occupy opposite poles consistent with a "head-to-tail" arrangement of adjacent modules. Indeed, 3D structures of the module pair consisting of CCPs 15 and 16 (*i.e.* FH-15-16) (50), and of the triple-module FH-6-8 (53) are elongated with small intermodular buried surface areas. The recently determined high resolution structure of the three N-terminal CCP modules of FH revealed similar structural characteristic bulges or loops, corresponding to insertions and areas of low sequence conservation, project laterally from the body of the module potentially contributing to binding specificity. Since stabilising interactions between neighbouring modules are limited, intermodular flexibility is possible. The degree of overall flexibility of FH is unknown as are the extents of any conformational changes upon interaction with binding partners.

1.4.2 Binding sites for C3b

The N-linked glycans of FH are dispensable for complement regulation (54) , although whether they electrostatically modulate interactions between FH and surface-borne polyanions requires further investigation. Early mapping of functional sites to specific FH CCP modules focused on a 38-kDa tryptic N-terminal fragment, with fluid-phase C3b-binding and cofactor activity (55) (56), corresponding to CCPs 1-5 plus part of CCP 6. A 42-kDa FH splice variant (FH-like 1) containing CCPs 1-7 is likewise able to regulate fluid-phase complement (57). To pinpoint key functional modules, fluid-phase cofactor activity was measured for (non-purified) module-deletion and truncation mutants of the 38-kDa fragment expressed recombinantly and secreted from Chinese hamster ovary cells (58) (Figure 1.11). The results imply that the four N-terminal CCPs are required for full cofactor activity in the fluid phase, although FH-1-3 (and to a lesser extent FH-2-4) retained residual activity. A subsequent study (59) (involving constructs prepared in a baculovirus expression vector) largely reinforced these findings: while the FH-1-4 construct displayed full fluid-phase cofactor activity, the triple-module constructs FH-1-3, FH-2-4, FH-1,2,4 and FH-1,3,4 lacked it, as did the four-module construct FH-1,6,7,4 (where CCPs 6 and 7 replace CCPs 2 and 3), and FH-1-4 with non-native linker lengths. Subsequently, CCPs 1-4 were also shown to be required for the decay accelerating activity of this molecule, although full-length FH is apparently ~100-fold more potent on surfaces than FH-1-4 or FH-like 1 in this respect. In summary (60), the N-terminal four CCPs of FH are necessary and sufficient to engage with C3b and C3 convertase in the fluid phase and thereby regulate amplification of the cascade via the alternative pathway. Interestingly, patients with an amino acid residue deletion in CCP 4 developed dense deposit disease (61), a renal pathology also seen in factor H-deficient humans, pigs and mice (63).

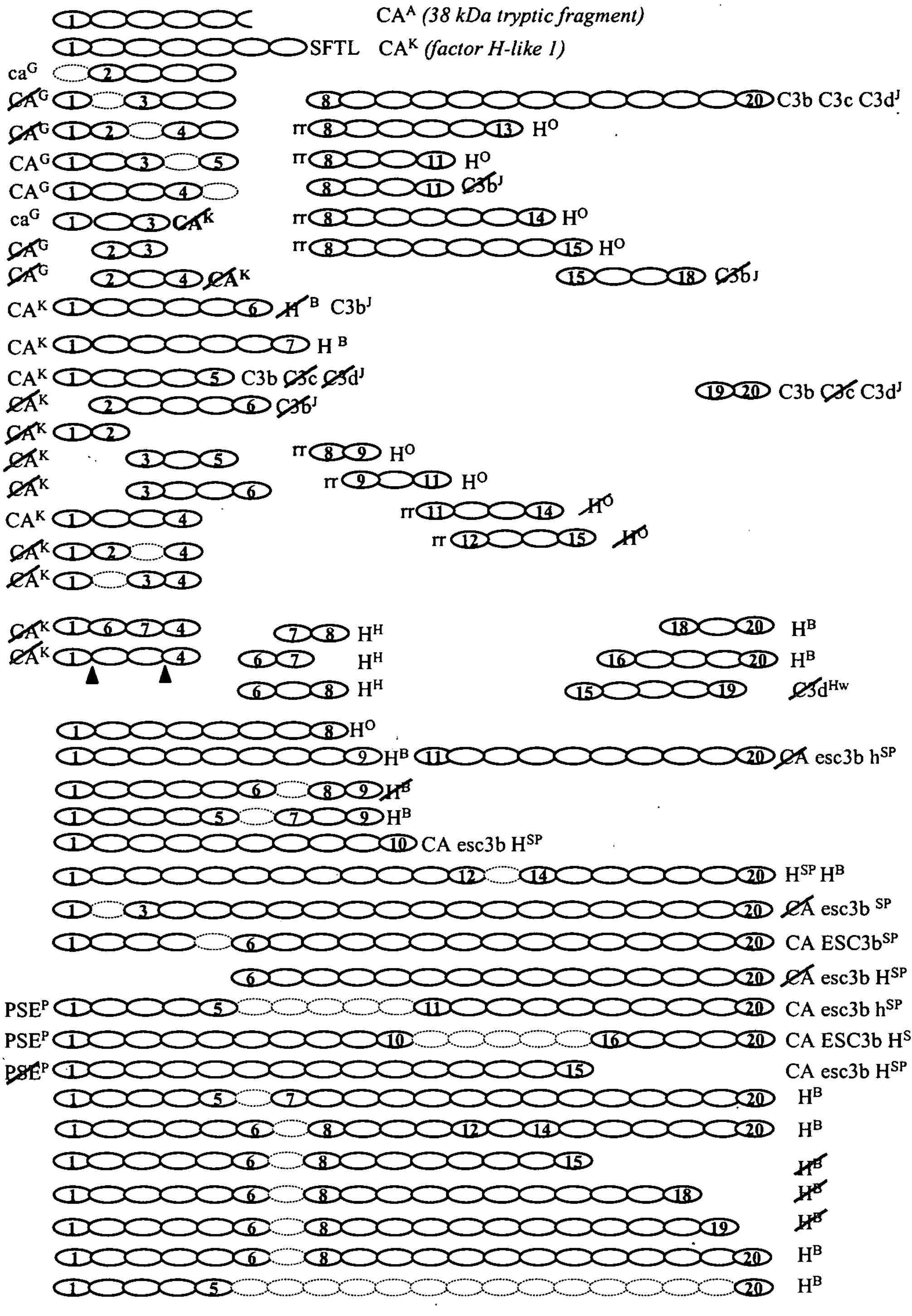

1 CA^A *(38 kDa tryptic fragment)*
1 SFTL CA^K *(factor H-like 1)*
ca^G 2
~~CA~~^{G} 1 3
8 20 C3b C3c C3d^J
~~CA~~^{G} 1 2 4
rr 8 13 H^O
CA^G 1 3 5
rr 8 11 H^O
CA^G 1 4
8 11 ~~C3b~~^{J}
ca^G 1 3 ~~CA~~^{K}
rr 8 14 H^O
~~CA~~^{G} 2 3
rr 8 15 H^O
~~CA~~^{G} 2 4 ~~CA~~^{K}
15 18 ~~C3b~~^{J}
CA^K 1 6 ~~H~~^{B} C3b^J
CA^K 1 7 H^B
CA^K 1 5 C3b ~~C3c~~ ~~C3d~~^{J}
19 20 C3b ~~C3c~~ C3d^J
~~CA~~^{K} 2 6 ~~C3b~~^{J}
~~CA~~^{K} 1 2
~~CA~~^{K} 3 5
rr 8 9 H^O
~~CA~~^{K} 3 6
rr 9 11 H^O
CA^K 1 4
rr 11 14 ~~H~~^{O}
~~CA~~^{K} 1 2 4
rr 12 15 ~~H~~^{O}
~~CA~~^{K} 1 3 4
~~CA~~^{K} 1 6 7 4
7 8 H^H
18 20 H^B
~~CA~~^{K} 1 4
6 7 H^H
16 20 H^B
6 8 H^H
15 19 ~~C3d~~^{Hw}
1 8 H^O
1 9 H^B 11 20 ~~CA~~ esc3b h^{SP}
1 6 8 9 ~~H~~^{B}
1 5 7 9 H^B
1 10 CA esc3b H^{SP}
1 12 14 20 H^{SP} H^B
1 3 20 ~~CA~~ esc3b ^{SP}
1 6 20 CA ESC3b^{SP}
6 20 ~~CA~~ esc3b H^{SP}
PSE^P 1 5 11 20 CA esc3b h^{SP}
PSE^P 1 10 16 20 CA ESC3b H^{SP}
~~PSE~~^{P} 1 15 CA esc3b H^{SP}
1 5 7 20 H^B
1 6 8 12 14 20 H^B
1 6 8 15 ~~H~~^{B}
1 6 8 18 ~~H~~^{B}
1 6 8 19 ~~H~~^{B}
1 6 8 20 H^B
1 5 20 H^B

Figure 1.11 Summary of module-deletion/truncations of FH. CCP modules are shown as ovals within a cartoon type representation of each deletion/truncation mutant. The black triangles indicate non-native linker lengths in one mutant. "CA" written alongside indicates it has cofactor activity; similarly "H", "C3b", "C3c" etc. indicate the mutant protein has binding affinity for heparin-affinity resin, C3b, C3c etc. (ESC3b signifies sheep erythrocyte-bound C3b, as opposed to fluid-phase or chemically immobilised C3b); "PSE": protects sheep erythrocytes from complement-mediated lysis. A strike-through means a particular activity was investigated

but found not to be measurable; lower-case letters signify substantial reduction in measured activity; SFTL = the four residues specific to the C-terminus of FH-like 1; "err" = non-native N-terminal sequence containing two Arg residues. Superscripts refer to cited works as follows: A = Alsenz et al. (55) (56); G = Gordon et al. (58); K = Kuhn (59); J = Jokiranta et al. (60); O = Ormsby et al. (61); B = Blackmore et al. (62) (63); H = Herbert et al. (1); HW = Hellwage et al. (64); SP = Sharma and Pangburn (65); and P = Pangburn (66).

Further C3b-binding sites were identified using immuno-affinity-purified module-deletion FH mutants from a baculovirus expression vector (62). Constructs lacked (Δ) modules: 2; 5; 1-5; 6-10; 11-15; 16-20; 1-10 or 11-20. All these deletion mutants exhibited C3b-cofactor activity except FHΔ2, FHΔ1-5, and FHΔ1-10. Crucially, FHΔ1-5 (and FHΔ2) nonetheless retained some binding affinity for cell-surface (sheep erythrocyte)-bound C3b (csbC3b), demonstrating that FH CCPs other than modules 1-4 bind C3b. Deletions of CCPs 16-20 decimated affinity for csbC3b, thus implicating a C-terminal region of FH as a second C3b-binding site. A third C3b-binding site was suggested because FHΔ6-10 exhibited a decreased affinity for csbC3b, similar to that of FHΔ1-5.

An antibody (131X) specific for CCPs 8-15 weakened interactions of full-length FH with csbC3b (64), suggesting a third C3b-binding region lies in these CCPs. Subsequently Jokiranta *et al.* (60) investigated binding of purified FH constructs, cloned in a baculovirus system, to C3b, and its fragments C3c and C3d, attached to Biacore "CM5" chips. In these surface plasmon resonance (SPR) experiments both FH-1-6 and FH-19-20 associated with immobilised C3b, confirming the presence of independent C3b-binding sites near both the N and C termini of FH. While FH-8-20 bound immobilised C3b, FH-8-11 and FH-15-18 did not (60). Hence this study could not confirm directly the existence of a third site, in the central segment of FH, able to independently bind C3b. On the other hand, no FH-12-14 construct was tested thus it remains possible that one or more of these three CCPs contribute to a putative third C3b-recognition region. According to the same study (60) (Figure 1.11) FH 1-6 bound to immobilised C3b but not to immobilised C3c or C3d (C3c and C3d are nonoverlapping proteolytic cleavage fragments of C3b; C3d corresponds to the thioester domain of C3b), while FH-19-20 recognised C3d in addition to C3b but not C3c (and the C-terminal C3b(C3d)-binding site was subsequently mapped to CCP 20 (66)). Intriguingly, FH-8-20 bound to both C3c and C3d suggesting (60) that the inferred, third, C3b-binding site is specific for C3c. A three-module, CCPs 10-12, C3b(C3c)-binding site would explain most of these results *i.e.* interference by the131X antibody, the loss of C3b-affinity by FH Δ6-10 (but not the full activity displayed by FH Δ11-15), and the results obtained with FH-8-11, FH-15-18 and FH-8-20; this hypothesis requires testing with the appropriate constructs. An alternative explanation consistent with the evidence is that measurable affinity for C3c requires simultaneous engagement of two sub-sites, one within CCP 10 and the other within modules 16-18. In summary, two distinct binding sites for C3b lie at the N (CCPs 1-4) and C termini (CCP 20) of FH; the latter is also able to bind C3d. Intervening modules participate in the binding process, but evidence for a distinct, third, C3b (C3c)-binding site remains circumstantial.

1.4.3 Binding sites for polyanions

Factor H binds to non-complement-activating surfaces through interactions with polyanions. This is fundamental to its ability to regulate complement on surfaces (37). In early studies, FH CCP 13 (which is highly basic) and CCP 14 were implicated using a photoaffinity-tagging heparin analogue (67). However, experiments on FHΔ13 and FHΔ11-15 (65) indicated that deletion of CCP 13 from FH results in only very slightly reduced ability to bind a heparin-agarose column (and negligible loss of binding to C3b-coated sheep erythrocytes). On the other hand FHΔ6-10 (65) showed significantly weaker heparin affinity implying a stronger GAG/sialic acid-binding site exists in the 6-10 region. A prominent role for CCP 7 in GAG binding was subsequently confirmed because FH-1-6 - one of a series of constructs generated in CHO cells (63) (Figure 1.11) - barely bound heparin while FH-1-7 was a good heparin-binder. Furthermore FH-1-6,8,9 (i.e. a module 7-deletion of FH 1-9) lost all affinity for heparin. Latterly, constructs of FH-6-8 were shown to bind GAGs and GAG analogues (1) (68).

Interestingly, FHΔ7 (and FHΔ7Δ13) bound heparin almost equally as well as FH, so module 7 is not the only one that binds polyanions (63). That FHΔ6–10, unlike FHΔ7 or FHΔ7Δ13, eluted from a heparin-affinity column at relatively low salt (65), suggests that modules 6, 8, 9 or 10 also participate in heparin binding. Some of these modules could contribute to the same heparin-binding site as CCP 7 or they could form a distinct, third site. A more recent study of constructs FH-8–9, FH-9–11 and FH-11–14 seems to support the notion of a third site centred on module 9; CCPs 8–9 bind most strongly, followed by CCPs 9–11, while CCPs 11–14 did not bind heparin (61). It is worth noting that in these experiments an artificial sequence containing two arginine residues (EFTWPSRPSRIGT) was apparently included at the N terminus (part of the cloning procedure) of FH-8–9 and FH-9–11; in combination with a native lysine residue prior to the Cys(I) of CCPs 8 or 9, this introduces a potential heparin-interacting artefact. Indeed, in a previous study FH-1–6, 8–9 (*i.e.* a construct containing both CCPs 8 and 9 but lacking CCP 7) had been shown not to bind heparin (62). In fact, the evidence suggests CCP 20 is the primary heparin-binding determinant in FHΔ7. The FHΔ7Δ20 construct eluted from a heparin-affinity column at low salt while FH-18–20 bound relatively tightly (62). Moreover, non-heparin-binding FH-1–5 was converted to heparin-binding FH-1–5,20 by inclusion of CCP 20 in the construct (62). Human FH from an individual with a mutation (FH-E1172Stop) resulting in a lack of module 20 bound weakly to a heparin-affinity column [49]. Highly purified, structurally characterized FH-19–20 (51) bound well to a heparin-agarose column. In summary, while GAG-binding sites in module 7 (with contributions from CCPs 6 and 8) and module 20 (with possible contributions from CCP 19) are well-established, current evidence for involvement of either CCPs 9 or 13 is inconclusive.

1.4.4 Dominant role of the C-terminus of Factor H in host recognition

The importance of the C-terminal heparin-binding site for self *versus* non-self discrimination was shown by experiments on FHΔ6–10, FHΔ11–15 and FH-1–15. Of these three constructs, only FH-1–15 could not protect sheep erythrocytes against lysis by human complement (Figure 1.11) (66). In a dramatic illustration of the role played by CCPs 19 and 20, Ferreira *et al.* (69) showed that purified, *P. pastoris*-produced, FH-19–20 competitively inhibited the action of FH on cell surfaces. This double-module construct overcame the protective effects of full-length FH and thereby promoted aggressive complement-mediated lysis of sheep erythrocytes. Further support for a dominant role of the FH C-terminus is provided by the ability of monoclonal CCP 20-specific antibodies (70) to block interactions of FH with endothelial cells. Thus this C-terminal polyanion and C3b-binding site is critical for the ability of FH to recognize and protect host cells bearing sialic acids and GAGs. Besides disrupting the protection of normally non-activating surfaces by full-length FH, FH-19–20 completely abolishes FH binding to immobilized C3b (but not to fluid-phase C3b) (69), despite the presence of the C3b binding site in CCPs 1–4. Moreover (70), CCP 20-specific antibodies blocked FH binding, in an enzyme-linked immunosorbent assay (ELISA), to C3b and C3d as well as to heparin (and to endothelial cells, as mentioned above). In order to reconcile these intriguing results with the multiple C3b- and GAG-binding sites identified by module-deletions and FH truncations, two models were proposed: (i) the C-terminus is unique among C3b-binding sites of FH in having a high affinity for cell surface-bound (csb) C3b (as opposed to fluid-phase C3b or C3b immobilized artificially on a chip or microtitre plate). The other C3b-binding site (in FH-1–4) has only poor affinity for C3b after the activated C3 fragment has become attached to a surface; this FH-1–4 site requires initial anchoring of FH via the C-terminus before it can engage, to a significant extent, with its binding site on csbC3b. Thus FH-19–20 competes with the sole csbC3b-binding site in full-length FH for binding to csbC3b. Because FH-19–20 has no complement regulatory region associated with it, the csbC3b to which it is bound is not destroyed by factor I. Similarly, if the csbC3b-binding site of FH is blocked by an antibody then FH will not be able to bind to csbC3b. In the fluid phase, binding of FH-19–20 to C3b is functionally irrelevant as the FH N-terminal modules bind well elsewhere on fluid-phase C3b and this latter interaction is sufficient for co-factor activity; hence - in agreement with experimental evidence – FH-19-20 does not inhibit fluid-phase co-factor activity. The C-terminus probably recognizes a composite site consisting of both GAGs and C3b; note that the C-terminal C3b-binding site is the only one that also binds GAGs. (ii) According to an alternative, or supplementary, model, the other FH C3b-binding site (in CCPs 1–4) is cryptic, only becoming available following occupation of the C-terminal site by C3b. The C-terminal site binds initially to the thioester domain of C3b inducing a conformational change within FH. For example, the binding site in CCPs 1–4 could be occluded initially by interactions with other CCPs in a compact conformation of FH. This notion of proximity between N- and C-terminal modules tallies with SPR experiments showing that FH-1–7 binds full-length FH on a Biacore chip [51] and with low-resolution structural studies discussed earlier. A problem with this notion of a cryptic site is that it predicts that fluid-phase C3b, in the presence of full-length FH and an excess of FH-19–20, will not be cleaved by factor I; this is because the FH-19–20 site on C3b would not be available for binding by FH. In fact, FH-19–20 does not inhibit the co-factor activity of FH in the fluid phase. It therefore remains necessary to invoke structural or accessibility differences between fluid-phase and csbC3b, as in model (i).

1.4.5 Other binding partners

As well as having binding sites for C3 fragments, polyanions, and factor I, FH reportedly binds to C-reactive protein (CRP) (71), fibromodulin (72) and adrenomedullin (73). Finally, it is bound by proteins borne on a huge number of microorganisms that surface-sequester FH to evade complement (74-76).

1.4.6 Factor H protein family

The Factor H protein family comprises seven multidomain proteins: Factor H (FH), Factor H-like protein 1 (FHL-1 or reconectin), and five Factor H-related proteins (FHR-1, -2, -3, -4, -5) (74). All these proteins are built exclusively from differing numbers of CCP modules. While FHL-1 – a splice variant of FH gene - is identical to the first seven N-terminal CCP modules of FH (with a C-terminal extension of four amino acid residues), the five FHR proteins are built from between four and nine CCP modules, all of which contain two modules highly homologous with FH-19-20 and other domains showing sequence homology to FH CCPs 6-7. Each of the FHR proteins are encoded by distinct genes located in the regulators of complement activation gene cluster (74). Since FHL-1 contains the four N-terminal regulatory domains of FH and bears a polyanion binding site in CCPs 6-7, it is not surprising that FHL-1 displays similar complement regulatory activities to FH in some respects. A unique feature of FHL-1 is its ability to act as an adhesion protein, an activity that is mediated by the RGD motif in CCP 4. Plasma levels of FHL-1 are 10 to 50-times lower than those of the main alternative pathway regulator FH (0.5 mg/ml). FHR proteins have been shown to bind to polyanions and C3b (74). Wide spread lack of DAA or CA among FHR proteins is a consequence of lacking sequence homology of any FHR protein with the N-terminal regulatory domain of FH. The exception is FHR-5 which exhibits weak CA and DAA (compared to FH) (77). FHR3 and FHR4B are not associated with CA or DAA but were shown to enhance these functional activities of FH (78). The exact role of FHR proteins is still under investigation.

1.4.7 Atypical haemolytic uraemic syndrome

Further support for a key biological role of the FH C-terminus (see 1.4.4 Dominant role of the C-terminus of Factor H in host recognition) in complement regulation *in vivo* derives from studies of atypical haemolytic uraemic syndrome (aHUS) (79). The majority of aHUS-linked FH mutations occur towards the C-terminus, with CCP 20 being a hotspot (80). Strikingly, a mouse model of aHUS was generated in FH knock-out mice (that develop a different renal pathology, dense deposit disease (81)) by transgenic expression of FH-1–15 (82). It was hypothesized that a predisposition to aHUS is linked directly to an inability of mutant forms of FH either to bind properly to C3b(C3d) or to recognize polyanionic markers on non-activating surfaces (or a diminishment of both these roles). This hypothesis has been tested by mutagenesis and structural studies. Table 1.4 lists laboratory-generated and naturally occurring FH mutants containing sequence changes within CCPs 19 and 20, all of which have been tested for function. Some were expressed recombinantly within the contexts of partial versions of FH such as FH-19–20 or FH-8–20, hence there is variation in the extent to which other binding sites within the protein can contribute to the functional outcome. Some changes coincide with aHUS-linked mutations while others were designed to identify residues participating in functional sites. All mutated proteins for which C3b(C3d)-binding data are reported exhibit decreased affinity; K1186A, which is not aHUS-linked, is the exception. Most mutants display loss or reduction in affinity for heparin-affinity resin and for human umbilical vein endothelial cells (exceptions are R1182A, K1186A, S1191L and V1197A). In fact, all the aHUS-linked mutants tested for function exhibited a deficiency in binding to either C3b(C3d) or GAGs, and in several cases to both. Pathophysiological insights based solely on the data in Table 1.4 are limited for several reasons. (i) No distinction was made between perturbation of a specific binding site and widespread structural disruption of the module. (ii) Contradictory results were obtained, *e.g.* W1183L in the context of FH-19–20 dimerised [as judged by mobility on sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)] and the dimer binds tightly to a heparin-affinity column; yet W1183L in the context of FH-8–20 binds more weakly to heparin-affinity resin than the native-sequence 8–20 construct. (iii) Glycosaminoglycans are diverse; their levels and patterns of sulphation, for example, vary between tissue-type and over the course of development and ageing (83). This raises a critical issue that has repercussions for the organ-limited nature of particular complement-associated diseases: can the ability of FH to protect host surfaces be attributed to specific GAG-recognition processes (84)? Unfortunately, heparin-affinity chromatography is a crude probe of the capacity to recognize specific surface polyanions.

Mutation	**Source of protein**	**aHUS link?**	**CA?**	**Binding to:**		
				C3b(C3d)	**GAG**	**HUVEC**
W1157R	In context of FH-8-20 (Baculovirus) (Jz 2006)	Yes	N/R	<< to C3b/C3d in CPA; none by SPR;	Slight reduction in heparin binding	Slightly reduced
E1172-stop	Purified from heterozygous patient (M 2003; J 2005)	Yes	N/R	<< to C3b by SPR	Slightly weaker than WT	No
R1182A	In context of FH-19-20 (*Pichia pastoris*) (J 2006)	Resembles R1182S	N/R	< to C3d in CPA; by SPR, sl. < C3b, < C3d	Binds to heparin-affinity column	N/R
W1183L	Ex heterozygous patient (& ex COS cells) (S-C 2002,4)	Yes	Yes *	< to C3b in CPA	N/R	N/R
W1183L	In context of FH-8-20 (Baculovirus) (Jz 2006)	Yes	N/R	< to C3b & C3d (SPR); < to C3b/C3d in CPA	Binds weakly to heparin column	Weak
W1183L	In context of FH-19-20 (*P. pastoris*); note - dimer on SDS-PAGE? (J 2006)	Yes	N/R	< to C3d in CPA < to C3d; anomalous result for C3b by SPR	Dimer binds more tightly than native FH-19-20?	N/R
K1186A	In context of FH-19-20 (*Pichia pastoris*) (J 2006)	No	N/R	Full binding to C3b in CPA; SPR N/R	Similar heparin affinity to FH-19-20	N/R
K1188A	In context of FH-19-20 (*Pichia pastoris*) (J 2006)	No	N/R	< to C3d in CPA; < to C3b & C3d by SPR	Sl < binding than FH-19-20 to hep.	N/R
S1191L	Ex heterozygous patient (H2006)	Yes	*	N/R	Full binding to heparin column	N/R
S1191L, V1197A	Ex heterozyg. patient (& in FH-18-20 context) (H 2006)	Yes	*	< to C3b and C3d by SPR	Full binding to heparin column	N/R
V1197A	Ex heterozyg. and also ex hemizyg. patients (H	Yes	*	N/R	Full binding to heparin column	N/R

	2006)					
V1197A	Ex homozygous patient (& ex COS cells) (S-C 2002,4)	Yes	Yes *	< to C3b in CPA	N/R	N/R
V1197A	In context of FH-8-20 (Baculovirus) (Jz 2006)	Yes	N/R	< to C3b/C3d-coated plates, and by SPR	Binds weakly to heparin column	Weak
E1198A	In context of FH-19-20 (*Pichia pastoris*) (J 2006)	Yes	N/R	< to C3d in CPA. SPR: sl < to C3b, spurious C3d	> to hep. column cf FH-19-20.	N/R
E1198K	Ex heterozygous patient (V-S 2006)	Yes	*	N/R	N/R	Weak
R1203E, R1206E, R1210S, K1230S, R1231A	In context of FH-15-20 (*P. pastoris*)(HW 2005)	R1210S resembles R1210C	N/R	Significantly < to C3d & C3b (by SPR)	No binding to heparin column	No
R1203E, R1206E, R1210S	In context of FH-15-20 (*P. pastoris*) (J 2005)	See above	N/R	N/R	No binding to heparin column	No
R1210C	Ex heterozyg. patient (di-S with other proteins) (& ex COS cells) (S-C 2002,4)	Yes	Yes *	< to C3b in CPA	N/R	N/R
R1210C	Ex heterozyg. patient & in context of FH-8-20 (Baculovirus) (M 2003; Jz 2006)	Yes	N/R	< to C3b/C3d-coated plates; < to C3d by SPR	Binds weakly to heparin column	Weak
R1215G	In context of FH-8-20 (Baculovirus) (M 2003)	Yes	N/R	Significantly < binding to C3d by SPR	Binds weakly to heparin column	Weak
P1226S	In context of FH-8-20 (Baculovirus) (Jz 2006)	Yes	N/R	Not to C3b/C3d-coated plates; None, by SPR	Binds weakly to heparin column	N/R
K1230S, R1231A	In context of FH-15-20 (*P. pastoris*) (J 2005)	No	N/R	N/R	Binds hep. column equal to FH-15-20	Weak

Table 1.4 Mutations in complement control proteins (CCPs) 19 and 20 of FH. aHUS: atypical haemolytic uraemic syndrome; GAG, glycosaminoglycan; HUVEC, human umbilical vein endothelial cell; NR, not reported; CPA, coated plate assay; SDS-PAGE: sodium dodecyl sulphate-polyacrylamide gel electrophoresis; SPR, surface plasmon resonance; sl, slight; hep, heparin; ex, from; <, reduced binding; CA, co-factor activity; *reduced ability to protect sheep erythrocytes from complement-mediated haemolysis. Superscripts refer to cited work as follows: Jz 2006, Jozsci *et al.* (85); M2003,Manuelian *et al.* (86); J 2005, Jokiranta *et al.* 2005 (87); J 2006, Jokiranta *et al.* 2006 (52); S-C 2002, 2004, (88) (89); H 2006, Heinen *et al.* (90); V-S 2006, Vaziri-Sani *et al.* (91).

Thus, unfortunately, it is difficult to test directly the hypothesis that mutations in FH predispose to aHUS if they disrupt regions of FH involved in recognition of specific GAG sulphation patterns in the glomerulus (for example). In an attempt to deal with some of these issues, the 3D structural basis of GAG recognition by CCPs 19 and 20 was investigated using nuclear magnetic resonance (NMR) (51) and crystallography (52). The two modules are organized in the linear, end-to-end arrangement observed in other structures of CCP-module pairs. A model GAG compound - pure, fully sulphated tetrasaccharide (dp4), enzymatically cleaved from heparin - was titrated into the FH-19-20 sample and the NMR frequencies, or chemical shifts, of protein backbone atoms were monitored for perturbations. The presence of bound dp4 induces changes in the magnetic field experienced by nearby nuclei and thus amino acid residues in or near the binding site will resonate with slightly different chemical shifts. A surface patch of FH-19-20 residues was thus implicated in binding to dp4. Strikingly, these coincided well with aHUS-linked mutations (51). Thus, these experiments support a disease model in which dysfunctional FH fails to protect fully the GAG-rich layers of the glomerular basement membrane from complement activation. In another hypothesis C3d (and C3b) binding might be disrupted by aHUS-linked mutations, as suggested by Jokiranta *et al.* (52). One possible explanation for these apparently conflicting results is that FH-19–20 has GAG- and C3d(C3b)-binding sites that are sufficiently close to interfere or co-operate with one another depending on the nature of the ligand tested (*e.g.* heparin *versus* a physiological GAG and C3b *versus* C3d) and the order of binding events. Thus heparin inhibits binding of C3d to FH-15–20 while C3d enhances binding of native-sequence FH-15–20 to heparin (87). These observations are not inconsistent with the intuitively feasible hypothesis that a tertiary complex containing GAGs, C3b and FH forms during the process of complement regulation on self-surfaces. A caveat to interpretation of these studies is that electrostatic 'steering interactions' (as probed here by mutating Arg and Lys residues to uncharged residues, or potentially neutralizing their side-chains by addition of a polyanion such as heparin) may represent just one step in the multi-step process whereby C3d interacts with CCPs 19–20. Electrostatic steering (92) could enhance the number of productive encounters between a pair of protein molecules by influencing their orientations as they approach one another, but the eventual complex could be stabilized by other interactions involving an entirely different set of side-chains. Another factor that complicates interpretation is the putative oligomerization of FH that could be important and may occur via the C-terminus (52). Thus despite much recent progress, in the absence of detailed thermodynamic studies of binding and 3D structures of binary and tertiary complexes, the picture remains obscured.

1.4.8 Dense deposit disease

Membranoproliferative glomerulonephritis type II (MPGN II), also known as dense deposit disease (DDD) (93, 94), is characterised by C3 deposition in the absence of immunoglobulin (95). In contrast, MPGN types I and III are associated with immune complex localisation in glomeruli (94) and account for the majority (~80%) of all MPGN cases. Deposition of C3 along the glomerular capillary and in mesangial regions is characteristic for MPGN II. In this disease, severely under-controlled systematic alternative pathway activation leads to dense deposits within the glomerular basement membrane (GBM). Within 10 years from diagnosis half of the patients develops end-stage renal failure. In many individuals with MPGN II similar deposits occur in the choriocapillaris-Bruch's membrane-retinal pigment epithelial interface predisposing for visual impairments (see also below 1.4.9 Age-related macular degeneration). Immunohistological studies identified C3c as the major constituent of dense deposit in patience with MPGN II (94). In a subtype of MPGN II, rapidly progressive MPGN II, dense deposits were also shown to react with anti-C3d antibodies.

The most common trigger of MPGN II is autoantibodies (C3 nephritic factors) that stabilise C3 convertases of the alternative pathway and shield them from FH-mediated inactivation. Consistent with mouse and pig models of the disease (in which FH is entirely lacking), mutations in FH of human patients have been associated with a predisposition to MPGN II (94). Many of these mutations (*e.g.* substitution of a consensus cysteine disrupting the formation of correct disulfide patterns) result in retention of FH in the endoplasmic reticulum and prevention of extracellular FH secretion. The resultant significantly reduced FH plasma levels lead to severely under-controlled alternative pathway activity, and the consequences are especially apparent at the GBM. The GBM is a layer of functionally specialized extracellular matrix rich in type IV collagen, laminin, other proteoglycans and glycoproteins (94), with an overall high content of glycosaminoglycans. The GBM is exposed to complement at fenestrations between the endothelial cells lining glomerular capillaries. The lack of membrane-bound RCAs on the GBM is consistent with necessity to recruit circulating FH as a means of avoiding amplification of C3b at the GBM surface, and with the potentially severe consequences for this structure of a lack of fully functional FH.

1.4.9 Age-related macular degeneration

Patients suffering from MGPN II often develop the soft ocular drusen regarded as a hallmark of early age-related macular degeneration (96) (AMD), which is a leading cause of geriatric vision-loss. A recently discovered link between the Y/H 402 polymorphism in CCP 7 of FH and both dense deposit disease (97) and AMD (98) hints at the physiological importance of this second (after CCP 20) polyanion-binding site in FH. The at-risk allele has received most attention; it is present in ~35% of individuals of European descent. Homozygous individuals have a 6-fold increased risk of developing AMD, whereas heterozygotes are only 2.5 times more susceptible (99). The at-risk sequence variation is also present in FH-like 1, where module 7 represents the sole polyanion-binding site. This GAG-binding module has been reported additionally to participate in binding sites for CRP (100), fibromodulin (101), DNA (102) and various pathogen-borne proteins (100, 102, 103).

Structural and chemical shift perturbation studies of CCP 7 - similar to those carried out on FH-19-20 – revealed that the H/Y 402 side-chain is positioned in such a way as to contribute specificity to a GAG-binding groove (1). Working with the single module, Y402 CCP 7 bound significantly more strongly than H402 CCP 7 both to a heparin-affinity column and to dp4. This appears to support the case for a causal link between the polymorphism and a mechanism for AMD involving insufficient complement regulation in the ageing choroid. When examined in the context of the triple-module FH-6–8, however, the two variants bound equally well to some GAGs (1) but differently to others (68), with either variant binding more tightly depending upon the GAG tested. Both Y402 and H402 versions of full-length FH bound equally well to a series of fully sulphated heparin fragments and (as might have been predicted from the aforementioned results obtained for FHΔ7) to a heparin-affinity column (1, 101, 104, 105). In all probability, what these results illustrate is that the FH–GAG interaction is both dual-site (*i.e.* involves two physiologically relevant GAG-binding sites on FH) and tissue-specific. The results obtained with isolated CCP 7 in complex with a chemically defined heparan sulphate analogue, together with the data for interaction of FH-6–8 with a range of heterogeneous GAGs, indicate strongly that the disease-linked polymorphism is tweaking the GAG-(self)-recognition capabilities of FH in a subtle fashion not apparent in the test tube with full-length FH and the 'blunt instrument' of heparin. Such a conclusion is supported by the detailed picture of protein–sugar interactions provided by the crystal structure of FH-6–8 in complex with sucrose octasulphate (53). A subtle difference between the two allotypic variants is consistent with the H402 (at-risk) variant of FH – which is present in 35% of Western populations - functioning adequately until at least old age. Indeed, the H402 variant is less tightly sequestered than the Y402 variant by the M6 protein of *Streptococcus pyogenes* and might confer an evolutionary advantage in this respect (105). Consistent with a 'GAG hypothesis' for a causal link with AMD there is also evidence for differential binding of the Y/H 402 variants to cell surfaces. Flow cytometry and confocal laser scanning microscopy revealed a slightly lower binding of the H402 variant to retinal pigment epithelial cells and to endothelial cells, and indeed (in the context of FH-like 1) the H402 variant has reduced co-factor activity at the cell surface even though there is no difference in fluid-phase cofactor activity between the two variants (106). In another flow cytometry study the H402 variant exhibited higher binding to necrotic Jurkat T cells, which may reflect a measured difference in DNA-affinity between the variants (with H402 FH-6–8 binding to DNA more tightly than Y 402 FH-6–8) (101). Controversy surrounds the effects of the polymorphism on the affinity of FH for CRP as measured by SPR or by ELISA. Five reports concur that the H402 allotypic variant (within the context of full-length FH, FH-like 1, FH-5–7 or FH-6–8) is the weaker CRP binder (1, 101, 104-106). Previous work showed that CRP is more abundant in the eyes of AMD patients with the H402 variant of FH than in those with the Y402 variant (107) and that CRP is present in drusen (101, 107). It has been suggested that FH recognizes CRP borne on the membranes of apoptotic cells and acts to ensure that apoptotic cell clearance proceeds in a non-inflammatory setting (21). Overlap has been reported for CRP- and GAG-binding sites on CCP 7 (100), suggesting that the relative affinities for these two ligands might be critical for FH function at self-surfaces in situations where both ligands are encountered; if the Y/H 402 side-chain contributes to both sites it could be a key player in this respect. So a 'GAG hypothesis' and a 'CRP hypothesis' need not be mutually exclusive. However, a study by Hakobyan *et al.* (108) casts doubt on the relevance of the FH–CRP interaction. In the hands of these authors, FH interacts with CRP only under circumstances where the CRP pentamer is disrupted through Ca^{2+} ion-removal, a situation that would never prevail in physiological circumstances.

1.4.10 Summary of literature survey

Taking all these data together, a mechanism for FH is emerging. The two terminal regions of CFH (specifically, modules 1–7 and modules 19–20) contain all the proven discrete binding sites and are also the sites of most disease-linked sequence variations. Towards the centre of CFH (CCPs 12–14) module 13 - the smallest CCP module in FH with respect to the number of amino acids - is joined with its neighbouring modules by long linkers (Figure 1.12) allowing CFH to kink so that the two terminal ends of FH are brought into proximity. The C-terminal modules are the only ones with the potential to bind to a composite site consisting of C3b and the polyanions on the self-surface to which C3b is attached, and hence are dominant in distinguishing self-surfaces from non-self (complement activating) ones. The N-terminal three or four CCPs, like other similar blocks of CCPs in other RCAs (*e.g.* CD46, CD55, CD35) are able to perform the task of disrupting the surface-associated C3/C5 convertases once the CFH is anchored in place by its C-terminus.

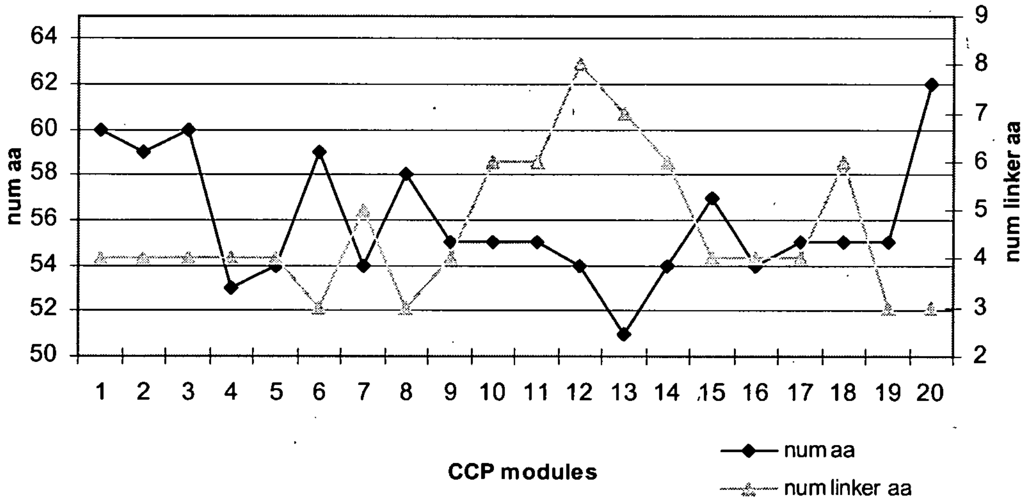


Figure 1.12 CCP module and linker size in FH. The five linkers of CCPs10-15 consist of six, seven or eight residues. The largest numbers of linker residues occur around CCP 13 while the size of CCP 13 is the smallest in FH (and among all known RCAs).

To position correctly the N-terminal modules in the GAG-convertase-CFH complex, a further interaction of CFH with GAGs is mediated by the specific recognition capabilities of

CCP 7 - indeed, this module could act as a 'proof-reader' to make it more difficult for bacteria to emulate the chemistry at self-surfaces.

1.5 Project aims

Our working hypothesis suggests the architecture of the middle portion of FH imposes or stabilizes a conformation of the molecule in which the terminal regions engage simultaneously with its key ligands (C3 fragments and polyanionic carbohydrates). To test our hypothesis, it is necessary both to re-examine the distribution of binding sites throughout the protein and to elucidate its structure. Localisation of binding sites will be achieved by expressing selected regions of FH and measuring their affinity for protein and carbohydrate ligands. With regard to structure determination, the large potentially flexible nature of FH rules out a direct approach based on NMR or X-ray diffraction. To circumvent this difficulty, a "dissect-and-rebuild" strategy based on studies of short fragments of FH is planned. Both functional and structural aspects of this study therefore depend upon recombinant expression of multiple-CCP module fragments of FH at micro-to milligram yield. Based on previous experience in the Barlow lab, *P. pastoris* was the vector of choice for this exercise.

The first aim is to undertake functional characterisation of FH ligand-binding regions where literature reports remain circumstantial or controversial; this includes investigation into the ability of CCP9 and CCP13 to act as binding partners for polyanions and CCPs 12-14 to act as a binding partner for C3b and C3c.

The second aim is to conduct quantitative studies on the well-established binding sites at either terminus of FH and the region of CCP6-8; the objective here is to obtain a more accurate understanding of the binding of FH to C3b and its fragmentation products than has been available hitherto

The third aim is to express multiple module constructs encompassing the centre of FH to test the notion that CCP modules 10-15 include a hinge (or rigid bend) that facilitates spatial proximity of the C- and N-termini. The top priority will be the double-module construct FH-12-13, in which the smallest module of FH, CCP 13 (51 amino acid residues) is connected to CCP 12 by the longest linker (eight amino acid residues). This construct will be expressed in ^{15}N and ^{13}C-enriched media to enable determination of a high-resolution solution structure using NMR. In addition, longer constructs embracing flanking modules (*i.e.* 10,11 and 14,15) will be subjected to analytical ultracentrifugation to provide an architectural context for the high resolution structural studies.

CHAPTER 2

MATERIAL AND METHODS

2.1 DNA amplification, cloning, protein production and purification

2.1.1 Polymerase chain reaction (PCR)[3]

The PCR was used to amplify coding sequences for full-length FH from a human cDNA library, and for amplifying the coding sequences for the various FH constructs. It was also used for screening of colonies, QuickChange® site directed mutagenesis (Stratagene®, CA, USA) and sequencing reactions.

All priming sequences[4] were purchased from Sigma®-Genosys and are listed in Table 2.1. The primers were shipped in a desalted and de-protected form. Exceptionally, some of the longer primers were ordered in an HPLC-purified form.

Oligonucleotide name	**DNA sequence**	**restriction site**
FH-1 F	CTATTTGTG*CTGCAG*AAGATTGCAATGAACTTCC	Pst1
FH-6 F	AA*CTGCAG*GACCTTGTGATTATCC	Pst1
FH-7 F	TCAAA*CTGCAG*GACTCAGAAAATGTTATTTTC	Pst1
FH-8 F	AA*CTGCAG*GAACATGTTCCAAATC	Pst1
Orm-FH-8 F	GTCGGTACCAAAACATGTTCCAAATCAAG	Kpn1
FH-10 F	AA*CTGCAG*GAGAATGCGAACTTCC	Pst1
FH-11 F	AA*CTGCAG*GATCATGTGGTCCACCTCC	Pst1
FH-12 F	AA*CTGCAG*GAACCTGTGGAGATATACC	Pst1
FH-13 F	AA*CTGCAG*GAAAGTGCAAATCATC	Pst1
FH-14 F	AA*CTGCAG*GATTATGCCCACCTCCACC	Pst1
a-FH-7 R	GC*TCTAGA*CTAGATGCATCTGGGAGTAGG	XbaI
b-FH-7 R	GC*TCTAGA*CTATTTGACACGGATGCATC	Xba1
FH-8 R	GC*TCTAGA*CTAAGATTTAATGCACGTG	Xba1
FH-9 R	GC*TCTAGA*CTAATAACATATGGGTAAATC	Xba1
Orm-FH-9 R	GG*TCTAGA*CTATTCATAACATATGGGTAAATCA	Xba1
FH-11 R	GC*TCTAGA*CTAAATACACACTGGTAAAG	Xba1
FH-12 R	GC*TCTAGA*CTACACACACTGGGGAAGTTGG	Xba1
FH-13 R	GC*TCTAGA*CTATGAGCAGTTCACTTCTGG	Xba1
FH-14 R	GC*TCTAGA*CTAAACACAGAGTGGTATTGAC	Xba1
FH-15 R	GC*TCTAGA*CTATTCACACTGAGGTGG	Xba1

---

[3] For composition of buffers, media and vector maps see appendicies

[4] Many of the primers listed here were designed by Claire Egan or Dr Andrew Herbert former members of the Barlow group.

FH-20 R	GCATGC*TCTAGA*CTATCTTTTGCACAAGTTGGATAC	XbaI
FH-7(H402Y) F	GGATATAATCAAAATTATGGAAGAAAGTTTG	n/a
FH-7(H402Y) R	AAACTTTCTTCCATAATTTTGATTATATCC	
FH-8(G446K) F	CACGTGGCCCAGCGCTCCGTCTGCTATCGGTACCAAAAC	n/a
FH-8(G446K) R	GTTTTGGTACCGATAGCAGACGGAGCGCTGGGCCACGTG	n/a
FH-12(N718Q) F	CAGTGGAATTTCAGTGCTCAGAATCATTTAC	n/a
FH-12(N718Q) R	CTGAGCACTGAAATTCCACTGAATC	n/a
FH-13(N802Q) F	GGATCCAGAAGTGCAGTGCTCAATGG	n/a
FH-13(N802Q) R	GTGCCATTGAGCACTGCACTTCTGG	n/a
FH-14(N822Q) F	CCCAATTCTCACCAGATGACAACC	n/a
FH-14(N822Q) R	GGTTGTCATCTGGTGAGAATTGGG	n/a
FH-13(Q802N) F	GGATCCAGAAGTGAACTGCTCAATGG	n/a
FH-13(Q802N) R	CCATTGAGCAGTTCACTTCTGGATCC	n/a
FH-14(Q822N) F	CCCAATTCTCACAATATGACAACC	n/a
FH-14(Q822N) R	GGTTGTCATATTGTGAGAATTGGG	n/a
alpha-Factor F	GGGGATTTCGATGTTGCTGTTTTG	n/a
AOX1 R	CCGGTCTTCTCGTAAGTGCC	n/a
pUB/Bsd-TOPO F	GCAGCTTATAATGGTTACAAATAAAGCAATAGC	n/a
pUB/Bsd-TOPO R	GGTAACGCCAGGGTTTTCCC	n/a

Table 2.1 Primer sequences. Restriction-enzyme sites are shown in italics. Primers used with the QuickChange site-directed mutagenesis kit for the generation of constructs with "mutated" primary sequences are named in such a way as to indicate the amino acid substitution. F and R indicate the forward and reverse priming directions, respectively. The expression "(X446K)" indicates the substitution of a cloning artefact residue with wild-type amino acid Lysine at position 446. Orm-FH-8 F and Orm-FH-9 R stand for primers use in the study of Ormsby et al. (61)

The generation of coding sequences was generally carried out according to the scheme in Table 2.2 by mixing 1 μl of forward and reversed primers (10 μmol), template DNA (40 ng), 5 μl Herculase® reaction buffer (10x concentrate from Stratagene®), 2.5 μl dimethylsulfoxide (DMSO), 1 μl deoxynucleotde triphosphates (NTPs) (10 mM from (Roche®, Mannheim, Germany), 0.5 μl Herculase® Hotstart polymerase (5 U/μl) (Stratagene®) and adjusting the volume to 50 μl using molecular biology-grade H_2O.

Step	Temperature (°C)	Time (min)	Repetitions
Initialisation	95	1.0	1x
Denaturation	95	0.5	14x
Touch-down annealing steps	T_m	0.5	
	gradient -0.5°	R=3.0°/s	
Elongation	72	1.0 / kb DNA	
Denaturation	95	0.5	19 to 24x
Annealing	T_m -7	0.5	
Elongation	72	1.0 / kb DNA	
Final Elongation	72	5.0	-
Hold	4	hold	-

Table 2.2 PCR cycling parameters. The melting temperatures (T_ms) of the whole primers with subtraction to account for the contribution from non-complementary regions of the primers (*i.e.* restriction enzyme site or mutation substitutions) determine the annealing temperatures. In cases where the resulting T_ms of forward and reverse matching sequences were different, the annealing temperature was set to the lower primer T_m.

2.1.1.1 Generation of full length FH coding sequence from c-DNA library

The coding sequence for full-length FH was amplified from the human universal Quick-clone™ cDNA library (Clontech®, CA USA) and cloned into pCR®4Blunt-TOPO® vector then sub-cloned into pPICZalphaB vector). These two full-length FH clones were used as a template for generation of coding sequences for different FH module constructs.

2.1.1.2 Generation of wild-type coding sequences for different FH module constructs

For the generation of wild-type coding sequences the PCR volumes and cycling program, as described above, were used. In case where no, or only a little, PCR product could be obtained with these PCR parameters, the annealing temperature was lowered by up to 5 °C; alternatively a subsequent PCR reaction was performed using a small proportion of the first-round PCR product.

2.1.1.3 PCR screening of colonies

Screening by PCR was performed with Mastermix® (0.05 U/μl) from Promega UK Ltd. (Southampton, UK). From overnight plates, roughly ten single bacterial colonies were picked and transferred to a new plate as well as being transferred to a PCR tube containing 20 μl "EB buffer" (10 mM Tris·Cl, pH 8.5), which was subsequently exposed to 90-100 °C for 300 s to lyse the bacteria and solubilise plasmid DNA. A 1 μl aliquot of this extraction was mixed with 5 μl PCR-Mastermix solution and 2 μl each of forward and reversed primer (10 μM) The PCR products were analysed on an agarose gel. For PCR conditions see Table 2.3.

Step	Temperature (°C)	Time (min)	Repetitions
Initialisation	95	1.0	1x
Denaturation	95	0.5	30x
Annealing	50	0.5	
Elongation	60	1.0 / kb DNA	
Hold	4	hold	-

Table 2.3 Mastermix PCR cycling parametres

2.1.1.4 Sequencing reactions

All of the DNA sequencing was carried out by the automated sequencing service of the School of Biological Sciences (University of Edinburgh). The primers described in Table 2.4, or sequencing primers supplied with the TOPO® cloning kit (Invitrogen), were used. The PCR-based sequencing was carried out according to the table below using primer (3.2 pmol), template DNA (150-300 ng), ABI prism BigDye terminator mix V3.1 (4 μL), and adjusting the volume to 20 μL. The resultant PCR produce was submitted for sequencing.

Step	Temperature (°C)	Time (min)	Repetitions
Initiation	95	0.5	1x
Denaturation	96	0.5	24x

Annealing	50	0.5	
Elongation	60	4.0	
Hold	4	hold	-

Table 2.4 Sequencing cycle programme

2.1.1.5 *Sequencing of DNA inserts from P. pastoris*

Small fragments of *P.pastoris* colonies were picked and transferred to a PCR tube containing 20 μl EB buffer, which was subsequently exposed to 90-100 °C for 5 min to lyse the yeast and solubilise DNA. A 1 μl aliquote of this extraction was used to perform a Touch-down PCR. The resultant PCR products were analysed on a DNA agarose gel, ethanol precipitated and resuspended in molecular biology-grade H_2O. The resuspended DNA was used as a template for a sequencing reaction.

2.1.1.6 *QuickChange® site-directed mutagenesis*

The PCR-based sequencing reactions were carried out according to Table 2.5 (below) by mixing 1 μl of forward and reversed primer (10 μmol), template plasmid DNA (40 ng), 5 μl Pfu buffer (10x concentrate), 1 μl dNTPs (10mM), 1 μl Pfu Turbo hotstart (2.5 U/μl) (Invitrogen) and adjusting the volume to 50 μL using molecular biology-grade H_2O. The QuichChange reaction was mixed with *Dpn*1 (10 U/μl) (New England Biolabs) and incubated for between one and two hours at 37 °C to allow for digestion of the methylated template DNA plasmid prior to transformation into competent bacterial cells.

Step	**Temperature (°C)**	**Time (min)**	**Repetitions**
Initiation	95	0.5	1x
Denaturation	96	0.5	19x
Annealing	T_m -5	1.0	
Elongation	68	1.0 / kb DNA	
Final Elongation	72	5.0	-
Hold	4	hold	-

Table 2.5 PCR cycling parameters for QuickChange® site-directed mutagenesis

2.1.2 Horizontal Agarose Gel Electrophoresis for Nucleic Acids.

A batch of triethanolamine-agarose (1.0 to 1.3 % (w/v) agarose) was heated until the agarose had completely dissolved. It was cooled to approximately 45 °C before addition of ethidium bromide to a final concentration of approximately 0.5 µg/ml and then poured into a gel tank to set. Alternatively, the ethidium bromide-substitute, CYBR® Safe stain from Invitrogen (Invitrogen, CA, USA), was added in a similar procedure and according to the protocol supplied by Invitrogen. DNA samples in loading buffer were loaded into wells alongside DNA molecular-weight markers that covered the range 250 – 1500, or 500 - 10,000, base pairs (New England Biolabs, MA, USA). The gel was run at 100 V until clear resolution of the markers was obtained, then it was visualised with UV radiation using a transilluminator.

2.1.3 TOPO® cloning reaction

The cloning of the PCR-product into the pUB/Bsd® Topo® or pCR® 4blunt TOPO® vector was achieved as follows.

An aliquot containing 2 µl of PCR product was mixed with 0.5 µl of the provided dilute-salt solution, 0.5 µl sterile H_2O and 0.5 µl vector solution. The mix was centrifuged and was allowed to sit for 300-600 s at room temperature. Then a 3-µl aliquot of this mixture were employed to perform a reaction using the TOPO® cloning kit prior to transformation of One Shot® Top10 chemically competent *E. coli* cells. The transformed cells were spread on LB Miller agar plates ampicillin (100 to 200 µg/ml (w/v)) agar plates and grown overnight at 37 °C.

2.1.4 Restriction enzyme single and double digests

The reactions were routinely carried out with NEB® enzymes and matching reaction buffers as follows:

Appropriate NEB-buffer + BSA	2 µl

vector	14 μl
restriction endonuclease 1	1.0 μl (20 U/ μl)
(restriction endonuclease 2	1.0 μl (20 U/ μl) only in case of a double digest)
H_2O	added to obtain a volume of 20 μl

In the case of double digests, the reaction buffer was chosen to maximise the efficiency of both endonucleases following the New England Biolabs recommendations. The reaction components were mixed and incubated at 37 °C for 3 h or overnight. The restriction endonucleases were then deactivated at the temperature and time recommended by the manufacturer. The DNA fragments obtained were purified and concentrated by phenol-chloroform extraction followed by an ethanol precipitation or by gel extraction using QIAquick® spin gel extraction kit (Qiagen, Crawley, UK). The restriction digests were analysed on a 1% agarose electrophoresis gel to ensure that the reaction had been carried to completion.

2.1.5 DNA purification and Plasmid DNA extraction

The QIAprep®miniprep or QIAprep®maxiprep kit for plasmid extraction and the QIAquick® spin gel kit for extraction and purification of DNA from agarose gels were obtained from Qiagen (Crawley, UK).

From a bacterial culture that had been growing over night at 37 °C, an aliquot (5 ml for miniprep or 150 ml for maxiprep) was removed and spun down (5 min at 4000g). From the pellet, plasmid DNA was extracted using QIAprep miniprep kits following the protocols provided. For gel extraction of cut DNA fragments, the QIAquick® spin gel extraction kit was used. The eluted DNA was stored at -20 °C.

2.1.6 Ethanol precipitation

Based on the measured volume of the DNA sample, the salt concentration was adjusted by adding sodium acetate, pH 5.2 to a final concentration of 0.3 M, or by adding 5 M ammonium acetate to a final concentration of 2.0-2.5 M. Then between 2 and 2.5 volumes (calculated after salt addition) of cold 100% ethanol was added. The mixture was vortexed and placed on ice or at -20 °C for 0.5-1 h. Subsequently it was centrifuged (Eppendorf Microfuge®) for 10-15 min at maximal speed. Then the supernatant was carefully removed and 1.0 ml of 70% ethanol was added. After a brief further centrifugation step the supernatant was carefully removed and the pellet was air dried. Finally the pellet was resuspended in an appropriate volume of TE-buffer or water.

2.1.7 Phenol-Chloroform extraction

Phenol-Chloroform extractions were carried out with Phenol:Chloroform:Isoamyl Alcohol mixture (in rates of 25:24:1 respectively, saturated with 10 mM Tris, pH 8.0, 1 mM EDTA) a under standard procedures. This mixture was purchased from Sigma-Aldrich®.

2.1.8 DNA quantification

DNA concentrations were determined by spectroscopy (Eppendorf BioSpectrometer, Eppendorf, Hamburg, Germany) at a wavelength of 260 nm. The purity of DNA samples were estimated by the absorbance ratios of A_{260nm} / A_{280nm} and A_{260nm} / A_{320nm}.

2.1.9 Ligation reactions

Roughly, 50 ng of cut vector pPICZαB (Invitrogen) combined with a three-fold molar excess of insert were mixed and the volume adjusted with dH_2O to 10 μl. Then 10 μl of 2x Quick Ligation Buffer (provided with the ligase) and 1 μl of Quick T4 DNA Ligase (400 U/ μl) (both from NEB) were added and mixed thoroughly. After a brief centrifugation step, the reaction mixture was incubated at room temperature (25°C) for five minutes prior to Top 10 transformation. Transformed cells were spread onto LB Lennox Zeocin™ (25 μg/ml) plates and grown overnight at 37 °C. This step utilises the Zeocin™-resistance gene present in the *P. pastoris* expression vector pPICZαB.

2.1.10 Transformations of plasmids into *E. coli* competent cells and culture growth

A microcentrifuge tube containing the frozen stock of cells were thawed on ice; the cell suspension was then swirled and incubated on ice for a further ten minutes. Plasmid DNA (1-2 μl of concentrations between 50 to 400 μg/μl) was added, and the two solutions were mixed together by tapping the tube. After incubation on ice for 30 minutes, the tube was placed in a water bath at 42 °C for precisely 45 s and then rapidly returned to the ice for two minutes. Pre-warmed SOC medium (250 μL, 42 °C) was added and cells were incubated at 37 °C, on a shaking platform (200 rpm) for one hour, then plated onto pre-warmed LB Lennox agar plates for Zeocin™ (25 μg/ml final concentration) resistance selection, or LB agar plates for ampicillin (100-200 μg/ml final concentration) resistance selection, and incubated at 37 °C overnight.

After overnight growth on agar plates, single colonies were picked and used to inoculate 5 ml LB Lennox containing Zeocin™ or LB containing ampicillin (concentrations as above) followed by incubation on the shaking platform (225 rpm) 37 °C. These cultures were used for DNA analysis and manipulation. When larger culture volumes were required, 100 μL of the starter culture was placed into 50-250 ml of the appropriate media.

2.1.11 Transformation of *P. pastoris*

P. pastoris is a yeast expression system. As a eukaryotic vector, it combines the desired features of protein processing, protein folding, and posttranslational modification (109, 110). *P. pastoris* is known to exhibit high heterologous levels of protein expression. The genetic nomenclature and numerous techniques developed for *Saccharomyces cerevisiae* have been carried forward to *P. pastoris*. As *P. pastoris* is methylotrophic, it is able to use methanol as a carbon source in the absence of glucose. Methanol is oxidised to formaldehyde and hydrogen peroxide using molecular oxygen and alcohol oxidase (AOX). This reaction takes place in the peroxisome. Hydrogen peroxide is consumed via the action of a catalase. The formaldehyde, which is generated by AOX, partly leaves the peroxisome and is further oxidised to formate and carbon dioxide by cytoplasmic dehydrogenases, providing the organism with energy. The formaldehyde remaining within the peroxisome is used to produce cellular constituents. The genome of *P. pastoris* encodes for two alcohol oxidases, AOX1 and AOX2. The two enzymes are 97% identical to each other and share approximately the same specificity (111). Methanol metabolism is mainly carried out through the catalytic action of the AOX1 gene product, which promotes a more rapid catabolism of methanol relative to AOX2. *P. pastoris* growing on methanol as a carbon source expresses AOX1 levels of more than 30% of total soluble protein. Two mechanisms are involved in the regulation of the AOX1 gene. First, there is a repression/derepression mechanism, *e.g.* in the presence of glucose the AOX1 gene is repressed. Absence of glucose on its own is not sufficient to stimulate the AOX1 expression as, secondly, the presence of methanol is required as an induction signal prior to AOX1 gene transcription (112). The loss of its main methanol catabolic enzyme AOX1 results in *P. pastoris* strains (MUT^S: methanol utilization slow) that have to rely on their facility to catabolise methanol through the more slowly metabolising AOX2. There are therefore two different *P. pastoris* methanol metabolising phenotypes, MUT^S and MUT$^+$. The *P. pastoris* strain used throughout this study was KM71H (Invitrogen) of the MUT^S phenotype.

Heterologous protein expressed in *P. pastoris* can be intracellular or secreted. Secretion requires a signal sequence in order to direct the protein to the secretion pathway. As *P. pastoris* only secretes very low levels of native protein into the medium, and is able to grow in minimal medium, secretion can be an important first step in purification of the heterologous protein (113). The *P. pastoris*-based expression system offers superior post-translational modifications, as it does not cause hyperglycosylation. Of relevance to RCA proteins (containing two disulfides per CCP module) native disulfide bonds patterns are generally obtained in the secreted product. Once the heterologous protein sequence has been transformed into the expression cassette and subsequently integrated in the genome, it is under the control of the strong and highly-inducible AOX1-promotor. Therefore expression of the protein can be regulated through the feeding of methanol to cell cultures. Usually this system expresses foreign proteins in high yields, although yields are generally not as high as for alcohol oxidase itself (114). In conclusion *the P. pastoris* expression system is appropriate for generation of micro- to milligram quantities of pure recombinant protein with "correct" disulfides and the option of isotopic enrichment for NMR studies.

The following protocol was followed:

1. A 5-ml aliquot of YPD was inoculated with the *P. pastoris* strain KM71H and grown overnight at 30 °C.
2. Then 500 ml of fresh YPD medium was poured into a two-litre flask, inoculated with 0.1-0.5 ml of the overnight culture and cells were grown overnight to an OD_{600} of 1.3-1.5.
3. The cells were subsequently centrifuged at 1500 g for five minutes at 4°C, following which the pellet was resuspended with 500 ml of ice-cold, sterile water.
4. The cells were once again centrifuged (as in Step 3), and the pellet resuspended with 250 ml of ice-cold, sterile water.
5. Cells were centrifuged again (as in Step 3), and the pellet resuspended with 20 ml of ice-cold 1 M sorbitol.
6. Finally cells were centrifuged (as in Step 3), and the pellet resuspended with 1 ml of ice-cold 1 M sorbitol for a final volume of approximately 1.5 ml.

An aliquot of 80 μl of the cells from Step 6 was mixed with 5-20 μg of linearized DNA (in 5-10 μl TE Buffer, or sterile water) and transferred to an ice-cold 0.2-cm electroporation cuvette. The cuvette containing the cells was incubated on ice for five minutes. Settings on the Bio-Rad GenePulser II were: charging voltage - 1500 V, capacitance - 25 μF, resistance 200 m??Ω. Immediately after pulsing for 6 s, 1.0 ml of ice-cold 1.0 M sorbitol was added to the cuvette. The contents of the cuvette were then transferred to a sterile microcentrifuge tube and incubated for 1 h at 30 °C. Finally 200-600 μl aliquots were spread on YPDS Zeocin^{TM} plates. These were incubated at 30 °C until colonies had appeared.

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

The following protocol was employed for shaker-flask scale protein expression in P. pastoris.

1. A single colony was used to inoculate 100 ml of BMG in a one-litre baffled flask. The mixture was grown at 28-30 °C in a shaking incubator (250-300 rpm) until the culture reached an OD_{600} of between two and six (this took approximately 16-18 hours).
2. The cells were harvested by centrifuging at 1500-3000 g for five minutes at room temperature. In order to induce expression, the supernatant was decanted and the cell pellet was resuspended in BMM using one-fifth to one-tenth of the original culture volume.
3. The 100-ml baffled flask was capped with a sterile sponge and returned to the incubator.
4. Every 24 hours, 100% methanol was added to maintain a final concentration of 0.5% to 1.2 % - this addition of methanol induces protein expression.
5. At several times during the induction period, 1.0 ml of the supernatant was transferred to a 1.5-ml microcentrifuge tube. These samples were used to analyze protein expression levels.
6. Prior to protein purification, PMSF and EDTA were added to final concentrations of 0.5 mM and 5 mM, respectively, in order to minimize protease activity.

Note that for the screening of expression levels for multiple clones (three or more) the protocol above was scaled down by a factor of ten.

2.1.13 *P. pastoris* recombinant protein expression in fermentors

Control and logging of temperature, pH, agitation, and air and oxygen supply allow for much higher expression yields in fermentors compared to shaker-flask growths. Any of the protein constructs that was required in larger (*i.e.* milligram) quantities for subsequent experiments, or was insufficiently expressed in shaker-flask growths, was produced in a fermentor. Protein expression in ^{15}N-enriched, or ^{15}N and ^{13}C-enriched, media (to obtain labelled samples for NMR studies) were exclusively performed in fermentors. The fermentations were carried out in two-litre or five-litre cylindrical Bioflow 3000 (New Brunswick Scientific®) fermentor vessels.

In a typical fermentation, the initial volume of growth media was 600 ml. For the inoculum, a 300-mL BMG *P. pastoris* culture was grown in a two-litre shaker flask, pelleted at 1500 g and resuspended in about 20 ml 100 mM potassium phosphate buffer, pH 6.0. The contents of the initial media, and the details of the feeding schedule, depended upon whether the protein product was to be unlabelled, ^{15}N-labelled, or ^{15}N,^{13}C-labelled:

Basal salts and initial medium:

$CaSO_4$ x 7.H_2O	0.95 g
$MgSO_4$ x 7.H_2O	12.0 g
K_2SO_4	6.0 g
H_2O	540 ml
1M potassium phosphate buffer, pH 6.0	60 ml
	600 ml

The prepared media was poured into the fermentor vessel and all the probes were attached prior to autoclaving the fully assembled fermentor unit. After autoclaving the dissolved (D) O_2-probe was charged over night. Air was bubbled continuously into the vessel through a sterile filter and agitation was set to 200 rpm in order to saturate the medium with oxygen. Also attached were pH- and temperature-probes, as well as a feed for a 2 M KOH solution (the "base feed"). The temperature was set to 30 °C. To the media were added 0.5 ml Antifoam 206 (Sigma-Aldrich) and 2.5 ml of high-purity grade fermentation trace mineral salts (PTM1 salt, Amresco®).

In the case of ^{15}N isotope-enriched fermentations, 7-8 g of $^{15}N\text{-}(NH_4)_2SO_4$ were dissolved in about 20 ml H_2O and sterile filtered (0.2 μm) into the fermentor. After addition of all reagents and pH adjustment (to pH 5.0) the dissolved oxygen under these conditions was normalised to 100. Before inoculation the agitation rate was set so as to maintain a relative oxygen level of 40 during cell growth and induction (*e.g.* the aeration is held at a constant level but the agitation rate is increased when yeast cells metabolise nutrients and the oxygen level drops below a relative value of 40). When, however, maximum agitation rates of 1000 rpm proved inadequate to maintain the relative oxygen level at 40, a spike of oxygen (supplied by an oxygen cylinder) was utilised (in fermentations of up to a one-litre scale).

In ^{15}N isotope-enriched fermentations the cells were grown, in the presence of glycerol as carbon source, to high densities (for one to two days) prior to reduction of the temperature to 15 °C and initial induction with 0.5% (of culture volume) methanol. After consumption of the initial methanol, further methanol feeds of 1.0% to 1.5% of culture volume were provided for three to four days. The frequency of methanol feeds was determined on the basis of carefully monitoring the agitation-rate and dissolved oxygen curves so as to avoid overfeeding or poisoning. Every glycerol and methanol feed was accompanied by a 0.1 ml addition of PTM1-salts.

For ^{15}N and ^{13}C enrichment, 7 g of $^{15}N\text{-}(NH_4)_2SO_4$ and 15 g ^{13}C-glucose were dissolved in about 50 ml H_2O and sterile-filtered (using a 0.2-μm cut-off filter) into the fermentor prior to inoculation. The ^{13}C-labelled glucose, rather than ^{13}C-glycerol, was used at this juncture because it represents a significantly cheaper source of ^{13}C. In cases where an insufficient cell density was reached after consumption of 15 g ^{13}C-glucose, another 5 to 10 g of ^{13}C-glucose were added, as before. Trace amounts of glucose suppress the methanolinducible promoter; hence after consumption of all the ^{13}C-glucose, 1 g of ^{13}C-glycerol was added to ensure its de-repression. Reduction of the temperature to 15 °C ensued. This was followed by induction with ^{13}C-methanol and then by additional ^{13}C-methanol feeds as required to maintain cell growth,

The protocol for fermentations with non-isotope enriched media differed slightly. A 34% NH_4OH solution was utilised as a combined nitrogen source and base for regulating the pH. This dictates a change in the composition of basal salts – to the values listed below - in order to avoid excessive salt concentrations that could have adverse effects on recombinant protein expression.

Phosphoric acid 85%	27 ml
$CaSO_4$ x $7.H_2O$	0.95 g
$MgSO_4$ x $7.H_2O$	15.0 g
K_2SO_4	18.2 g
KOH	4.2
Glycerol	25 ml
Add H_2O up to	1L

Prior to calibrating the DO_2 probe, the pH was adjusted to pH 5 by addition of the 34% NH_4OH solution.

In all cases, the supernatant was harvested by spinning the cell suspension initially for 10 min at 5000 x g. The supernatant was decanted into second set of centrifuge bottles and spun again for 30 min at 8000 x g prior to sterile filtration (0.2 μm). Prior to protein purification, PMSF and EDTA were added to final concentrations of 0.5 mM and 5 mM, respectively.

2.1.14 Estimation of protein concentrations

Protein concentrations were determined by UV spectroscopy (Eppendorf BioSpectrometer, Eppendorf, Hamburg, Germany) at a wavelength of 280 nm. The purity of DNA samples were estimated by the absorbance ratios of A_{280nm} / A_{260nm} and A_{280nm} / A_{320nm}. Extinction coefficients for the protein constructs were calculated from the primary sequence with online ExPASy ProtParam tool (115). Concentrations were calculated according to the Beer-Lambert equation.

2.1.15 Trichloroacetic acid (TCA) precipitation to concentrate protein samples

To the protein sample was added an equal volume of a 20% (v/v) solution of TCA . The mixture was incubate for 30 minutes on ice and then spun in the microfuge for 15 min at 4 °C. The supernatant was removed carefully and 300 μl of cold acetone was added prior to spinning for 5 min at 4 °C. The supernatant was, again, decanted carefully. After air-drying the pellet, it was resuspended in SDS-PAGE loading buffer and loaded onto a SDS-PAGE gel after heating at 65 °C for 180 s.

2.1.16 Protein concentration in membrane spin concentrators

Buffer-exchanges for NMR samples, and all the concentration steps for the variously labelled protein samples were performed in 0.5-ml, 6.0-ml or 20-ml Vivaspin™ concentrators (Sartorius Mechatronics UK Ltd, Epsom,United Kingdom), with an appropriate molecular weight cut-off membrane (3000 - 10000 Da).

2.1.17 Sodium dodecyl-sulphate polyacrylamide gel elctrophoresis (SDS PAGE)

Polyacrylamide gradient gels (4–20% Criterion and from Bio-Rad, 4-12% NuPage® Bis-Tris gels from Invitrogen) were prepared following the protocol provided. Equal volumes of protein samples and (double concentration) protein loading buffer were mixed and heated prior to gel loading. Broad range protein marker was loaded as a reference. The gel was run at 100-200 V in TGS-buffer (for the BioRad system) or NuPAGE® MES-buffer (for the Invitrogen system) until good separation of the markers was obtained. Visualisation of protein bands was obtained by staining gels with Bio-Safe Coomassie stain (Bio-Rad, CA, USA).

2.1.18 Glycoprotein detection

After SDS-PAGE, gels were fixed by complete immersion in the Fixing Solution. The Fixing Solution was subsequently replaced with ultra-pure water and the gels agitated gently, then transferred to the provided periodic acid solution followed by a further short period of gentle agitation. The periodic acid solution was removed in a second washing step with water, then gels were immersed in Schiff's Reagent (Fuchsin-Sulfite Reagent) that stains glycans pink. This was, in turn replaced by provided reduction-solution (sodium metabisulfite). Finally the gels were subjected to a series of two or three washing steps with water.

2.1.19 Enzymatic deglycosylation of N-linked glycans

Proteins that exhibited N-linked glycosylation according to the above test were deglycosylated either prior to purification, or between the first and second purification steps, by incubating 100 ml supernatant with

6000 U EndoHf (New England Biolabs, MA, USA) at 37 °C for three hours. Yields were typically in the region of 0.1-0.5 mg of pure protein/g of wet cells.

2.1.20 Western blot

A two-step Western blot was adopted to recognise c-myc-containing protein constructs. The primary anti-myc-tag polyclonal antibody (Cell Signalling Technology®) used was raised rabbit. The secondary anti-rabbit IgG, conjugated with horseradish peroxidase (HRP) (Cell Signalling Technology®), was raised in goat. First, proteins were resolved using SDS-PAGE as described in above. The staining step was omitted. Then, the gel was soaked in Towbin-buffer and electro-transferred onto a nitrocellulose membrane (BioRad Laboratories). Electro-transfer was carried out in a Mini Trans-Blot cell (BioRad Laboratories) at a constant current of 150 mA for 90 minutes in Towbin-buffer. Pre-stained molecular weight markers allowed monitoring of the transfer onto nitrocellulose membrane. Blotted membranes were blocked for two hours in 100 ml of PBS containing 5% non-fat dried milk (blocking buffer). Overnight incubation at 4 °C on an orbital rocking platform with 5 ml of 1:1000 dilution of the primary antibody in blocking buffer followed. After rinsing the membrane with PBS it was washed with PBS containing 0.05% TWEEN20 for 20 min. After another washing step in PBS for 10 minutes, the membrane was incubated for two hours with 25 ml of a 1:3000 dilution of the secondary antibody on an orbital rocking platform. Membrane washing steps identical to the ones used after the addition of the first antibody (described above) were used. A SuperSignal® West PICO-Chemiluminescent substrate Trial Kit (Pierce®) was used to generate the enhanced chemiluminescence (at 425 nm) for the detection of the secondary antibody. The signal was detected by an X-ray film.

2.1.21 Protein chromatography

Buffer reagents and solvents used for protein purification were purchased from Sigma-Aldrich® and Fisher-Scientific® (Fisher Scientific UK Ltd, Loughborough, UK). Unless stated otherwise, all pre-packed columns and affinity resins were purchased from GE Healthcare Bio-Sciences AB, Sweden. Protein samples were buffer-exchanged either by multiple dilutions and concentrations of samples in Vivaspin concentrators, or by applying them to a 5-ml PD-10 desalting column or a 5-ml HiTrap™ desalting column (both from Amersham Biosciences®).

For the initial purification steps, bench-top columns (graduated Econo-Pac® polypropylene column from BioRad Laboratories, CA, USA) were packed with SP 6 FF Sepharose or Heparin 6 FF Sepharose. Prior to loading, the crude supernatant from fermentors or shaker flasks were either concentrated and buffer-exchanged as described above, or simply diluted and adjusted to a pH suitable for sample binding. After washing with low-salt concentration buffer, samples were eluted from the resin with suitable buffers supplemented with 1.0 M NaCl. Subsequent chromatography steps were carried out on BioCad® 700E Reperfusion chromatography system or on a ÄKTAdesign™-FPLC system (pump P-920, UV-detector unit UPC-900).

Prior to ion-exchange chromatography, samples were buffer-exchanged into binding buffer conditions (20 mM potassium phosphate, 20 mM sodium acetate or 20 mM sodium carbonate buffers at pH values suitable for the chosen type of ion-chromatography and sample pI). Elution of samples from ion chromatography columns was achieved by applying a gradient from 100% binding buffer to 100% elution buffer. The elution buffer generally consisted of the same buffering agent and pH as the binding buffer, but was substituted with NaCl to yield a concentration of 1.0 M NaCl. For cation-exchange the Tricorn™ MonoS column 4.6/100 (1.7 ml CV) was used. For anion-exchange the Tricorn™ MonoQ 4.6/100 (1.7 ml CV) column and HiTrap™ CaptoQ (dimensions of each: 7 mm x 25 mm) were used (all from Amersham Biosciences®).[5]

Ni^{2+}-affinity chromatography was employed to purify recombinant proteins that had been expressed with a hexa-histidine tag. Binding buffer consisted of 20 mM potassium phosphate buffer at pH 7.0, supplemented with NaCl to a final concentration of 0.5 M to inhibit non-specific binding. The elution buffer was made by adding imidazole to the binding buffer to a final concentration of 0.5 M. For Ni^{2+} affinity chromatography a XK 16/20 column was packed with 25 ml of IMAC Sepharose 6 FF resin .

Unless stated otherwise, gel filtration chromatography was performed in a buffer containing 0.5 M to 1.0 M NaCl and 20 mM potassium phosphate buffer at a pH of 6.6 or 7.4. Gel filtration was performed on a HiLoad™ 16/60 Superdex™ 75 prep grade column.

Heparin-affinity chromatography, a subtype of cation exchange chromatography that additionally features selection based on affinity for carbohydrate polymers, was performed with 1-ml and 5-ml HiTrap™ heparin-affinity columns (dimensions of each: 7 mm x 25 mm) or a PorosHE heparin-affinity chromatography column (4.6 x 100 mm, Applied Biosystems, Warrington, UK).

---

[5] All columns (apart from reversed-phase HPLC columns) and chromatography-resins were purchased from Amersham Biosciences®.

Reversed-phase chromatography steps were carried out on a Waters HPLC system comprising a Waters' 600 Controller fitted with a Waters' 486 tuneable absorbance detector. Separation was achieved by a gradient elution from the initial conditions of 95:5 v/v H_2O:acetonitrile to final conditions of 5:95 v/v H_2O:acetonitrile. Column used was Supelco Discovery® BIO Wide Pore C5 HPLC column.

2.1.22 Mass spectroscopy

2.1.22.1 MALDI-TOF peptide-mass fingerprinting

Purified proteins were suspended in 50 mM ammonium bicarbonate, then reduced and alkylated using 5 mM DTT and 15 mM iodoacetamide. Following digestion with trypsin (1:100 (w/w) trypsin:protein, two hours, 37 °C), a 0.5-μl aliquot of the digest was mixed with 0.5 μl α-cyano-4-hydroxycinnamic acid matrix (10 mg/ml in 50% (v/v) acetonitrile in water containing 0.1% (v/v) trifluoroacetic acid) on a MALDI sample plate. Samples were then analyzed on a Voyager-DE STR biospectrometry workstation MALDI-TOF mass spectrometer (Applied Biosystems), and processed spectra were searched against the National Centre for Biotechnology Information non-redundant database or in-house database using ProteinProspector (University of California, San Francisco, http://prospector.ucsf.edu) or Mascot (Matrix Science, http://www.matrixscience.com).

2.1.22.2 Electrospray ionisation mass spectrometry

The mass spectra were recorded on a Micromass (Manchester, UK) single quadrupole mass spectrometer fitted with an electrospray ionisation source. This was used in the positive-ion mode. Tuning parameters were as follows: ES+ was 2, the capillary voltage was 3.50 kV; HV:Lens 0.50 kV; 4, Cone: 50 V; 5, Skimmer Offset 5 V.

Prior to injecting the sample into the mass spectrometer, the sample was desalted during a LC step on a Luna™ (Phenomenex, Hurdsfield, UK) 5u-C5 column (dimensions 50 x 2 mm) using the following distilled water-acetonitrile gradient: the initial conditions of 95:5 v/v H2O:acetonitrile was adjusted over 30 min to the final conditions of 5:95 v/v H2O:AcCN.

2.1.23 N-terminal sequencing

N-terminal sequencing was performed by the Dr Nick Morrice from the University of Dundee.

2.2 Functional characterisation

2.2.1 Heparin binding studies

2.2.1.1 Heparin chromatography

Protein samples (50–78 μg, 1 ml) in 20 mM potassium phosphate buffer (pH 7.4) were loaded individually onto either a HiTrap heparin-affinity chromatography column (7 x 25 mm, GE Healthcare) or a Poros 20HE heparin-affinity chromatography column (4.6 x 100 mm, Applied Biosystems) equilibrated with 20 mM potassium phosphate buffer (at pH 7.4) and subsequently eluted with a linear gradient of 0–1 M sodium chloride in 20 mM potassium phosphate buffer (pH 7.4).

2.2.1.2 GMSA[6]

Oligosaccharides were prepared from low-molecular-weight heparin by partial digestion with heparinase I followed by size fractionation on a Bio-Gel P10 gel filtration column (Bio-Rad) (116). Fluorophore-labeled species were produced by attachment of 2-aminoacridone

---

[6] Apart from sample preparation, GMSA analyses were setup and performed by Bärbel Blaum.

to the oligosaccharide reducing end (116), and GMSA were performed, as described previously (117). Briefly, 2-aminoacridone tagged oligosaccharides were combined with the recombinant segments of FH at a range of concentrations in a volume of 10 μl of PBS containing 25% (v/v) glycerol for 15 minutes (at room temperature). Samples were then loaded on a 1% agarose gel in 10 mM Tris-HCl (pH 7.4) and 1 mM EDTA. Electrophoresis was performed (200 V, 8–15 minutes) in a horizontal agarose electrophoresis system using an electrophoresis buffer comprising 40 mM Tris/acetate, 1 mM EDTA (pH 8.0). Immediately thereafter, the fluorescent oligosaccharides were visualized.

2.2.2 Co-factor activity

An endpoint fluid phase assay was used to measure co-factor activity for factor I mediated cleavage of C3b into iCb. Factor I and FH (positive control) were purchased from Complement Technology Inc., TX, USA. Equimolar amounts (2 μM in the final 30 μl reaction) of Factor I and C3b (4 μg and 10 μg, respectively) of were mixed with Factor H (5 μg; 1 μM in 30μl) or FH-1-4 (2.9 μg in PBS; 3.3 μM in 30 μl) and PBS was added to a volume of 30 μl. The reaction mix was vortexed and incubated in a waterbath at 37 °C for one hour. Reducing SDS loading buffer was added immediately and the mix was heated prior to loading on a gel.

2.2.3 Binding to C3 fragments C3b, C3c and C3d

Factor H binding to C3b, C3c and C3d was monitored by surface plasmon resonance (SPR) using a Biacore T100 instrument (GE Healthcare). The sensor surfaces were prepared by immobilizing human C3b, C3c or C3d (Complement Technology) in two or three of the four flow cells of Biacore series S carboxymethylated dextran (CM5) or carboxymethylated matrix-free (C1) sensor chips (GE Healthcare), using standard amine coupling; the reference surface in each case was prepared in the remaining flow cell(s) by performing a dummy coupling reaction in the absence of any proteins. Experiments were performed at 25 °C using a flow rate of 30 μl/min (after performing a flow-rate study to check for mass transport limitation). Duplicate injections of FH samples in 10 mM HEPES-buffered 150 mM saline with 3 mM EDTA and 0.05% (v/v) surfactant p20 (HBSEP) were performed at the concentrations indicated. A contact time of 90 s was used, as this was found to be sufficient to achieve steady-state conditions, followed by a dissociation time of 600 s with HBS-EP as running buffer. The chips were regenerated between sample-injections by two injections of 1 M NaCl for contact times of 45 s. Data were processed using Biacore T100 evaluation software version 1.1. Reporter points for affinity measurements were set to 2 s before injection began and 2 s before the injection period finished. Dissociation constants were calculated by fitting steady-state binding levels derived from the background-subtracted traces to a one-to-one binding steady-state model

2.3 NMR structural studies

One- and two-dimentional NMR spectra were normally recorded to evaluate the quality of recombinantly expressed constructs. A suite of two- and three-dimensional NMR experiments was recorded for the structure determination of FH-12-13. As no NMR development was undertaken in this project, a brief overview of NMR theory will exclusively deal with spectra processing and interpretation, followed by an outline of structure calculation methods used in this study. Comprehensive experimental setup guidance and pulse sequence documentation are may be accessed online at (http://nmr-linux.chem.ed.ac.uk/highfield/highfield.html).

2.3.1 Sample preparation

Overexpressed constructs were purified as described in CHAPTER 3. Purified samples were concentrated in spin concentrators and buffer exchanged into 20 mM potassium phosphate buffer pH 6.6 or pH 7.4. Concentrations for 1-D experiments ranged from 50 to 125 μM. For

2- and 3-dimensional experiments, sample concentrations were 100 to 620 μM. In case of low expression or purification yields, samples were concentrated to 300 μl and transferred into Shigemi tubes. D_2O and NaN_3 were added to all NMR samples to yield 5 to 10 % (v/v) and about 0.01 %, respectively.

2.3.2 Data collection

All spectra were recorded on Bruker AVANCE™ 14.1 Tesla (600 MHz) spectrometer or a Bruker AVANCE™ 18.8 Tesla (800 MHz) spectrometers fitted with 5-mm triple-resonance probe/cryoprobes. All 2- and 3-dimensional spectra were recorded at 37 °C. All spectra were acquired by the author with the assistance of Dr Dusan Uhrín, Mr Juraj Bella and Dr Andrew Herbert (University of Edinburgh). Spectra recorded for the structure determination of FH-12-13 are listed in Table 2.6.

Experiment	Experiment type	labelling required	Dim 1	Dim 2	Dim 3
^{15}N-HSQC (118)	through-bond	^{15}N	^{1}H	^{15}N	n/a
^{13}C-HSQC (119)	through-bond	^{13}C	^{1}H	^{13}C	n/a
CBCA(CO)NH (120)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}C$_{\alpha/\beta}$	^{15}N
CBCANH (121)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}C$_{\alpha/\beta}$	^{15}N
HBHA(CO)NH (122)	through-bond	^{15}N, ^{13}C	^{1}H	^{1}H$_{\alpha/\beta}$	^{15}N
HBHANH (123)	through-bond	^{15}N, ^{13}C	^{1}H	^{1}H$_{\alpha/\beta}$	^{15}N
HNCO (124)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}CO	^{15}N
HN(CA)CO (125)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}CO	^{15}N
H(C)(CO)NH-TOCSY (126)	through-bond	^{15}N, ^{13}C	^{1}H	^{1}H$_{\text{sidechain}}$	^{15}N
(H)C(CO)NH-TOCSY (126)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}C$_{\text{sidechain}}$	^{15}N
^{15}N-TOCSY (127)	through-bond	^{15}N, ^{13}C	^{1}H	^{1}H	^{15}N
HCCH-TOCSY (128)	through-bond	^{15}N, ^{13}C	^{1}H	^{1}H	^{13}C
(HB)CB(CGCD)HD (129)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}C$_{\text{arom}}$	n/a
(HB)CB(CGCDCE)HE (129)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}C$_{\text{arom}}$	n/a
Aromatic ^{13}C-HSQC (119)	through-bond	^{15}N, ^{13}C	^{1}H	^{13}C$_{\text{arom}}$	n/a

^{15}N-HSQC-NOESY (127)	through space	^{15}N	^{1}H	^{1}H	^{15}N
^{13}C-HSQC-NOESY (130)	through space	^{15}N, ^{13}C	^{1}H	^{1}H	^{13}C

Table 2.6 NMR experiments used in this study.

2.3.3 NMR software

The Bruker software suit Topspin™ was used for data acquisition and initial processing. Apart from the ^{13}C-HSQC-NOESY, all the NMR spectra were finally processed using AZARA (W. Boucher, Department of Biochemistry, University of Cambridge, UK). Data visualisation and resonance assignment were carried out in the NMR data analysis software package, CCPNMR Analysis (131).

2.3.4 Data processing

Fourier transformation and several window functions for maximised resolution were applied to collected raw data spectra resulting in frequency spectra. Spectra processing with *process* command in AZARA require a parameter file (ser.ref) and a script input file (scr).

2.3.5 General resonance assignment strategy

An assignment of all the NMR resonances in a protein generally requires a suite of three-dimensional (3D) experiments recorded on 13C,15N-labelled sample. The reference spectrum for most 3D experiments is the 2D 15N-HSQC, in which every cross-peak corresponds to an amide (i.e. (CO)-NH-) (Figure 2.1) and thereby correlates an amide nitrogen with its attached proton. The various 3D experiments allow these correlations to be extended through-bonds to include other atoms. In this way, nuclei, as identified by their resonance frequencies, may be assigned to clusters corresponding to covalently linked groups of atoms. The redundancy of information in four pairs of 3D experiments allows overlapping clusters to be placed in sequential order and then matched to specific stretches of amino acid residues within the polypeptide sequence Figure 2.2).

The 3D ^{15}N-TOCSY, 3-D H(CC)(CO)NH-TOCSY and 3-D (H)CC(CO)NH-TOCSY are used to extend assignments from the backbone amides into the side-chain spin systems. Together with the HCCH-TOCSY experiment, which yields cross-peaks for every aliphatic (attached to ^{13}C) side-chain ^{1}H, these experiments should in theory enable the assignment of all of the non-aromatic side-chain atoms. Aromatic atoms can be identified by a set of 2D experiments that correlate aromatic HD and HE shifts with CB shifts of the same side-chain.

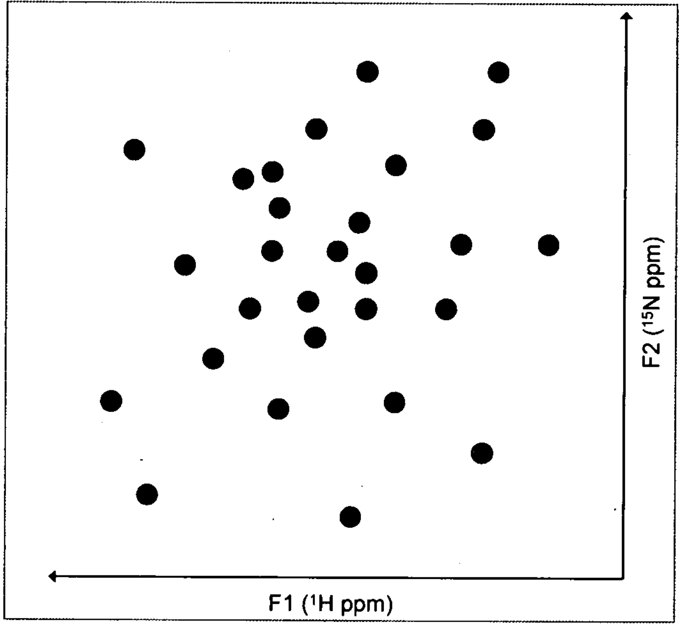


Figure 2.1 15N-HSQC spectrum. Each cross-peak (blue dot) represents an amide proton. Its correlated proton and nitrogen shifts (frequencies) can be extrapolated from the two axes, where F1 is the first dimension (the amide proton dimension) and F2 is the second, indirectly detected dimension (the nitrogen dimension).

These assignment are subsequently transferred to the ^{13}C- and ^{15}N-edited NOESY spectra to facilitate assignment of NOESY cross-peaks that arise from NOE transfers between non-covalently linked protons which are close in space.

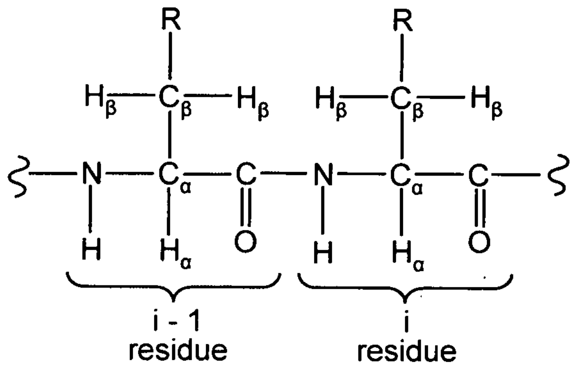


Figure 2.2 **Nomenclature of sequential amino acids in a polypeptide chain The residue i is preceded by the residue i-1. α- and β-atoms are indicated.**

2.3.6 Assignment of backbone resonances

For ^{13}C,^{15}N-labelled proteins, backbone assignment of FH-12-13 relied on the matching of cross-peaks in pairs of complementary experiment. In this way, signals (in the 3rd dimension) from the (i) and the (i-1) residue, which are transferred to the (i)-amide proton, were correlated with matching signals of the (i-1) residue, which are also transferred to the (i)-amide proton; identification of matching shifts in the 3rd dimension of the experiment pairs CBCA(CO)NH - CBCANH, HBHA(CO)NH - HBHANH and HN(CA)CO – HNCO allowed sequential assignment of the two amino acid residues involved and the atomic assignment of the (C^{α}, C^{β}) resonances, the (H^{α}, H^{β}) resonances and the CO resonances, respectively.

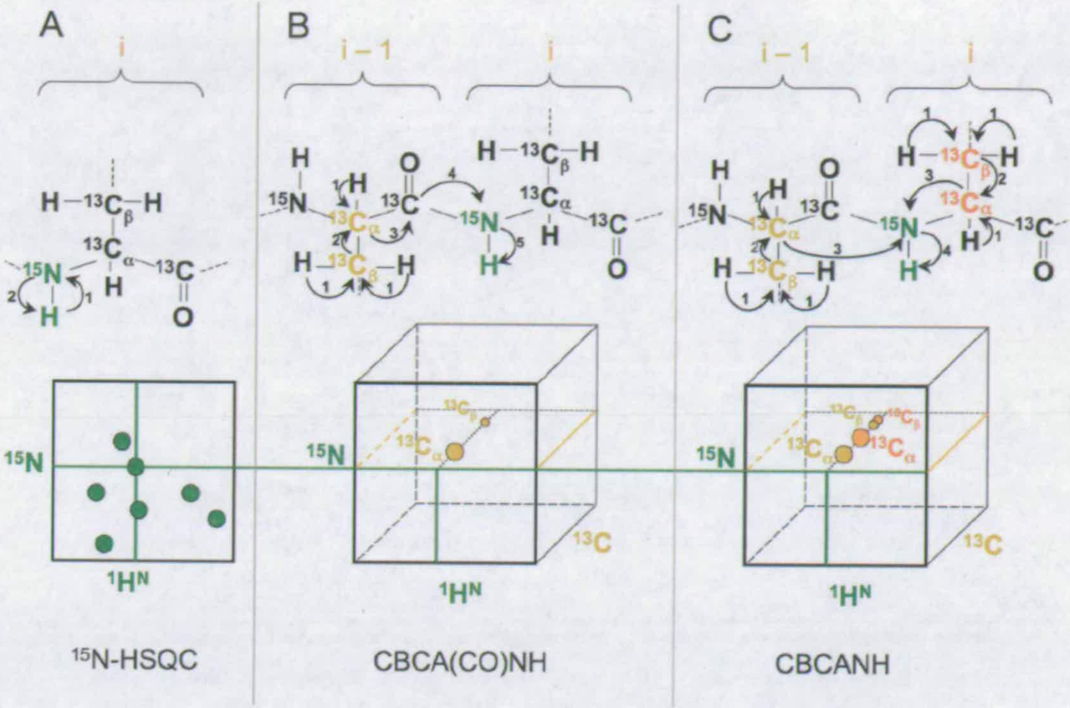


Figure 2.3 Magnetisation transfer and 3-D spectra representation of CBCA(CO)NH and CBCANH experiments. Magnetisation transfer (top-panels) and the multi-dimensional spectrum representations (bottom panels) of (A) the 15N-HSQC, (B) the CBCA(CO)NH and (C) the CBCANH experiments. Atoms labelled during the evolution periods of the pulse programs and whose dimensions are detected in the spectra are colour coded for clarity. Arrows indicate the magnetisation transfer steps of the experiment. Only one strip is shown for the 3-D CBCA(CO)NH and CBCANH along with the corresponding amide cross-peak in the 2-D ^{15}N-HSQC.

The first two dimension of each of these experiments were the amide proton (H_N) and amide nitrogen (N_H) frequencies – hence the plane described by these two axes depicts the reference ^{15}N-HSQC spectrum. Figure 2.3[7] describes magnetisation transfer of the CBCA(CO)NH and CBCANH experiments and illustrates representatively dimension organisation in 3-D spectra used for protein backbone assignment. Figure 2.4 shows matching connectivity between CBCA(CO)NH and CBCANH planes of the construct FH-12-13.

---

[7] Figure and legend are reproduced from PhD Thesis "The structure of an active N-terminal fragment of human complement factor H" by Hocking, HG

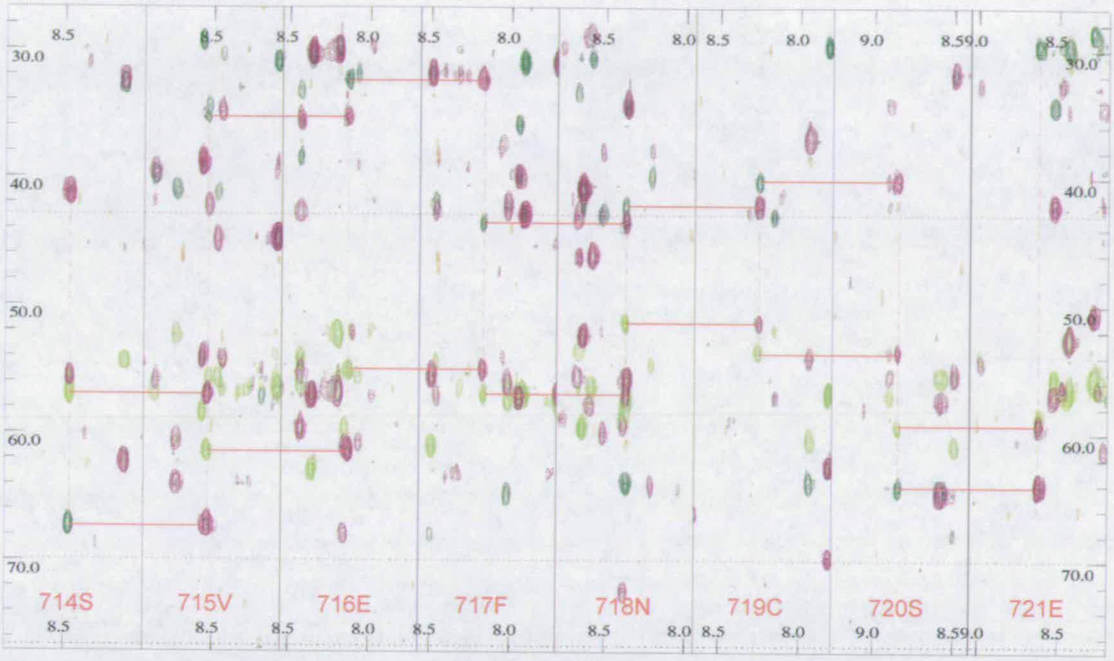


Figure 2.4 Triple resonance NMR CBCA(CO)NH (purple) and CBCANH (green) spectra of FH-12-13. Spectra allow to trace sequential backbone-connectivities. Sequence number and amino acids identifier are show in red print.

2.3.7 Assignment of side-chain resonances

TOCSY (Total Correlation Spectroscopy) experiments were employed for the assignment of ^{1}H and ^{13}C side-chain resonances. The ^{15}N-TOCSY-HSQC (Figure 2.5[8]) correlates aliphatic side-chain proton shifts from the (i) residue with root resonances (H^{N} and N^{H} shifts) of the same residue. H(CC)(CO)NH-TOCSY and (H)CC(CO)NH-TOCSY experiments correlate aliphatic proton shifts and aliphatic carbon shifts of the (i-1) residue, respectively, with the root resonances of following amino acid (i). These side-chain assignments were transferred to the ^{13}C-HSQC spectrum which then served as a reference point for the assignment of the HCCH-TOCSY spectrum. Similarly to the ^{15}N-TOCSY-HSQC, the HCCH-TOCSY spectrum correlates (Figure 2.5) all proton shifts from the (i) residue with ^{13}C-HSQC root resonances (H^C and C^H shifts) of the same residue.

---

[8] Figure and legend are reproduced from PhD Thesis "The structure of an active N-terminal fragment of human complement factor H" by Hocking, HG.

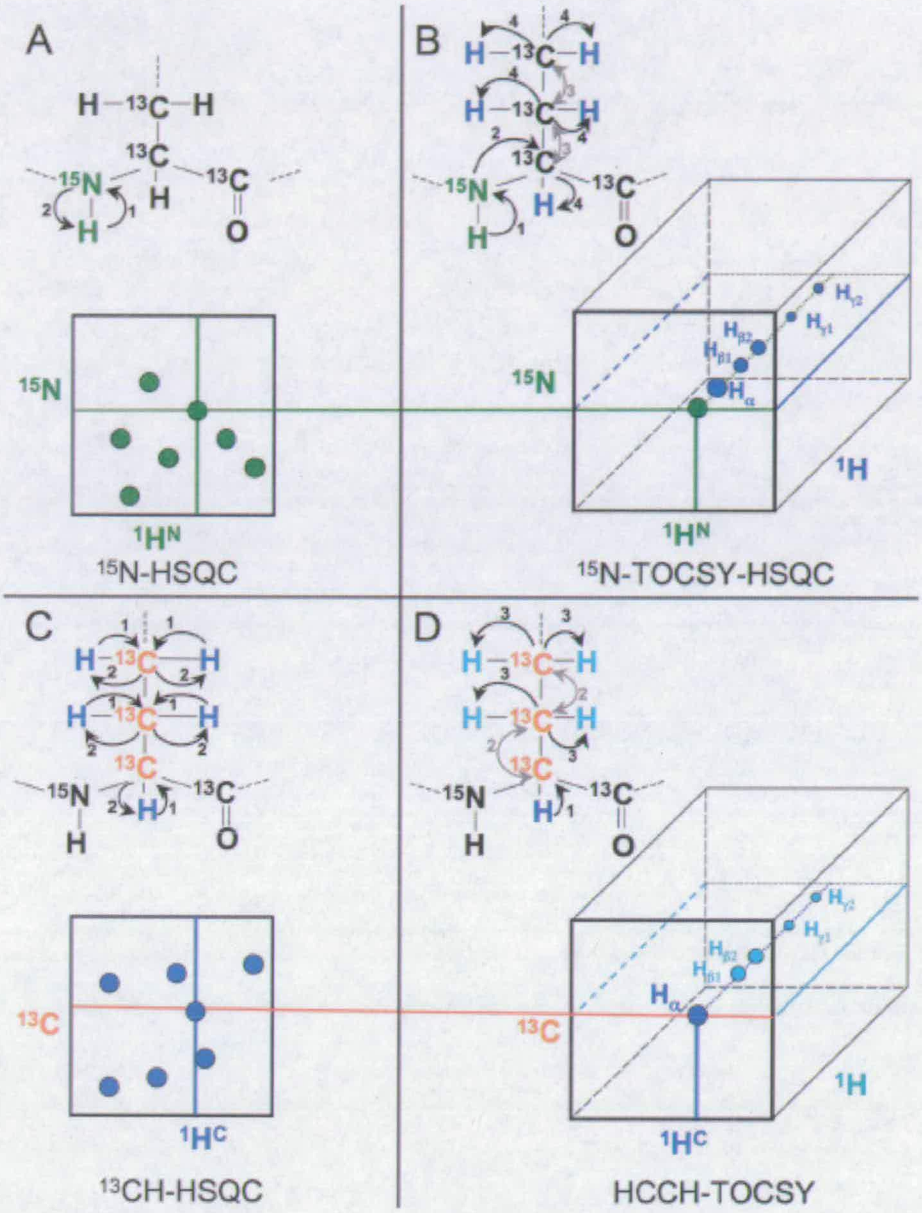

A
B
C
D
15N
1HN
15N-HSQC
15N
1HN
1H
Hα
Hβ1
Hβ2
Hγ1
Hγ2
15N-TOCSY-HSQC
13C
1HC
13CH-HSQC
13C
1HC
1H
Hα
Hβ1
Hβ2
Hγ1
Hγ2
HCCH-TOCSY

Figure 2.5 Aliphatic proton assignment using 15N-TOCSY-HSQC and the HCCH-TOCSY. Magnetisation pathways (top panels) and multi-dimensional spectrum representations (bottom panels) for (A) the ^{15}N-HSQC, (B) the ^{15}N-TOCSY-HSQC, (C) the ^{13}CH-HSQC and (D) the HCCH-TOCSY. Atoms labelled during the evolution periods of the pulse programs and whose dimensions are detected in the spectra are colour coded for clarity. Arrows indicate the magnetisation transfer steps of the experiment. Only one strip is shown for the 3-D ^{15}N-

TOCSYHSQC and the 3-D HCCH-TOCSY, along with their corresponding amide cross-peak in the 2-D ^{15}N-HSQC, and ^{13}CH cross-peak in the 2-D ^{13}CH-HSQC.

2.3.8 Assignment of aromatic resonances

Two 2-D experiments specialised for detecting aromatic proton shift, (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE, were employed for the shift assignment from hisitidine, phenylalanine, tyrosine and tryptophan residues. These experiments correlate C^β shifts of aromatic amino acids with H^δ or H^ε ring protons. In case of assigned C^β shifts and not significantly overlapped 2-D aromatic spectra, the H^δ or H^ε ring protons could be correlated to the C^β shift in the ^{13}C-HSQC spectrum.

2.3.9 Atom assignment – summery

The above discussed assignment strategies yielded a near-complete atomic assignment (98.3 % of all atoms) of FH-12-13. Chemical shift assignment of more than 90% of all backbone amide and non-labile proton shifts (132) is a requirement for correct automated NOE cross-peak assignment by the program CYANA employed in this study[9].

2.3.10 Cis-trans- Proline assignment

The construct FH-12-13 contains five proline residues. *Trans*- or *cis*-configuration of the proline peptide bonds was determined by calculating the difference between the C^β and C^γ shifts (ΔC^β-C^γ) of each proline residue. A statistical analyses of ^{13}C chemical shifts from high resolution NMR structures (133) showed ΔC^β-C^γ in *trans* prolines to be 4.51 ± 1.17 ppm and in *cis* prolines to be 9.64 ± 1.27 ppm. Table 2.7 lists ΔC^β-C^γ values for all proline residues in FH-12-13.

---

[9] Chemical shift tables and all distance restraints and NMR data will be deposited in the BioMagResBank (http://www.bmrb.wisc.edu/)

Pro-sequence number	ΔC^{β}-C^{γ}	ΔC^{β}-C^{γ}	ΔC^{β}-C^{γ}	conformation
695	32.32	27.13	5.18	*trans*
707	30.57	26.41	4.15	*trans*
708	34.73	24.817	9.91	*cis*
742	32.47	26.25	6.22	*trans*
799	34.83	25.95	8.89	*cis*

Table 2.7 List of ΔC^{β}-C^{γ} values and conformation assignment for all proline residues in FH-12-13.

Prolines 708 and 799 have ΔC^{β}-C^{γ} values of 9.91 and 8.89, respectively, and were configured *cis*. The remaining prolines were configured *trans*. The configuration of proline peptide bonds by chemical shift differences was cross-validated by analysing H^{α} to Pro-H^{δ} and H^{α} to Pro-H^{α} NOE-patterns in the ^{13}C-edited NOESY spectrum. For a *trans*-proline peptide bond strong H^{α} to Pro-H^{δ} and weak H^{α} to Pro-H^{α} NOE cross-peaks are expected; for a *cis*-proline peptide bond weak H^{α} to Pro-H^{δ} and strong H^{α} to Pro-H^{α} NOE cross-peaks are expected. The NOE patterns confirmed the configuration of proline-peptide bonds.

2.3.11 Assignment of NOE cross-peaks

Assignments from the two reference spectra ^{15}N-HSQC and ^{13}C-HSQC were transferred to both 3-D NOESY spectra (^{15}N-HSQC-NOESY and ^{13}C-HSQC-NOESY) by overlaying planes of the 3-D NOESY spectra onto the relevant reference spectra. Intra-residue NOE cross-peaks were identified by overlaying proton-proton planes of the ^{15}N-HSQC-NOESY and ^{13}C-HSQC-NOESY spectra on top of the ^{15}N-TOCSY and HCCH-TOCSY spectra, respectively.

A NOE cross-peak between two protons is usually corroborated by its symmetry-related cross-peak as described in Figure 2.6.

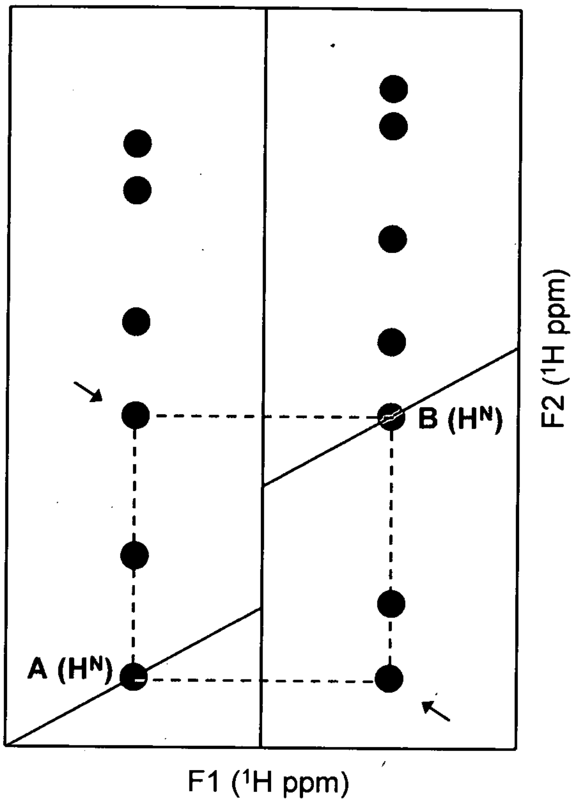


Figure 2.6 Symmetry related NOE cross-peaks. Proton A experiences "through space" dipolar coupling of proton B resulting in a NOE cross-peak at the proton frequency of B. Due to the mutual nature of the NOE effect proton B experiences a corresponding effect resulting in a NOE cross-peak at the proton frequency of A. The two symmetry related cross-peaks are indicated by an arrow.

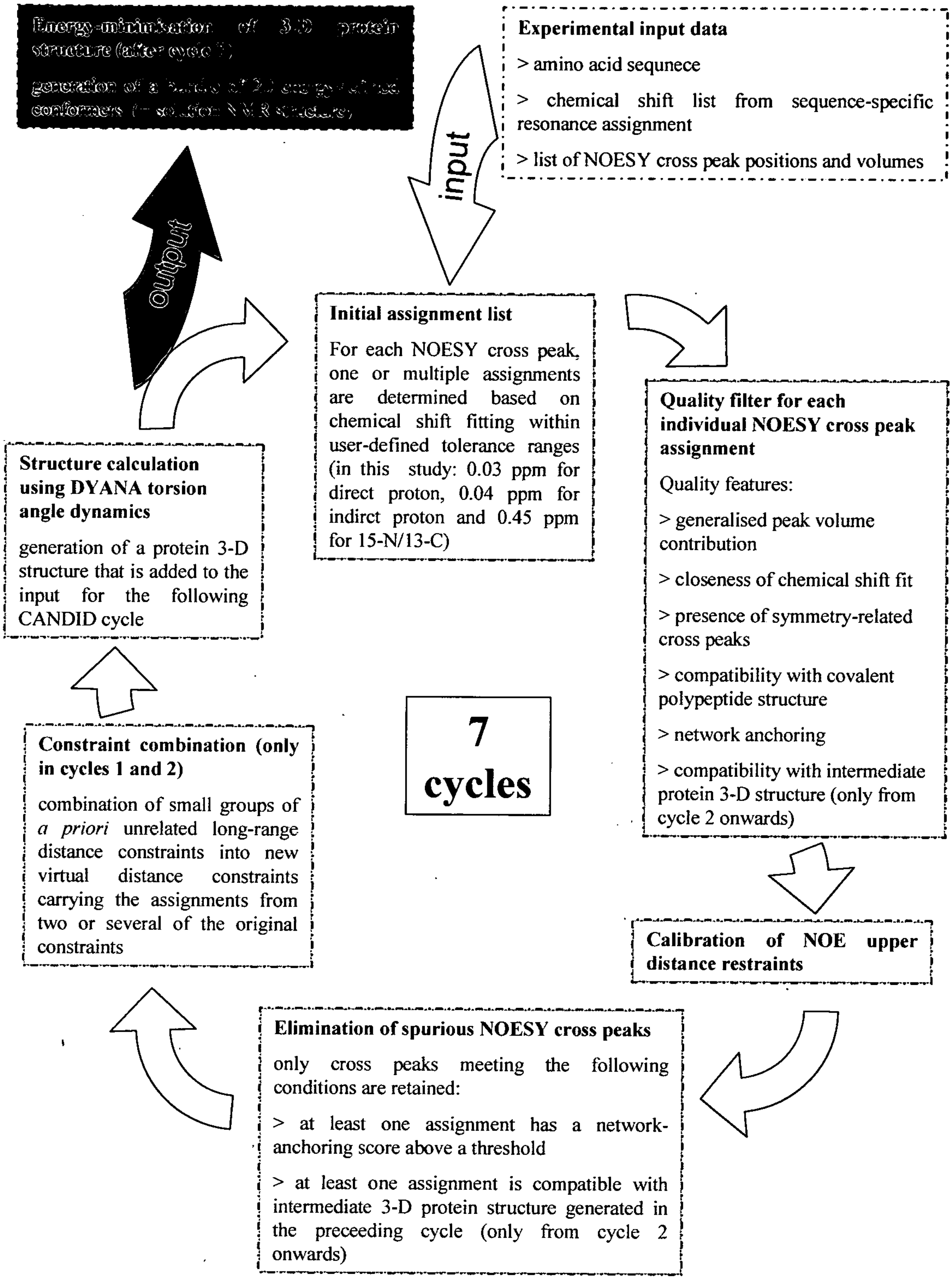
Second-dimension proton shifts (^{1}H in the F2 dimension) were only assigned to specific nuclei where the two corresponding symmetry-related NOE cross-peaks could easily be identified. In cases of peak overlap or degenerate symmetry-related cross-peaks, the peaks were left unassigned and treated as ambiguous by the automated assignment and structure calculation software. Unambiguous assignment of proton shifts in the F2 dimension was mainly focused on amide-to-amide, amide-to-side-chain and aromatic-to-any proton cross-peaks.

Peak lists for both NOESY spectra alongside with a chemical shift list table of all assigned atoms in FH-12-13 were exported from the CCPNMR ANALYSIS project in the XEASY format. NOE intensities were calculated using the "box sum" routine in CCPNMR

ANALYSIS software package. A process of indirect magnetisation transfer – known as spin diffusion – between two protons far in space via a dense network of protons can result in not completely accurate NOE intensities and consequently partly erroneous distance restrains. To minimise these effects and restrict magnetisation transfer to close spin pairs, both NOESY pulse sequences were implemented with minimal mixing times.

2.3.12 Structure calculation in CYANA

The CYANA program for NMR structure calculation is based on its predecessor DYANA. DYANA is an acronym for DYnamics Algorithm for Nmr Application and calculates structures from distance restraints, as well as torsion angle constraints, collected by NMR (134). For initial structure calculation and semi-automated NOE assignment the structure-calculation software CYANA (135) (version 2.1) was used. The program employs multiple parallel sampling of conformational space to drive the minimisation of a potential energy target function in combination with fulfilling the experimentally derived distance constraints. In other words empirical restraints (*i.e.* bond lengths, bond angles and prochiralities) are kept at their optimal value while the employed torsion angle dynamics minimise a variable potential energy target function (in torsion angle space) to satisfy the experimental distance restraints. The energy minimisation function uses simulated annealing to prevent the process from getting trapped in local energy minima. By applying repeated *in silico* heating steps followed by slow, progressing cooling steps local energy minima can be overcome to reach the global energy minimum. The CYANA 2.1 software (135), which is partly based on DYANA, embeds the *Combined automated NOE assignment and structure determination* (CANDID) module (132) enabling NMR structure determination of proteins by automated assignment of the NOESY spectra. An iterative approach runs through multiple cycles of NOE cross-peak assignment (by the CANDID module) followed by DYANA-driven structure calculation (torsion angle dynamics). In the current study, seven such cycles assigned a user-supplied, partly manually assigned NOE cross-peak list, filtered all submitted ambiguous NOE cross-peaks for cross-peaks meeting certain quality features (*i.e.* closeness of chemical shift match or presence of symmetry-related cross peaks) and calculated a three-dimensional protein structure. Figure 2.7 schematically illustrates the automated NOE cross-peak assignment and 3D structure calculation cycling by CYANA.



Experimental input data

> amino acid sequnece

> chemical shift list from sequence-specific resonance assignment

> list of NOESY cross peak positions and volumes

Initial assignment list

For each NOESY cross peak, one or multiple assignments are determined based on chemical shift fitting within user-defined tolerance ranges (in this study: 0.03 ppm for direct proton, 0.04 ppm for indirct proton and 0.45 ppm for 15-N/13-C)

Quality filter for each individual NOESY cross peak assignment

Quality features:

> generalised peak volume contribution

> closeness of chemical shift fit

> presence of symmetry-related cross peaks

> compatibility with covalent polypeptide structure

> network anchoring

> compatibility with intermediate protein 3-D structure (only from cycle 2 onwards)

Calibration of NOE upper distance restraints

Elimination of spurious NOESY cross peaks

only cross peaks meeting the following conditions are retained:

> at least one assignment has a network-anchoring score above a threshold

> at least one assignment is compatible with intermediate 3-D protein structure generated in the preceeding cycle (only from cycle 2 onwards)

Constraint combination (only in cycles 1 and 2)

combination of small groups of *a priori* unrelated long-range distance constraints into new virtual distance constraints carrying the assignments from two or several of the original constraints

Structure calculation using DYANA torsion angle dynamics

generation of a protein 3-D structure that is added to the input for the following CANDID cycle

Energy-minimisation of 3-D protein structure (after cycle 7)

generation of a bundle of 20 energy-refined conformers (= solution NMR structure)

Figure 2.7 NOE cross-peak assignment and 3D structure calculation cycling by CYANA; adapted from (132)

After completion of the first cycle every subsequent CANDID cycle takes into account the three-dimensional protein structure from the previous cycle, in addition to the complete input used for the first cycle. Network-anchoring and constraint-combination are built in routines to prevent biased structures triggered by artefacts in the input data.

NOE cross peaks are embedded in a hugely redundant network of NOE connectivities. Network anchoring rates the initial, chemical shift-based assignment on the basis of how well the assigned NOE-connection is embedded in a (consistent) set of neighbouring NOE cross-peak assignments.

Constraint-combination combines two or several NOE (upper) distance restraints to reduce the influence of single incorrect distance restraints originating from NOE artefacts. Since after the completion of the first cycle the NOESY cross-peak evaluation will (in addition to the initial inputs) also be driven by compatibility with the intermediate 3-D structure, it is critical that the correct polypeptide fold is found during the first cycle. This was accounted for by verifying that the backbone RMSD for each individual CCP domain (of the structures generated in cycle 1) was equal or lower to 3 Å (132).

Initially two sets of structure calculation were performed in parallel; one where the conserved disulfide bonds in each module of FH-12-13 were not defined, and the second with the disulfide bonds (disulfide connectivities: Cys^{691}-Cys^{733}, Cys^{719}-Cys^{744}, Cys^{753}-Cys^{792} and Cys^{781}-Cys^{803}) were defined. The backbone RMSDs between the two sets were calculated for the single domains (CCP12, CCP13) and were found to be 0.744 Å and 0.462 Å, respectively. Therefore, this exercise allowed confirmation of the defined disulfide bonds.

2.3.13 Structure calculation in CNS

Structure refinement in a thin layer of water solvent has been shown to significantly improve structural quality criteria such as Ramachandran-plot statistics, packing quality, the number of unsatisfied H-bond donors and acceptors and backbone conformation (136). The refinement of protein structures in explicit solvent is not an option in CYANA, but it has been implemented in the Crystallography and NMR systems (CNS) structure calculation software (137). Nonetheless, it was decided to avoid employing CNS at the initial stage of FH-12-13 structure calculations because it does not have an explicit automated-NOE cross-peak assignment capability; moreover CNS-based structure calculations require more CPU time. Therefore NOE assignments were derived from the final CYANA structure calculations then imported (through FormatConverter) into ANALYSIS. Within ANALYSIS NOE intensities were converted into distance bounds (Table 2.8). Finally, the restraints list was exported into CNS format (again, using FormatConverter).

relative NOE intensities	upper distance restraint
> 3.5	2.5 Å
1.3 – 3.5	2.8 Å
0.3 – 1.3	4.0 Å
0.1 – 0.3	5.0 Å
up to 0.1	6.0 Å

Table 2.8 Relative NOE intensity rating and translation into distance restraints. Standard "Bin" NOE cross-peak volume integration into relative intensities was performed in ANALYSIS. A relative intensity of 1.0 reflects the average over all observed NOE cross-peak intensities. Upper distance restraints for each interval of relative intensities are indicated. For all intervals, the minimum distance was set to the sum of the Van-der-Waal radii of two protons, 1.8 Å.

Structure calculation by CNS (version 1.2 was used in this study) employs restrained molecular dynamics in Cartesian space (as opposed to the torsion angle space utilised in CYANA) and is driven by the minimisation of a hybrid energy target function of empirical and experimental restraints (138, 139). The function describing the overall energy target ($E_{overall}$) thus combines two energy terms, one derived from empirical restraints expressed by a force field (E_{FF}) and one derived from experimental, NOE-derived, distance restraints (E_{NOE}):

$$E_{overall} = E_{FF} + E_{NOE}$$

The energy term of empirical restraints (E_{FF}) is described by: bond lengths (E_{bond}), bond angles (E_{angles}), dihedral angles ($E_{improper}$), chiralities ($E_{improper}$) and a van-der-Waals repulsive term (E_{vdW}) that prevents overly close contacts between atoms:

$$E_{FF} = E_{bond} + E_{angles} + E_{improper} + E_{vdW}$$

The CNS protocol used in the current study (see structure calculation process involves different stages, along which the weighting for NOE distance restraints progressively increases: generation of random starting structure, local geometry optimisation of starting structure, prochiral swapping (randomly swapping of methyl, methylene and side-chain amide protons to prevent the energy minimisation from getting trapped in local energy minima) and refinement.

In each round of CNS structure calculations, 100 structures for FH-12-13 were calculated and ranked according to overall energy and NOE energy and plotted *versus* $E_{overall}$ and E_{NOE}. Converged structures were scanned for violated NOE distance restrains (NOE violations >0.5 Å). NOE cross-peaks responsible for such NOE violations were inspected and depending on the identification of peak artefacts (i.e. extensive peak overlap, embedment in noise signals, closeness to diagonal) the upper distance value derived from these NOE-cross peaks were adjusted.

Three iterative CNS calculation rounds with minimum adjustments yielded the 3-D structure of FH-12-13, which was subjected to a final CNS refinement step in water solvent. Water refinement was carried out using the RECOORD protocols (http://www.ebi.ac.uk/msd/recoord); A 7 Å water shell was created around the protein and a molecular dynamics force field, including electrostatic and Lennard-Jones non-bonded potentials, was employed to mimic interaction with the solvent.

2.3.14 Molecule visualisation programs

Representations of protein structures were made in the molecular visualisation programs MOLMOL 2k.2 (140) and PyMol (141). Both programs were also used for the calculation of backbone root mean square deviation (RMSD) values. Electrostatic surface representations were calculated and displayed using GRASP (142).

2.4 Analytical ultracentrifugation

All studies were performed (20 ± 0.5 °C) in a Beckman XL-A analytical ultracentrifuge. For sedimentation velocity analysis cells were loaded with 400 μl of sample or appropriate buffer solution, and the rotor accelerated to 45,000 or 50,000 rpm. Radial scans at 273 nm were taken every two minutes and data analysed using the program Sedfit (143) to compute c(s) distributions. Resolution was set at 150, with an F (confidence) factor at 0.68. Over a series of fits an average value for the frictional ratios was determined: this was used as default for all fitting, to minimize artifactual variations. The baseline, the meniscus and cell base radial positions were floated in the fitting. The final profile was analyzed using software pro Fit™ (Quantum Soft, Zürich). A value of the partial specific volume (Vbar) (in ml/g) was computed via the program SEDNTERP (144), which was also employed to compute the density and viscosity properties of the buffer solution.

CHAPTER 3

PROTEIN PRODUCTION AND CHARACTERISATION

3.1 Introduction and overview of construct

A variety of different FH constructs covering the different putative functional sites of FH (as described in the Introduction) have been prepared for this study. Figure 3.1 illustrates these constructs in terms of their CCP module-composition and their position within the FH protein.

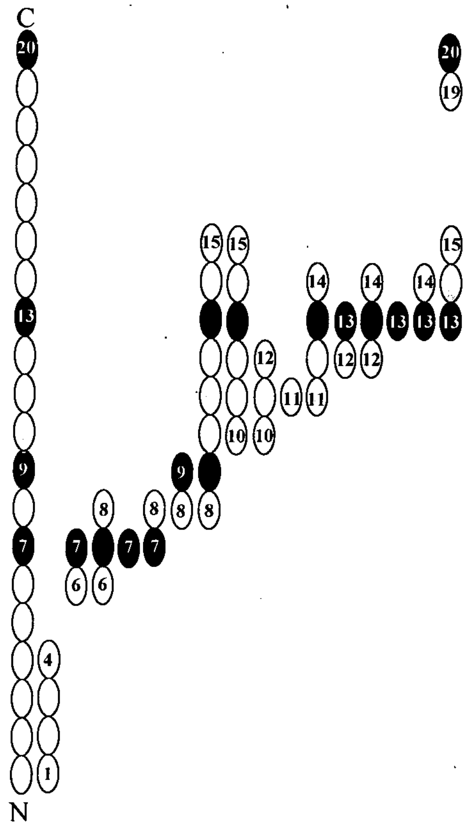


Figure 3.1 Summary of FH segments employed in this study[10]. A schematic representation of the FH molecule is shown on the far left with CCP modules numbered 1-20; to its right are cartoons

---

[10] FH-1-4 was cloned by Dr. David Kavanagh (148). FH-19-20 was prepared (cloned, expressed, purified) by Dr Andrew Herbert (51). The DNA preparation prior to transformation into *P. pastoris* of constructs FH-7-8 and FH-6-8 was carried out by Claire Egan. The DNA for FH-7_{H402}, FH-13 and FH-12-14 was cloned by Claire Egan. Batches of FH-13 protein were used which had been expressed and purified by Claire Egan or by the author of this thesis. All those constructs that included CCP module 7 were prepared in both, with and without the at-risk polymorphism for AMD at position 402 (residues numbered according to their its position within intact FH prior to cleavage of the signal

of the constructs prepared for the current study. Modules previously implicated in GAG binding are shaded in black.

Table **3.1** summarises N-terminal and C-terminal non-native sequences of all constructs and specifies the range of native FH residues within each construct by sequence number.

Construct	***Non-native sequence*** (N-terminal)	***Factor H residues***	***Non-native sequence*** (C-terminal)
FH-1-4	AGEQKLISEEDL	19-263	HHHHHH
FH-6-8$_{Y402}$, FH 6-8$_{H402}$	AG	322-508	-
FH-6-7$_{Y402}$, FH-6-7$_{H402}$	AG	324-443	-
FH-7$_{Y402}$, FH-7$_{H402}$	AG	386-445	-
FH-7-8$_{Y402}$, FH-7-8$_{Y402}$	AG	386-508	-
FH-8-9	AG	447-565	-
FH-8-9 $_{K446}$	A	446-565	-
FH-rr8-9	EFTWPSRPSRIGT	446-566	-
FH-8-15	AG	447-927	-
FH-10-12	AG	568-745	-
FH-10-15	AG	568-927	-
FH-11	AG	629-685	-
FH-11-14	AG	629-865	-
FH-12-13	AG	690-804	-
FH-12-14	AG	690-865	-
FH-13	AG	752-804	-
FH-13-15	AG	752-927	-

---

sequence - UniProtKB/Swiss-Prot entry P08603). The DNA preparation prior to transformation into *P. pastoris* of constructs FH-7-8 and FH-6-8 was carried out by Claire Egan. FH-7$_{H402}$, FH-13 and FH-12-14 was cloned by Claire Egan. Batches of FH-13 were used which had been expressed and purified by Claire Egan or by the author of this thesis. Two and three N-glycosylation sites present in constructs FH-13-14 and FH-12-14 (respectively) were knocked out by substituting asparagine residues for glutamines in N-glycocylation sites. FH-19-20 was prepared (cloned, expressed, purified) by Dr Andrew Herbert (51).

FH-19-20	AG	1107-1231	-

Table 3.1 Sequences of recombinant proteins used in this study. Numbering is on the basis of the encoded protein sequence including the 18 amino acid residues of the natural signal sequence.

Overview of protein production

The following sections describe cloning, expression, purification and characterisation of all FH segments prepared for this study. Unless stated otherwise, the coding sequences for all constructs were amplified from human FH cDNA that had been cloned into pUB/Bsd TOPO or pPICZαB. The appropriate DNA sequences were cloned into TOPO vector (for details see Figure 3.2). Sub-cloning into pPICZαB vector deviated from standard procedures (see CHAPTER3, Construct FH-6-7).

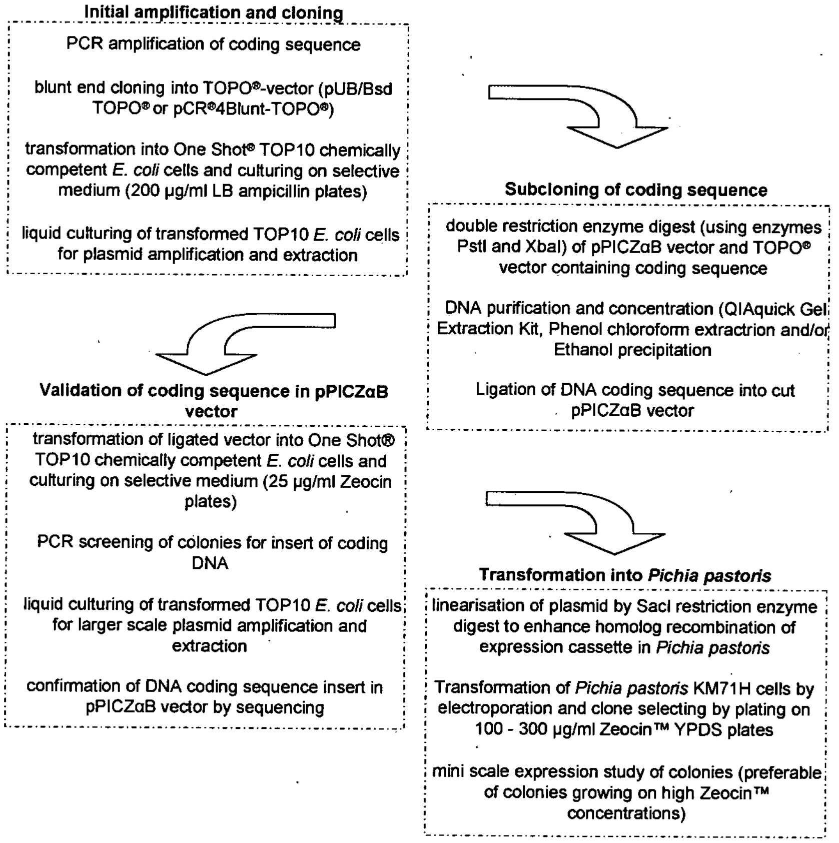

Initial amplification and cloning
PCR amplification of coding sequence
blunt end cloning into TOPO®-vector (pUB/Bsd TOPO® or pCR®4Blunt-TOPO®)
transformation into One Shot® TOP10 chemically competent E. coli cells and culturing on selective medium (200 µg/ml LB ampicillin plates)
liquid culturing of transformed TOP10 E. coli cells for plasmid amplification and extraction
Subcloning of coding sequence
double restriction enzyme digest (using enzymes PstI and XbaI) of pPICZαB vector and TOPO® vector containing coding sequence
DNA purification and concentration (QIAquick Gel Extraction Kit, Phenol chloroform extractrion and/or Ethanol precipitation
Ligation of DNA coding sequence into cut pPICZαB vector
Validation of coding sequence in pPICZαB vector
transformation of ligated vector into One Shot® TOP10 chemically competent E. coli cells and culturing on selective medium (25 µg/ml Zeocin plates)
PCR screening of colonies for insert of coding DNA
liquid culturing of transformed TOP10 E. coli cells for larger scale plasmid amplification and extraction
confirmation of DNA coding sequence insert in pPICZαB vector by sequencing
Transformation into Pichia pastoris
linearisation of plasmid by SacI restriction enzyme digest to enhance homolog recombination of expression cassette in Pichia pastoris
Transformation of Pichia pastoris KM71H cells by electroporation and clone selecting by plating on 100 - 300 µg/ml Zeocin™ YPDS plates
mini scale expression study of colonies (preferable of colonies growing on high Zeocin™ concentrations)

Figure 3.2 Flowchart showing steps involved in the generation of genetically modified *P. pastoris* clones for recombinant protein expression.

Successfully transformed *P. pastoris* (strain KM71H) clones were tested for protein expression in mini-scale trials. The best clones were selected for expression in shaker flasks or fermentor. Protein was harvested from supernatant either by concentration or batch absorption. Protein was purified using one or two of the following types of column chromatography: size-exclusion, ion exchange, Ni^{2+}-affinity and heparin-affinity. All of the standard protocols are described in CHAPTER 2 while specific parameters are reported on a case by case basis below. A summary of the protein production undertaken for the current study is presented in Table 3.2.

Protein X	**Starting material**	**Mini-scale**	**Shaker flask/ Fermentor**	**Endo Hf?**	**Pre-purification**	**Purification step 1**	**Purification step 2**	**Sequence validation**	**Fold validation**	**Notes**
FH-1-4	Overexpressing clone [1,2]		Fermentor	No	None	Nickel affinity[3]	Size exclusion[4] (Figure 3.4)	Accurate mass determination by MS[5] (Figure 3.5)	Cofactor assay	
FH-6-7$_{Y402}$	FH DNA[6]	Figure 3.8		No						
FH-6-7$_{H402}$	FH-6-8$_{H402}$ DNA in pPICzαB	Figure 3.8		No						
FH-6-8$_{Y402}$	Site-directed mutagenesis	Figure 3.9	Expressed by A. P. Herbert (APH)	No					^{15}N HSQC recorded by APH	Taken forward by APH
FH-6-8$_{H402}$	DNA sub-cloned in pPICzαB	Figure 3.9	Shaker flask (Figure 3.10)	No	Heparin sepharose[7]	Cation exchange (CE) (Poros) (Figure 3.10)	None	MS-tryptic fingerpr. (Figure 3.11)	^{15}N HSQC recorded by APH	
FH-7$_{Y402}$	FH-7$_{Y402}$ DNA in pPICzαB	Figure 3.12	Shaker flask (Figure 3.12)	No	SP-sepharose[7]	CE (Mono-S) (Figure 3.12)	None		Structure determined by APH	
FH-7$_{H402}$	Overexpressing clone		Expressed by APH	No					Structure determined by APH	
FH-7-8$_{Y402}$	FH-7-8$_{Y402}$ DNA in pPICzαB	Figure 3.13	Shaker flask (Figure 3.13)	No	Heparin Sepharose	CE (Poros) (Figure 3.14)		MS-tryptic fingerpr. (Figure 3.15)	^{15}N HSQC recorded by B. Blaum	
FH-7-8$_{H402}$	FH-7-8$_{H402}$ DNA in pPICzαB	Figure 3.13	Expressed by APH	No						
FH-8-9	FH DNA	Figure 3.16	Shaker flask + ^{15}N fermentor)	Yes	None	Anion exchange (AE) (CaptoQ) (Figure 3.18)	Size exclusion (Figure 3.19)	MS-tryptic fingerpr. (Figure 3.20)	^{15}N HSQC (Figure 3.20)	K446 version also made -
FH-$_{RR}$8-9	FH DNA	Figure 3.21		Yes	None	AE (CaptoQ) (Figure 3.22)	Size exclusion (Figure 3.22)	MS-tryptic fingerpr. (Figure 3.23)		Different cloning procedure –
FH-8-15	Overexpress-ing clone		Shaker flask	Yes	None	AE (MonoQ) (Figure 3.24)	Size exclusion (Figure 3.25)	MS-tryptic fingerpr. (Figure 3.26)	1D 1H-NMR (Figure 3.26)	
FH-10-12	FH DNA	Figure 3.27	Shaker flask	Yes	None	AE (MonoQ)	Size exclusion	MS-tryptic fingerpr.	1D 1H-NMR	Some sugar

						(Figure 3.28)	(Figure 3.29)	(Figure 3.30)	(Figure 3.30)	contamination
FH-10-15	FH DNA	Figure 3.31	Shaker flask (Figure 3.31) + fermentor	Yes	None	(Figure 3.32)	AE (Mono Q) (Figure 3.33)	MS-tryptic fingerpr. (Figure 3.34)	1D ^{1}H-NMR (Figure 3.34)	Tendency to oligomerise
FH-11	FH DNA	Figure 3.35		No						Forms dimers?
FH-11-14	FH DNA	Figure 3.36	Shaker flask (Figure 3.36)	Yes	None	Size exclusion (Figure 3.37)	AE (Mono Q) (Figure 3.38)	MS-tryptic fingerpr. (Figure 3.39)	1D ^{1}H-NMR (Figure 3.39)	Tendency to oligomerise
FH-12	FH DNA	Figure 3.40		Yes						
FH-12-13	FH DNA	Figure 3.41	Shaker flask (Figure 3.41) + ^{15}N and ^{15}N,^{13}C ferm.	Yes	Heparin Sepharose	Heparin-affinity (Poros)		MS-tryptic fingerpr. (Figure 3.43); Accurate MS	Structure determined	
FH-13-14	CHAPTER 3	Figure 3.45	Shaker flask	Yes	Heparin Sepharose	Size exclusion (Figure 3.46)		MS-tryptic fingerpr. (Figure 3.47)		Strong tendency to oligomerise
FH-13-14 Q802,Q822	CHAPTER 3	Figure 3.44								As above
FH-13	Overexpressing clone		Shaker flask (Figure 3.48)	Yes	SP-Sepharose	CE (Mono S) (Figure 3.48)			Structure determined by C. Fenton	
FH-13-15	FH DNA	Figure 3.49	Shaker flask	Yes		Heparin (HiTrap) (Figure 3.50)	Size exclusion (Figure 3.50)	MS-tryptic fingerpr. (Figure 3.51)	1D ^{1}H-NMR (Figure 3.51)	
FH-14	FH DNA	Figure 3.52								Very low or no expression

Table 3.2 Overview of recombinantly expressed proteins

[1] All of the recombinant proteins in this study were expressed in *P. pastoris* strain KM71H

[2] N-terminal c-myc tag; C-terminal hexa-His tag

[3] XK16/20 column loaded with IMAC Sepharose charged with Ni^{2+} (Amersham Biosciences)

[4] Hiload Superdex 75 prep grade size-exclusion column (16 mm x 600 mm) Amersham Biosciences

[5] By liquid chromatography-(12-Tesla) Fourier-transform ion cyclotron resonance mass spectrometry

[6] ”FH DNA” = FH-coding sequence amplified from cDNA and sub-cloned into either pUB/Bsd TOPO vector or pPICzαB

[7] “Fast Flow Sepharose” Amersham Biosciences

3.1.1 Construct FH-1-4

A *P. pastoris* KM71H clone over-expressing FH-1-4 was cloned by Dr David Kavanagh. The construct was designed with an N-terminal c-myc tag and a C-terminal *hexa*-His tag for ease of purification. The sequence was partially validated by peptide mass fingerprinting, as described in (145). The N-terminal c-myc tag allowed Western-blot screening for this construct using commercially available polyclonal anti-c-myc tag antibodies (Cell Signalling Technology)

Protein from four-litre fermentation was initially purified on a Ni^{2+}-affinity column. . The elution profile showed a dominant protein peak at roughly one third of the imidazole gradient (corresponding to ~215 mM imidazole eluting from the column). Samples of 5 µl of the crude supernatant and of fractions corresponding to the elution peak were analysed by SDS-PAGE (Figure 3.3-A).

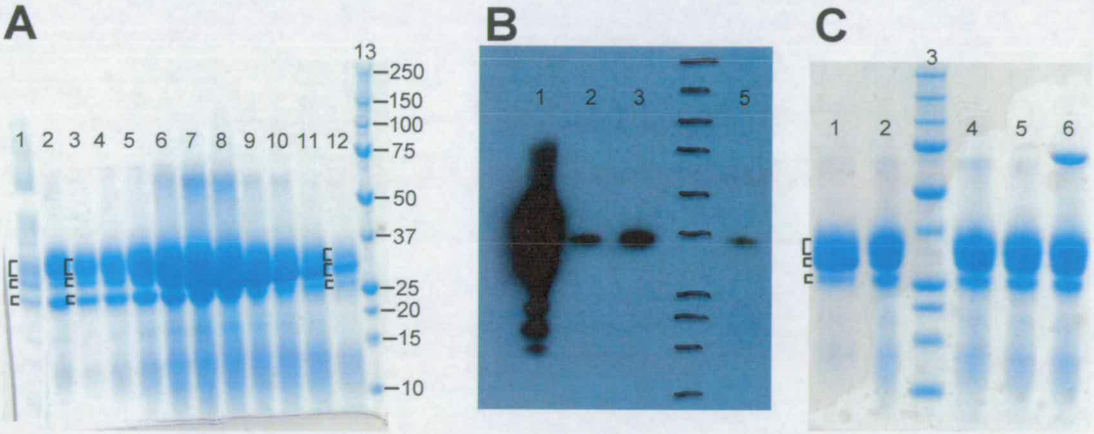


Figure 3.3 Gradient SDS-PAGE of FH-1-4 following fermentor-expression and purification. (A) Each lane was loaded with 5 µl of protein sample mixed with reducing loading buffer. Lane 1: crude fermentor supernatant. Lanes 2-12: samples corresponding to the protein peak eluting at ~215 mM imidazole. Lane 13: Protein ladder with molecular-weight (kDa) standards as indicated. Note that fractions corresponding to the middle of the peak are overloaded (B) Western blot performed with primary anti-c-myc tag antibodies. Lane 1 was overloaded with FH-1-4 corresponding to sample of lane 7 from gel in (A); lanes 2, 3 and 5 contain shaker-flask

expressed FH-1-4 variants[11] - I62V, R53H and R78G - for comparison. (C) FH-1-4 corresponding to sample of lane 7 from gel in (A): flow-through after application to ConA-lectin column (lane 1); untreated (lane 2); treated with 500 U of PNGase F (lane 4); treated with 500 U of Endo H_f (70 kDa) (lane 5); treated with about 1000 U of Endo H_f (lane 6); molecular weight markers are shown in lane 3.

The gel revealed three dominant bands between molecular weight standards of 25 and 37 kDa. The higher molecular weight band is broad and diffuse, the one in the middle is sharp and corresponds to the expected molecular weight of FH-1-4 (~30.0 kDa) and the third band is probably a degradation product. A sample from the Ni^{2+}-affinity elution was analysed by Western blot (anti-c-myc tag detection) to confirm that the second band corresponds to FH-1-4. Although the sample (lane 1- Figure 3.3-B) was too concentrated, it is still apparent that both the diffuse upper band and the "degradation band" correspond to proteins containing a c-myc tag, as does the band corresponding to the expected molecular weight for FH-1-4of ~30 kDa. Detection by the antibody of the diffuse upper band could be explained, if it is a glycosylation product of the FH-1-4 construct. N-linked and O-linked high-mannose containing glycosylation is a well known post-translational modification observed in the *P. pastoris* expression system (146, 147). Although an N-linked glycosylation site is present in FH-1-4 (N_{217}-G_{218}-S_{219}), this was found not to be glycosylated in shaker-flask expressions of the same clone (148). It could not be ruled out, however, that glycosylation might occur under fermentation conditions. To investigate if the diffuse, upper band does indeed derive from glycosylation, the FH-1-4 sample was applied to a column packed with convallatoxin A lectin (149) that retards molecules containing high-mannose carbohydrate structures (Figure 3.3-C). A small proportion was also mixed and incubated with different, commercially available endoglycosidase enzymes that trim N-linked glycans (Figure 3.3-C). Neither treatment, however, resulted in any depletion of the diffuse band; therefore, no conclusions regarding the nature of this band could be drawn.

---

[11] The FH-1-4 variants were prepared by Isabel Pechtl.

The SDS-PAGE gel analysis of fractions following Ni^{2+}-affinity chromatography shows that the proportion of the ~30 kDa protein begins to dominate towards the last third of the elution peak. This partial resolution was exploited by re-application of the relevant fractions onto the Ni^{2+}-affinity column followed by elution with 40 mM imidazole-containing buffer over eight column volumes. This procedure resulted in an elution peak with a long tail. While the diffuse and degradation bands predominated in the early fractions of this peak, the desired species (~30 kDa) is clearly enriched in the later fractions (Figure 3.4-A). The late-eluting fractions were pooled, concentrated and applied to gel-filtration chromatography (Figure 3.4-B). A small peak eluted at ~48 ml just before the large main peak (at 56 ml). Fractions spanning the two peaks were analysed on gradient SDS-PAGE. Fractions corresponding to the earlier peak contain the diffuse and degradation bands whereas fractions collected under the main peak contain almost exclusively the target material.

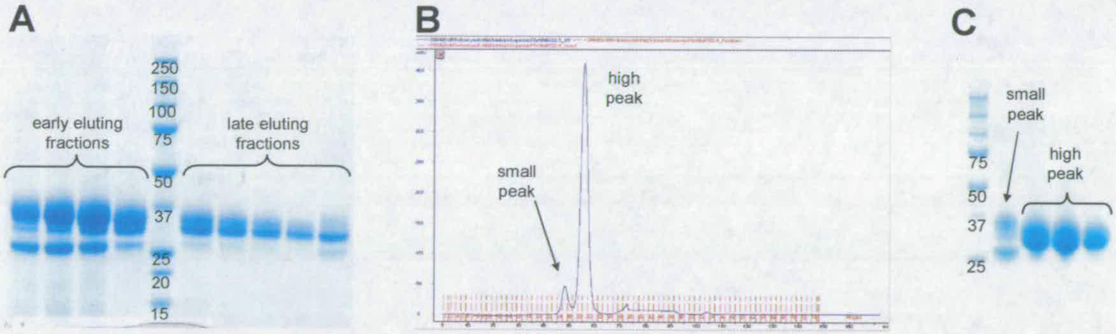


Figure 3.4 Purification of FH-1-4; (A) Gradient SDS-PAGE of fractions following elution of the FH-1-4-containing material from a nickel-affinity column using 40 mM imidazole. (B) Gel filtration of late-eluting fractions from the nickel-affinity elution profiled in (A). (C) Gradient SDS-PAGE of fractions corresponding to the two peaks in the chromatogram of (B). The small peak eluting before the main peak contains the diffuse upper band and degradation product while the larger peak corresponds mostly to the target material.

Fractions corresponding to the major peak in the gel-filtration chromatogram were pooled and submitted for N-terminal sequencing, mass spectrometry (Figure 3.5) and an assay of factor I co-factor activity (see Methods) (Figure 3.6).

N-terminal sequencing reported the amino acids "EQKLIS" that may be compared to the expected N-terminal amino acid sequence "AGEQKLIS", where AG corresponds to the cloning artefact and "EQKLIS" is the initial sequence of the c-myc tag. Such a finding is consistent with several literature reports of N-terminal variations in *P. pastoris*-expressed proteins that are fused with the secretion signal sequence of the *S. cerevisiae* alpha–factor. These arise from aberrant processing of the signal sequence by dibasic endopeptidase Kex2, and dipeptidyl amino peptidase Ste13 (150-159). Another explanation for the missing AG could be trimming of N-terminal sequences by non-specific proteases that are present at low levels in culture media (160). Taking the N-terminal truncation of AG into account, the molecular weight of FH-1-4 is predicted to be 29663.2 Da. Liquid chromatography-mass spectrometry analysis (Figure 3.5) revealed one major species (A: 29663 Da) corresponding precisely to the predicted mass, and three minor species of slightly higher molecular weight (B: 29791.8 Da, C: 29826.2 Da and D: 29950 Da). The difference between peaks A and B is 128.8 Da. Therefore peak B reflects a species that still retains the glycine of the cloning artefact dipeptide "AG" accounting for a theoretical mass of 128 Da. Differences between peaks C-A, and D-B are 163.2 Da and 158.2 Da, respectively. Species C and D most likely correspond to species A and B that bear one hexose sugar unit; expected molecular mass for such a modification is 162 Da. A possible explanation could be non-specific linkage of recombinantly expressed protein with sugars present in *P. pastoris* culture medium (Maillard-reaction).

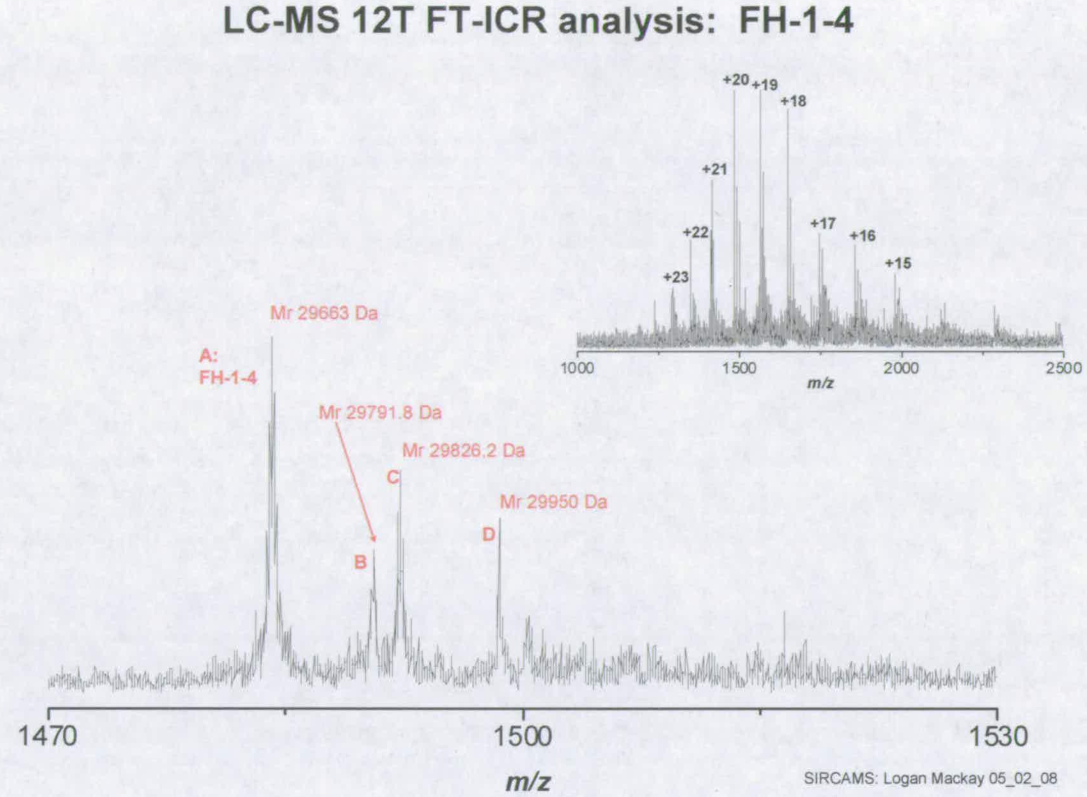


Figure 3.5 Mass spectrometric analysis of FH-1-4; Major peak A corresponds exactly to the predicted mass of FH-1-4 of 29663 Da.

The purified FH-1-4 was submitted for an assay of cofactor activity with respect to cleavage of C3b by Factor I to investigate functional activity of the *P. pastoris*-expressed construct (Figure 3.6).

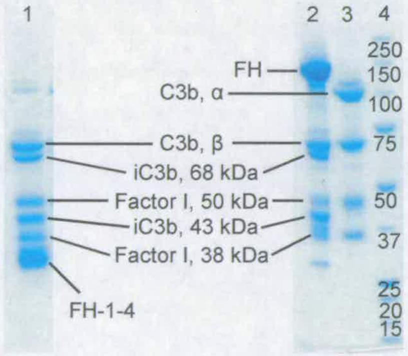

1
2
3
4
FH
C3b, α
C3b, β
iC3b, 68 kDa
Factor I, 50 kDa
iC3b, 43 kDa
Factor I, 38 kDa
FH-1-4
250
150
100
75
50
37
25
20
15

Figure 3.6 Cofactor activity assay of FH-1-4 for factor I-mediated cleavage of C3b (see Methods). FH-1-4 was loaded in lane 1 of a SDS-polyacrylamide gel. The positive control (FH)

and the negative control (C3b and factor I only) were loaded in lanes 2 and 3, respectively. Lane 4 shows molecular weight markers (kDa).

Cofactor activity could be demonstrated for the FH-1-4 expressed in this study. This is evident from the presence of the 68 and 43 kDa cleavage products on the gel (Figure 3.6) This is consistent with literature reports (58),(59) that the cofactor activity of FH is present within its four N-terminal CCP modules.

3.1.2 Constructs FH-6-7$_{Y402}$ and FH-6-7$_{H402}$

Due to difficulties experienced with the standard sub-cloning methodology modifications were instituted. The outline of the initial cloning steps are summarised in Figure 3.2.

"Quick" ligation reactions (New England Biolabs Ltd., Hitchin, UK), without performing a gel purification after restriction-enzyme digestion, were carried out. Simple re-ligation of the picaαB vector, without incorporation of the target construct, is a possibility. In order to minimise this, a large excess of cut FH-6-7 DNA was used for the ligation reaction. A PCR-based screen (shown for FH-6-7$_{H402}$ and FH-6-7$_{Y402}$ in Figure 3.7-A and Figure 3.7-B, respectively) was performed on the colonies, which were grown overnight. Forward and reverse primers, sandwiching the cloning site, were used for the PCR screening. The expected bop-length of relegated picaαB vector is 330 bp, whereas the expected bp-length of the pPICZαB vector containing FH-6-7 is ~ 620 bp. Two out of nine, and five out of five, screened colonies showed a DNA band of the right size for FH-6-7$_{H402}$, and FH-6-7$_{Y402}$, respectively.

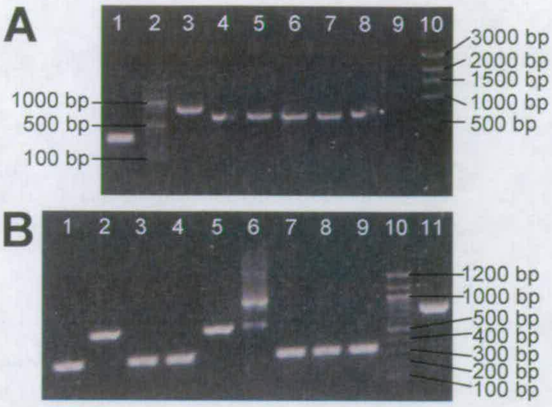

A
1 2 3 4 5 6 7 8 9 10
1000 bp
500 bp
100 bp
3000 bp
2000 bp
1500 bp
1000 bp
500 bp
B
1 2 3 4 5 6 7 8 9 10 11
1200 bp
1000 bp
500 bp
400 bp
300 bp
200 bp
100 bp

Figure 3.7 PCR screen of *E. coli* colonies transformed with the ligation product of the pPICZαB sub-cloning step; (A) control: the 330-bp band in lane 1 corresponds to "empty" pPICZαB vector; 100-bp ladder (lane 2); 800-bp band of positive control (lane 3); 620-bp bands of vector containing FH-6-7$_{\text{H402}}$ (lanes 4 to 8); empty (lane 9); 1-kb ladder (lane 10); (B) the five bands of 330-bp correspond to re-ligated vectors (lane 1, 3, 4, 7 and 8); the two bands of 620-bp correspond to vector containing FH-6-7$_{\text{Y402}}$ (lane 2 and 5); diffuse bands (lane 6); control: 330-bp band corresponds to "empty" pPICZαB vector (lane 9); 100-bp ladder (lane 10); 800-bp band of positive control (lane 11).

Colonies showing a DNA band at the expected size were selected for plasmid extraction. Confirmation of the successful ligation was achieved by sequencing. After transformation, a small-scale expression test was performed on those *P. pastoris* colonies that grew on YPDS plates containing Zeocin (300 µg/ml) (Figure 3.8).

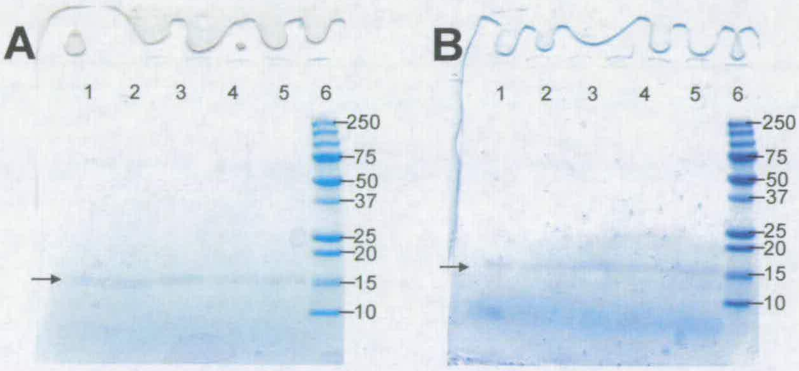

A
1 2 3 4 5 6
250
75
50
37
25
20
15
10
B
1 2 3 4 5 6
250
75
50
37
25
20
15
10

Figure 3.8 Mini-scale expression study of FH-6-7 (A) Gradient SDS-PAGE of mini-scale expression study of five *P. pastoris* clones expressing FH-6-7$_{\text{Y402}}$; an arrow indicates the protein

bands at the expected mobility; molecular weight markers (kDa) as indicated. (B) Gradient SDS-PAGE as in (A, but for construct FH-6-7$_{\mathrm{H402}}$.

3.1.3 Constructs FH-6-8$_{\mathrm{Y402}}$ and FH-6-8$_{\mathrm{H402}}$

The coding DNA for FH-6-8$_{\mathrm{H402}}$ sub-cloned into *P. pastoris* expression pPICZαB vector was provided by Claire Egan. FH-6-8$_{\mathrm{Y402}}$ was obtained by changing the 402-H coding sequence of the construct FH-6-8$_{\mathrm{H402}}$ (in pPICZαB vector) to the 402-Y coding sequence using the QuickChange™ mutagenesis kit (Stratagene®). Successful amplification of FH-6-8$_{\mathrm{Y402}}$ in the QuickChange™-PCR reaction was monitored by agarose gel analysis (Figure 3.9 A).

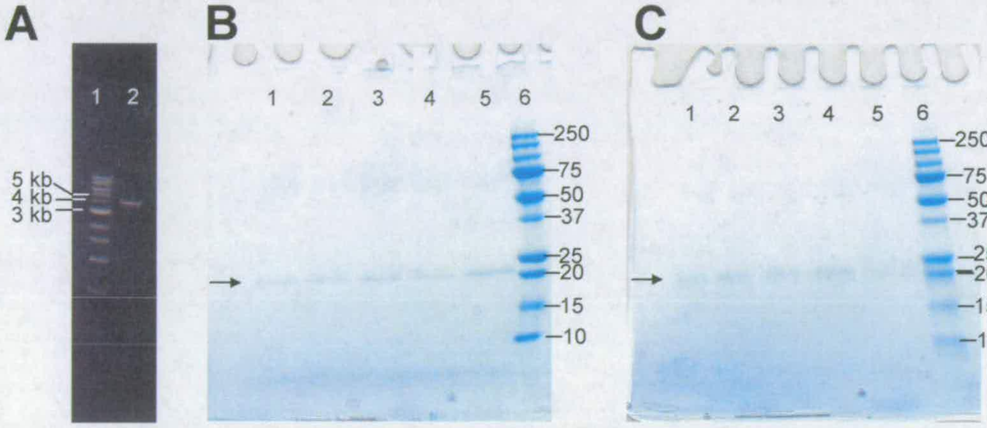


Figure 3.9 Plasmid DNA preparation of construct FH-6-8$_{\mathrm{Y402}}$ and mini-scale expression study of *P. pastoris* clones FH-6-8$_{\mathrm{H402}}$ and FH-6-8$_{\mathrm{Y402}}$; (A) Agarose gel shows QuickChange-PCR product (FH-6-8$_{\mathrm{Y402}}$ in the pPICZαB vector) at the expected size of 4.1 kb (pPICZαB vector: 3.6 kb; FH-6-8$_{\mathrm{Y402}}$ insert: 540 bp). (B) Gradient SDS-PAGE of the mini-scale expression study; an arrow indicates the protein bands at the expected mobility for FH-6-8$_{\mathrm{H402}}$; molecular weight markers (kDa) are indicated. (C) Gradient SDS-PAGE as in B) but for construct FH-6-8$_{\mathrm{Y402}}$.

After insertion of the expression cassettes into competent *P. pastoris* KM71H cells, a mini-scale expression study was performed on colonies growing on YPDS plates containing 300 µg/ml Zeocin (Figure 3.9-B and -C).

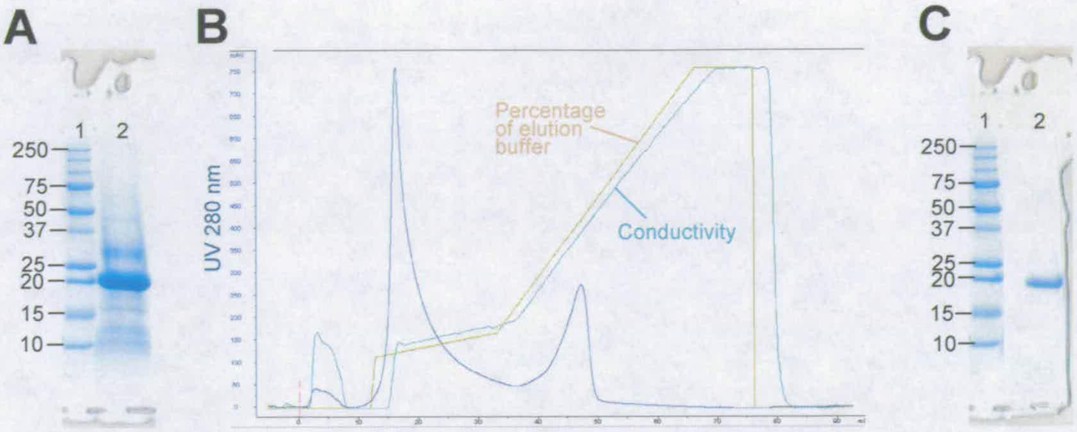
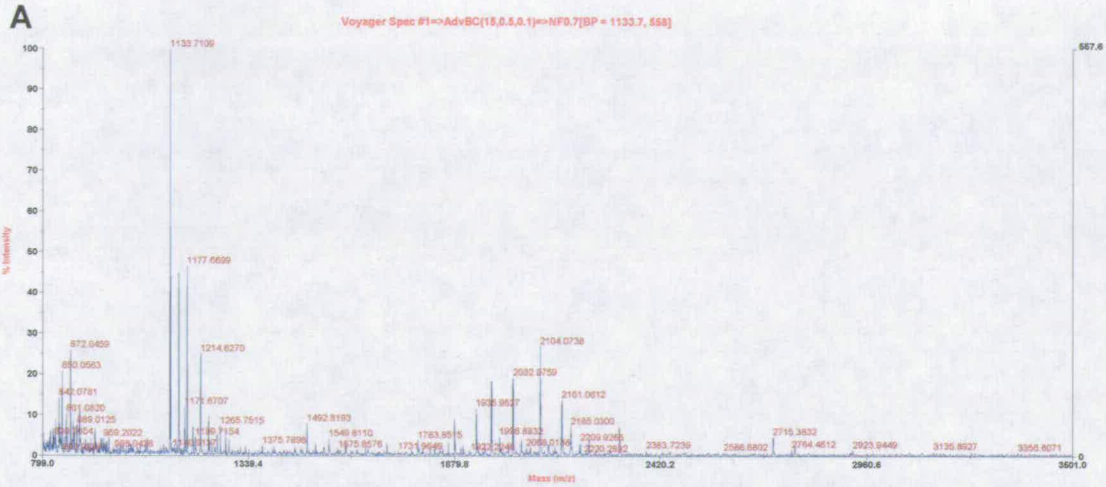


Figure 3.10 Expression and purification of FH-6-8$_{H402}$; (A) Gradient SDS-PAGE of concentrated supernatant of shaker flask expression (lane 2). (B) Pre-purified FH-6-8$_{H402}$ was buffer exchanged into equilibration buffer, mixed with a protease inhibitor cocktail (which was purchased from Sigma) and loaded onto the POROS 20S cation-exchange column equilibrated with 20 mM potassium phosphate buffer (pH 6.0) and eluted in three gradient steps to 1 M sodium chloride. FH-6-8$_{H402}$ was eluted at about 50% concentration of elution buffer. (C) Gradient SDS-PAGE of the FH-6-8$_{H402}$ peak fraction (lane 2).

FH-6-8$_{H402}$ expressed in shaker flasks was harvested by batch absorption onto heparin sepharose at pH 5.0. It was eluted in 1 M salt, diluted and adjusted to pH 6.0 and then subjected to cation-exchange chromatography (Figure 3.10). SDS-PAGE revealed a single band under reducing and non-reducing conditions. Note that slightly increased mobility occurs under non-reducing conditions, consistent with disulfide formation (86).

Partial validation of primary structure was obtained from a tryptic-digest MALDI-TOF mass spectrometry-based peptide fingerprint analysis (Figure 3.11).



B

```
Sequence coverage: 36% (69 of 189 amino acids)
1          11         21         31         41         51         61         71
AGLKPCDYPD IKHGGLYHEN MRRPYFPVAV GKYYSYYCDE HFETPSGSYW DHIHCTQDGW SPAVPCLRKC YFPYLENGYN
81         91         101        111        121        131        141        151
QNHGRKFVQG KSIDVACHPG YALPKAQTTV TCMENGWSPT PRCIRVKTCS KSSIDIENGF ISESQYTYAL KEKAKYQCKL
161        171        181
GYVTADGEPS GSITCGKDGW SAQPTCIKS
```

Figure 3.11 (A) Mass spectrum of tryptic-peptide fingerprint of FH-6-8$_{H402}$ (B) Sequence coverage in peptide-mass fingerprinting. The amino acid sequence of FH-6-8$_{H402}$ is shown; highlighted in red are matched peptide masses, and underlined are trypsin cleavage points.

3.1.4 Constructs FH-7$_{Y402}$ and FH-7$_{H402}$

The KM71H *P. pastoris* clone FH-7$_{H402}$ was prepared by Claire Egan. Coding DNA for FH-7$_{Y402}$ sub-cloned into *P. pastoris* expression pPICZαB vector was provided by Claire Egan, too. This vector was transformed into *P. pastoris*. Colonies that grew on YPDS plates containing 300 µg/ml Zeocin were subjected to a mini-scale expression study (Figure 3.12-A).

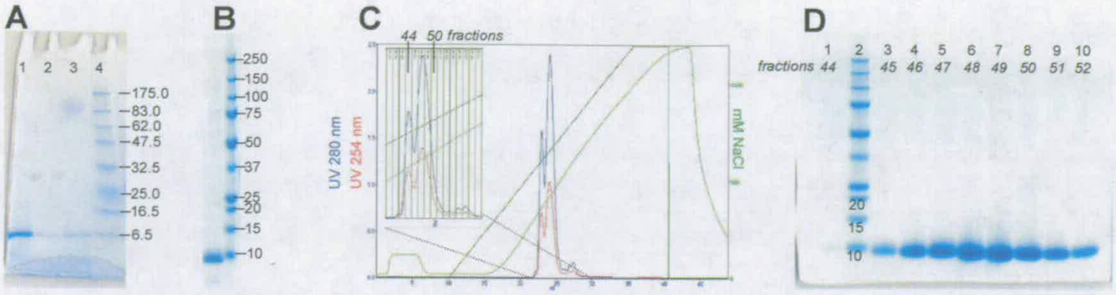


Figure 3.12 Expression and purification of FH-7$_{Y402}$. (A) Gradient SDS-PAGE from mini-scale expression trial. (B) Shaker flask expression of colony lane 1 (of A) with highest expression level. (C) Pre-purified (SP-Sepharose) FH-7$_{Y402}$ was buffer-exchanged into 20 mM sodium acetate pH 5.0 and loaded onto the Mono S column. Protein was eluted in a linear gradient to 1 M NaCl over 15 column volumes. (D) Gradient SDS-PAGE of fractions from the cation-exchange chromatography. Both peaks in the cation-exchange chromatogram correspond to bands with similar mobility on the gel.

Following shaker-flask expressions of FH-7$_{Y402}$ (Figure 3.12-B), the supernatant was diluted five-fold and the pH was adjusted to 5.0 for batch-absorption to SP-Sepharose. After elution (1 M NaCl) protein was subjected to cation exchange chromatography (Figure 3.12-C).

The elution profile shows two overlapping peaks. SDS-PAGE analysis of the fractions spanning both peaks show similar gel-mobility and integrity (as judged by the absence of degradation bands). This likely reflects N-terminal variations originating from differential processing of the *Saccharomyces cerevisiae* α–factor mating secretion signal. This commonly results in a ragged N-terminal sequence leaving parts of the Ste13 peptidase signal "EAEA" (157). The presence of either one or two "EA" repeats would result in different net-charges, and partial separation in ion-chromatography. The purity and integrity of recombinantly expressed protein was monitored by SDS-PAGE (Figure 3.12-D). SDS-PAGE under non-reducing conditions also yielded a single band with slightly increased mobility (not shown) compared to reducing conditions, as expected.

Protein expressed by this clone was used for the determination of the high resolution NMR structure of FH-7$_{Y402}$ by Dr Andrew Herbert (1). This published work obviously validates the integrity and structure of the recombinantly expressed protein.

3.1.5 Constructs FH-7-8$_{Y402}$ and FH-7-8$_{H402}$

The coding DNA sub-cloned into the *P. pastoris* expression vector pPICZαB was provided for both FH-7-8 constructs by Claire Egan. Following transformation of *P. pastoris* four FH-7-8$_{H402}$ clones and only one FH-7-8$_{Y402}$ clone grew on YPDS plates containing 300 µg/ml Zeocin, and these were subjected to mini-scale expression trials (Figure 3.13).

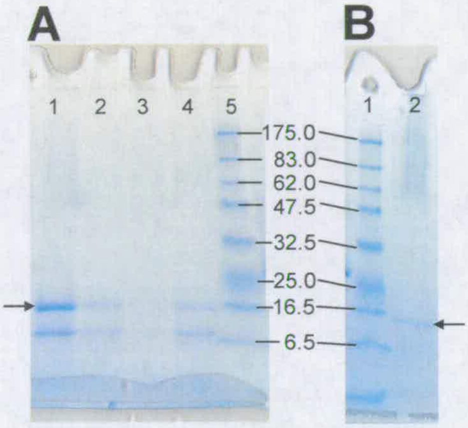


Figure 3.13 Mini-scale expression study of FH-7-8; (A) Gradient SDS-PAGE of four clones expressing FH-7-8$_{H402}$; an arrow indicates the protein bands at the expected mobility for FH-7-8$_{H402}$. (B) As in A) but for one colony of construct FH-7-8$_{Y402}$.

Following shaker-flask culture, supernatant (diluted five-fold, pH 5.0) was batch-absorbed onto heparin resin and eluted (1 M NaCl). SDS-PAGE (reducing conditions) of the eluted material showed the presence of some degradation, while SDS-PAGE under non-reducing conditions reveals a single band (Figure 3.14-A). A possible explanation of the degradation pattern observed would be a situation in which a certain portion molecules are 'clipped' but remain held together by one or more disulfide bonds.

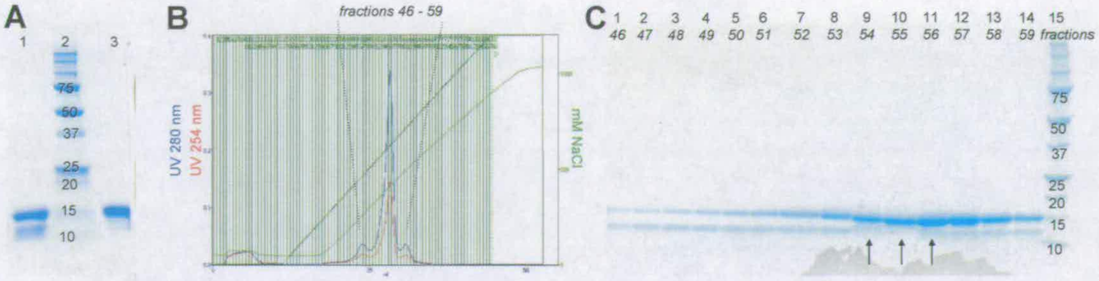
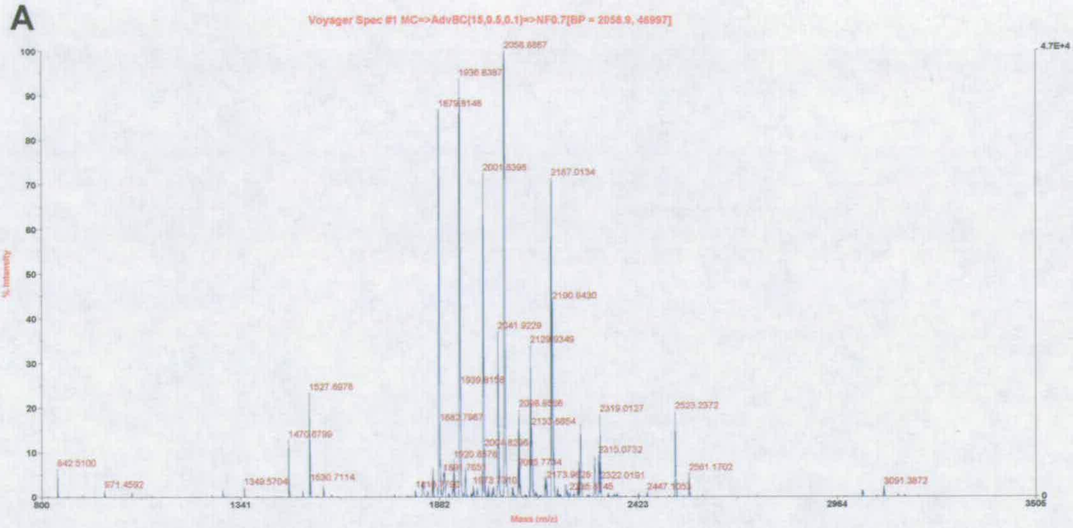


Figure 3.14 Expression and purification of FH-7-8$_{Y402}$; (A) Gradient SDS-PAGE of FH-7-8$_{Y402}$ (shaker flask expression) pre-purified on Heparin-Sepharose; reducing and non-reducing and gel loading conditions (lanes 1 and 3 respectively). (B) Heparin affinity chromatogram; pre-purified FH-7-8$_{Y402}$ was buffer-exchanged into 20 mM potassium phosphate buffer, loaded onto a Self Pack POROS 20 Heparin affinity column at pH 7.4 and eluted in a linear gradient to 1 M NaCl over 20 CV. (C) Gradient SDS-PAGE of fractions from the heparin-affinity chromatography shown in B). Lanes and fractions as indicated. Fractions highlighted by arrows correspond to the major peak.

The eluted material was flowed over 5 ml of Con-A lectin Fast Flow resin (Amersham Biosciences) which was pre-equilibrated with 20 mM sodium acetate buffer at pH 5.0. The flow through was collected and buffer-exchanged into 20 mM potassium phosphate buffer, pH 7.4, for heparin-affinity chromatography.

The chromatogram shows two smaller peaks flanking the main peak (Figure 3.14-B). SDS-PAGE (Figure 3.14-C) reveals that only the fractions corresponding to the main peak show a single band, and these were collected for downstream applications. However, to increase the yield fractions on either side were pooled, desalted and re-applied to the heparin- affinity column. Partial validation of the primary structure was obtained from a tryptic digest MALDI-TOF mass spectrometry peptide fingerprint analysis (Figure 3.15).



B

```
Sequence coverage: 66% (83 of 125 amino acids)
1          11         21         31         41         51         61         71
AGLRKCYFPY LENGYNQNYG RKFVQGKSID VACHPGYALP KAQTTVTCME NGWSPTPRCI RVKTCSKSSI DIENGFISES
81         91         101        111        121
QYTYALKEKA KYQCKLGYVT ADGETSGSIT CGKDGWSAQP TCIKS
```

Figure 3.15 (A) Mass spectrum of peptide fingerprint of construct FH-7-8$_{Y402}$ (B) Sequence coverage in peptide mass fingerprinting. The amino acid sequence of FH-7-8$_{Y402}$ is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points.

2-D NMR ^{1}H,^{15}N HSQC experiments (performed by Bärbel Blaum (3)) were consistent with correctly folded 3-D structure.

3.1.6 Constructs FH-8-9 and FH8-9$_{K446}$

The FH-8-9 coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (Figure 3.2).

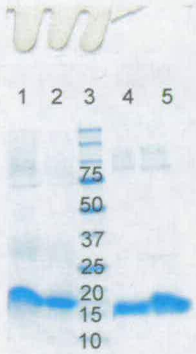


Figure 3.16 Gradient SDS-PAGE of mini-scale expression study of two *P. pastoris* clones expressing FH-8-9; samples run under reducing conditions and non-reducing conditions are left and right of the molecular weight marker, respectively.

Small-scale expression tests were performed for those colonies that grew on YPDS plates containing 300 µg/ml Zeocin (Figure 3.16); EndoH$_f$-treated samples were analysed by SDS-PAGE.

Batches of FH-8-9 were expressed (unlabelled) in shaker flasks, or ^{15}N-isotopically enriched in 1-litre scale fermentation. Figure 3.17 shows the log of fermentation parameters for the duration of the fermentation run (~ 116 h).

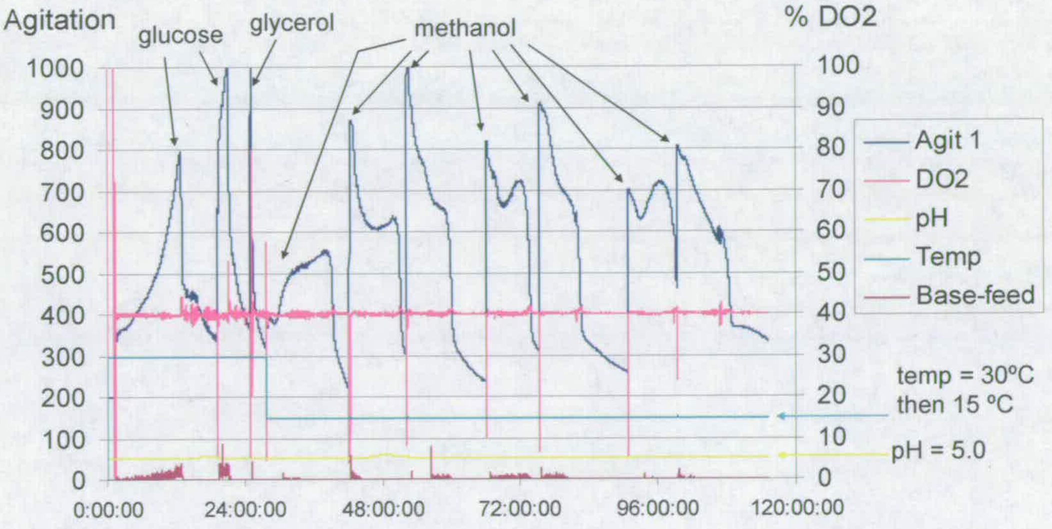


Figure 3.17 Log of ^{15}N-FH-8-9 fermentation. Control parameters shown are agitation (dark blue), relative percentage of dissolved oxygen (DO_2 – pink), temperature (cyan), pH (yellow) and

potassium hydroxide feed (purple). Black arrows indicate respiratory response of *P. pastoris* to feeding. Metabolism of energy sources (glucose, glycerol. methanol – metabolism peaks indicated by black arrows) leads to consumption of dissolved oxygen. In order to keep the DO_2 value set at 40%, agitation value is increased to enhance oxygen dissolution. $^{15}N\text{-}NH_4SO_4$ was used as nitrogen source and was present in the initial medium. Cell mass was grown at 30 °C on two glucose feeds, followed by a glycerol feed for de-repression of the alcohol oxidase promoter. For induction with methanol the temperature was lowered to 15 °C to increase stability of recombinantly expressed protein. *P. pastoris* metabolism acidifies the culture medium. This was balanced with potassium hydroxide feeds to maintain the pH at a constant value of 5.0.

The purification protocols for (unlabelled) shaker-flask expression and for ^{15}N-labelled fermentations, were identical. The deglycosylated supernatant was flowed over 5 ml of Con-A lectin resin, which was equilibrated with 20 mM potassium phosphate buffer at pH 7.4. The flow-through was diluted five-fold, adjusted to pH 7.0 and applied to an anion-exchange column (Figure 3.18).

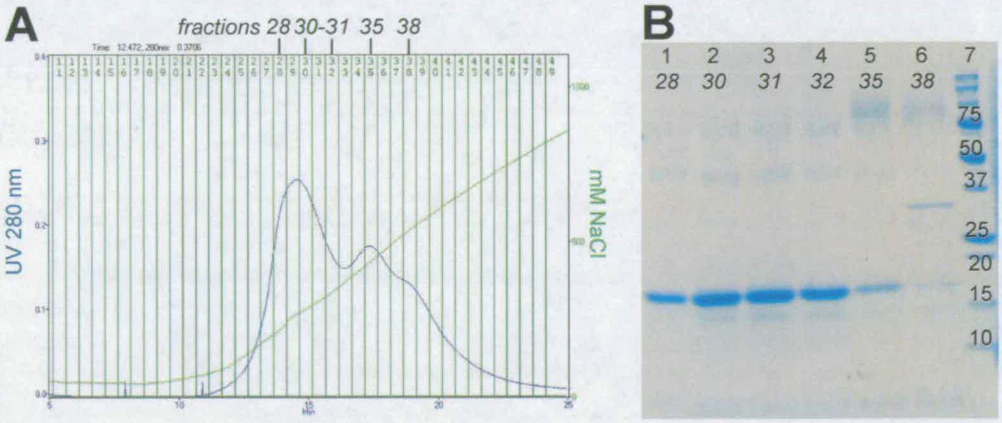


Figure 3.18 Initial purification of FH-8-9 (A) diluted and deglycosylated fermentor supernatant FH-8-9 was passed through a HiTrap CaptoQ column, equilibrated with 20 mM potassium phosphate at pH 7.0 and eluted in a linear gradient, over 20 CV, to 1 M NaCl. (B) Gradient SDS-PAGE of fractions indicated in A).

The elution profile shows two broad overlapping peaks. Fractions analysed by SDS-PAGE (Figure 3.18-B) indicated that the first peak - eluting at the beginning of the gradient - corresponds mainly to FH-8-9 (with a small proportion of degraded material present) contaminated with significant impurities originating from *P. pastoris* proteins. The second elution peak corresponds mostly to *P. pastoris* proteins. The elution of recombinantly expressed FH-8-9 at the beginning of the gradient, combined with the significant amounts of impurities present, suggests that the initial protein capturing and purification step (CaptoQ) is not ideal. However, the low pI[12] (4.7) of the FH-8-9 construct ruled out cation-exchange chromatography as a first step. Therefore care was taken to maximise yield from the CaptoQ column by reapplying the flow-thorough onto the column. Size-exclusion chromatography (in 20 mM potassium phosphate buffer, pH 7.4, 500 mM NaCl) was applied as a second purification step to separate *P. pastoris* proteins and degraded material from intact FH-8-9 protein. The high molecular weight *P. pastoris* proteins eluted early from the Superdex 75 column (Figure 3.19-A). Fractions representing the major peak at 78.9 ml were analysed on SDS-PAGE (Figure 3.19-B).

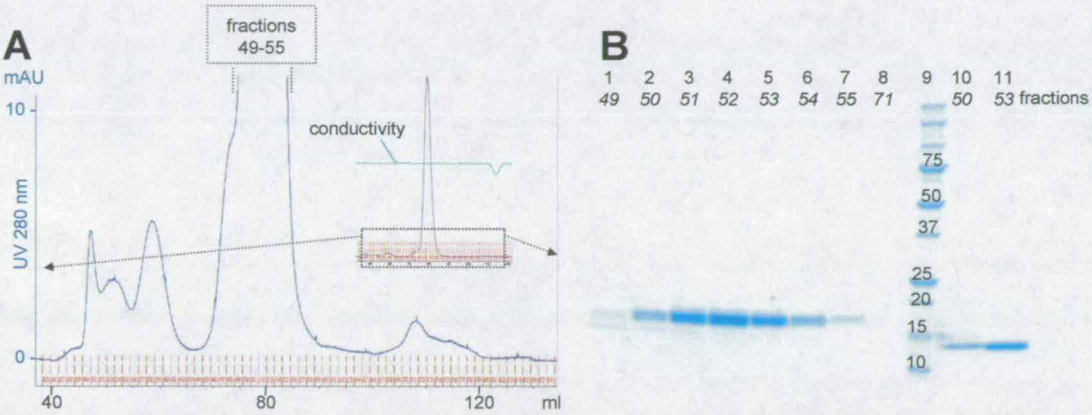


Figure 3.19 Second purification step for FH-8-9; (A) Concentrated and semi-purified FH-8-9 sample was loaded on a HiLoad Superdex 75 size-exclusion column. (B) Gradient SDS-PAGE of fractions from size exclusion-chromatography shown in A). Reducing and non-reducing conditions applied in lanes 1-9, and 10-11, respectively.

---

[12] Theoretical pI values were calculated by using the online server Expasy swiss prot protparam tool (115).

Fractions from the leading shoulder of the main peak correspond to degraded FH-8-9, while other fractions from this peak contain intact and pure material. Partial validation of primary structure was obtained from a tryptic digest MS (Figure 3.20).

A

```
Sequence coverage: 59% (71 of 119 amino acids)
1          11         21         31         41         51         61         71
TCSKSSIDIE NGFISESQYT YALKEKAKYQ CKLGYVTADG ETSGSITCGK DGWSAQPTCI KSCDIPVFMN ARTKNDFTWF
81         91         101        111
KLNDTLDYEC HDGYESNTGS TTGSIVCGYN GWSDLPICY
```

B

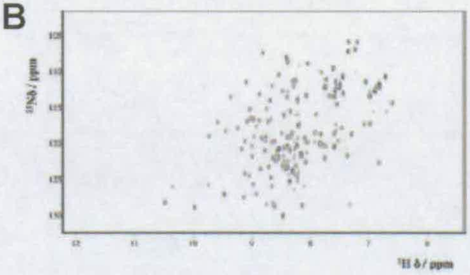


Figure 3.20 Validation of FH-8-9 (A) Sequence coverage in peptide mass fingerprinting for FH-8-9. The amino acid sequence of FH-8-9 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points. (for mass spectrum of peptide fingerprint of construct FH-8-9 Appendices) (B) ^{15}N-HSQC spectrum of FH-8-9. Spectrum shows good dispersion with well resolved peaks. NMR sample is 110 μM in 20 mM potassium phosphate buffer pH 6.6.

Lineshape and good peak dispersion in the 2-D NMR ^{1}H,^{15}N HSQC spectrum are consistent with a correctly folded 3-D structure (Figure 3.20-B).

For the preparation of construct FH-8-9$_{K446}$, which features a change of the non-native Gly from the cloning artefact into the native Lys at position 446, a QuickChange™ mutagenesis (Stratagene®) strategy was employed. A preparation of electrocompetent *P. pastoris* strain KM71H was transformed with the successfully produced expression cassette. High plasmid copy number-containing colonies were selected on YPDS plates containing

200 µg/ml Zeocin. A mini scale expression study (EndoH$_f$-treated) allowed selection of a clone with high expression levels (Figure 3.21).

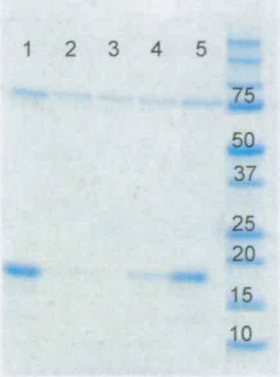


Figure 3.21 Gradient SDS-PAGE of mini-scale expression trial of four *P. pastoris* clones expressing FH-8-9$_{K446}$; Lanes 1-4 reducing conditions; lane 5 corresponds to sample shown in lane 1, but loaded under non-reducing conditions.

The supernatant of the mini-scale expression study was concentrated and buffer-exchanged into 20 mM potassium phosphate, pH 7.4.

3.1.7 Construct FH-$_{RR}$8-9

FH-$_{RR}$8-9 corresponds to the construct reported in a study by Ormsby *et al.* (61) and is characterised by a long N-terminal cloning artefact including two arginines. By-and-large the same cloning strategy was applied as reported above. However, two deviations from this protocol (Figure 3.2) were made in order to reproduce exactly the construct reported (61). The 5'-primer contained a *Kpn*1-restriction enzyme site (instead of *Pst*1 site as for all other constructs) and the *P. pastoris* expression vector used was pPICZαA (instead of pPICZαB as for all other constructs). Briefly, coding DNA for FH-$_{RR}$8-9 was amplified in a PCR reaction using pPICZαB sub-cloned FH-8-9 DNA as a template. The resulting DNA was ethanol precipitated and cloned (blunt-end cloning) into the pCR-4Blunt-TOPO vector. One Shot TOP10 chemically competent *E. coli* cells were transformed with the vector and selected on LB ampicillin (200 µg/ml) plates. The plasmid was amplified (by liquid culturing of

LB ampicillin (200 µg/ml) plates. The plasmid was amplified (by liquid culturing of transformed TOP10 cells), extracted and submitted to double restriction-enzyme digestion using *Kpn*1 and *Xba*1 restriction enzymes. After transformation of *P. pastoris,* colonies that grew on YPDS plates containing 300 µg/ml Zeocin were subjected to mini-scale expression studies. EndoH$_f$-treated samples were analysed by SDS-PAGE (Figure 3.22-A).

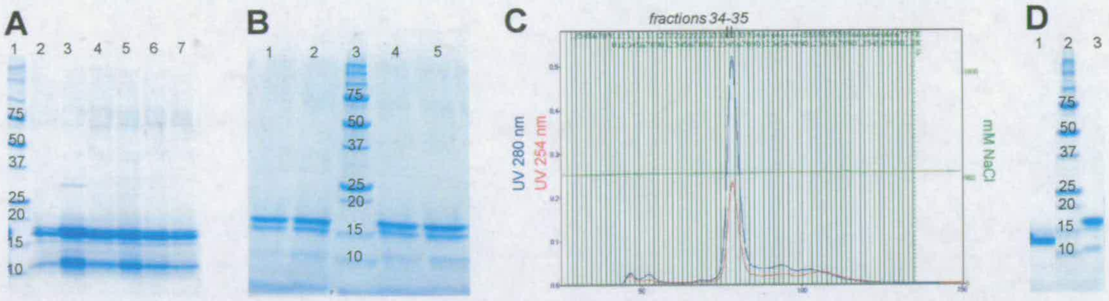


Figure 3.22 Expression and purification of FH-$_{\mathrm{RR}}$8-9; (A) Gradient SDS-PAGE of mini-scale expression study of six *P. pastoris* clones. (B) Gradient SDS-PAGE of fractions from anion-exchange chromatography. (C) The concentrated, semi-purified FH-$_{\mathrm{RR}}$8-9 sample was loaded on the HiLoad Superdex 75 size exclusion column equilibrated with 20 mM potassium phosphate, pH 7.4, containing 500 mM NaCl. (D) Gradient SDS-PAGE of pooled peak fractions 34 and 35 from size exclusion chromatography; non-reducing and reducing gel loading conditions were applied to lanes 1, and 3, respectively).

Despites the presence of the two Arg in the N-terminal cloning artefact of FH-$_{\mathrm{RR}}$8-9 the theoretical pI of the CCP module pair remained quite acidic (pI= 5.0). This allowed shaker flask-expressed FH-$_{\mathrm{RR}}$8-9 to be purified in the same way as FH-8-9. Figure 3.22-B shows SDS-PAGE analysis of fractions eluted from the HiTrap™ CaptoQ column. A small amount of impurities from *P. pastoris* proteins is visible as are distinct FH-$_{\mathrm{RR}}$8-9 degradation products. For further purification, size-exclusion chromatography was used (Figure 3.22-C). Central peak fractions were pooled and analysed on a gradient SDS-PAGE (Figure 3.22-D). While under non-reducing conditions a single protein band at the expected mobility was obtained, under reducing conditions a small amount of degradation is still evident. Partial validation of primary structure was obtained from a tryptic digest MS analysis (Figure 3.23).

```
Sequence coverage: 52% (71 of 134 amino acids)
1          11         21         31         41         51         61         71
EFTWPSRPSR IGTKTCSKSS IDIENGFISE SQYTYALKEK AKYQCKLGYV TADGETSGSI TCGKDGWSAQ PTCIKSCDIP
81         91         101        111
VFMNARTKND FTWFKLNDTL DYECHDGYES NTGSTTGSIV CGYNGWSDLP ICYE
```

Figure 3.23 (A) Sequence coverage in peptide mass fingerprinting for FH-$_{RR}$8-9. The amino acid sequence of FH-$_{RR}$8-9 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points (for mass spectrum of peptide fingerprint of construct FH-$_{RR}$8-9 see Appendices.

3.1.8 Construct FH-8-15

The FH-8-15 coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (by Carina Gandy working under the supervision of the author of this thesis) (Figure 3.2).

Batches of FH-8-15 were expressed in shaker flasks. The supernatant was concentrated and submitted to deglycosylation by treatment with endoglycosidase EndoH (New England Biolabs Ltd., Hitchin, UK). The deglycosylated sample was buffer exchanged into 20 mM sodium carbonate buffer, pH 9.0, and applied to a MonoQ anion-exchange chromatography column and eluted in a linear gradient of 20 column volumes to 1 M sodium chloride (Figure 3.24). Fractions analysed by SDS-PAGE show several impurities, originating mainly from suspected degradation of FH-8-15 with some possible contribution from *P. pastoris* proteins. Samples analysed under reducing conditions showed more impurity bands than those analysed under non-reducing conditions. This implies that a portion of the molecule might be ‘clipped’ but held together by disulfide bridges.

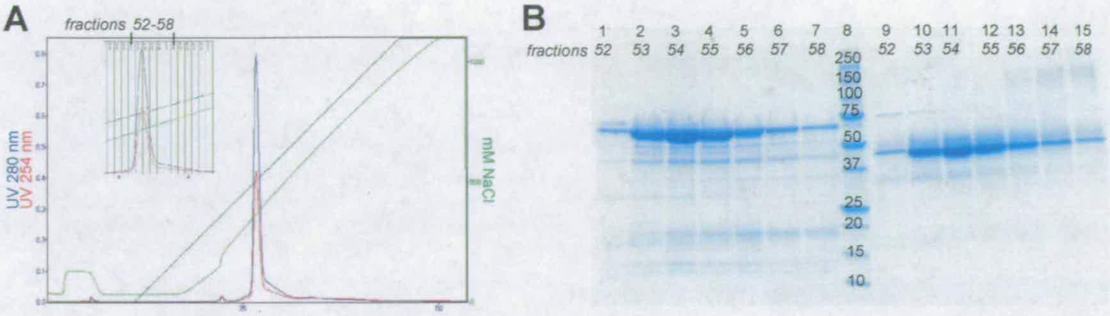


Figure 3.24 First purification step of FH-8-15; (A) Concentrated, deglycosylated and buffer-exchanged FH-8-15 was loaded onto the MonoQ anion-exchange column. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing conditions applied in lanes 1-8, and 9-15, respectively.

Peak fractions eluted from the Mono Q column were pooled and concentrated. As a second purification step, size-exclusion chromatography (using the HiLoad Superdex 75 size-exclusion column) was applied (in 20 mM potassium phosphate buffer, pH 7.4, 500 mM NaCl). Peak fractions of the eluted protein were monitored by SDS-PAGE (Figure 3.25).

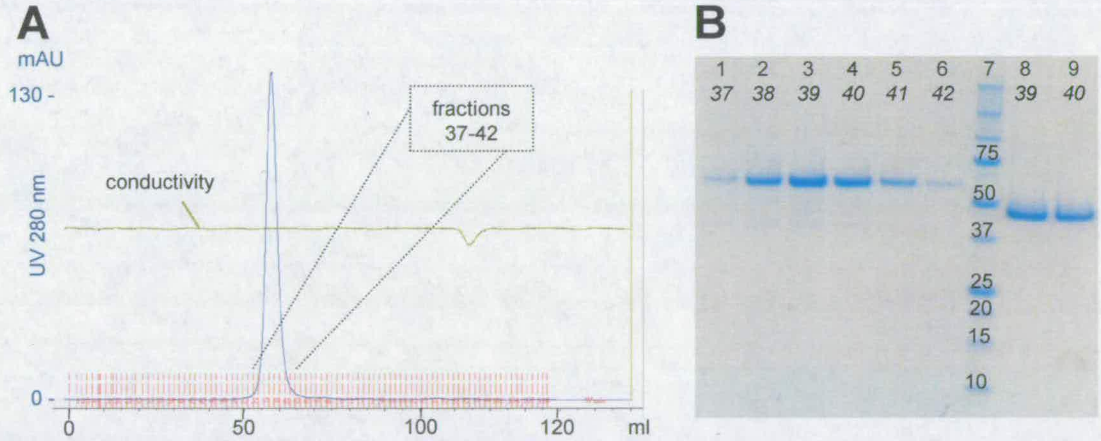


Figure 3.25 Second purification step of FH-8-15 (A) The concentrated and pre-purified FH-8-15 sample was loaded on the size-exclusion column. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing gel loading conditions applied to lanes 1-7, and 8-9, respectively.

Samples loaded under non-reducing conditions showed a pure single protein band, while under reducing loading a very small contamination consisting of two faint, putative degradation products, was noticeable. Partial confirmation of the primary sequence was achieved by a tryptic- digest MS. Folding was examined through acquisition of a 1-D ^{1}H-NMR experiment (Figure 3.26).

A

```
Sequence coverage: 14% (70 of 483 amino acids)
1          11         21         31         41         51         61         71
AGTCSKSSID IENGFISESQ YTYALKEKAK YQCKLGYVTA DGETSGSITC GKDGWSAQPT CIKSCDIPVF MNARTKNDFT
81         91         101        111        121        131        141        151
WFKLNDTLDY ECHDGYESNT GSTTGSIVCG YNGWSDLPIC YERECELPKI DVHLVPDRKK DQYKVGEVLK FSCKPGFTIV
161        171        181        191        201        211        221        231
GPNSVQCYHF GLSPDLPICK EQVQSCGPPP ELLNGNVKEK TKEEYGHSEV VEYYCNPRFL MKGPNKIQCV DGEWTTLPVC
241         51        261        271        281        291        301        311
IVEESTCGDI PELEHGWAQL SSPPYYYGDS VEFNCSESFT MIGHRSITCI HGVWTQLPQC VAIDKLKKCK SSNLIILEEH
321        331        341        351        361        371        381        391
LKNKKEFDHN SNIRYRCRGK EGWIHTVCIN GRWDPEVNCS MAQIQLCPPP PQIPNSHNMT TTLNYRDGEK VSVLCQENYL
401        411        421        431        441        451        461        471
IQEGEEITCK DGRWQSIPLC VEKIPCSQPP QIEHGTINSS RSSQESYAHG TKLSYTCEGG FRISEENETT CYMGKWSSPP
481
QCE
```

B

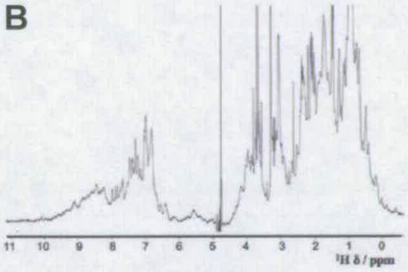


Figure 3.26 Validation of FH-8-15; (A) Sequence coverage in peptide mass fingerprinting for FH-8-15. The amino acid sequence of FH-8-15 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points. (for mass spectrum of peptide fingerprint of construct FH-8-15 see Appendices) (B) ^{1}H-NMR spectrum of FH-8-15; Spectrum shows good dispersion with well-resolved peaks. NMR sample is 68.3 µM in 20 mM potassium phosphate buffer pH 6.6.

Lineshape and good peak dispersion in the ^{1}H-NMR spectrum are consistent with a sample containing > 80 % folded material.

3.1.9 Construct FH-10-12

The FH-10-12 DNA coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (Figure 3.2). A small-scale expression trial was performed for those colonies that grew on YPDS plates containing 150 µg/ml Zeocin (Figure 3.27); EndoH$_f$-treated samples were analysed by SDS-PAGE.

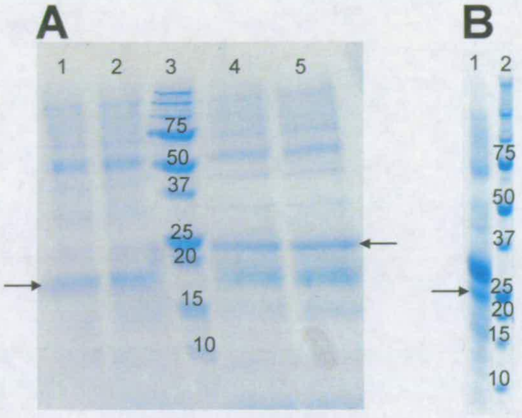


Figure 3.27 Gradient SDS-PAGE of recombinantly expressed FH-10-12. (A) Mini-scale expression trial of two *P. pastoris* clones expressing FH-10-12; non-reducing and reducing gel loading conditions were applied to lanes 1-2, and 4-5, respectively. (B) EndoH$_f$-treated supernatant from the shaker-flask expression. Two dominating bands are visible, one at the expected mobility for FH-10-12 co-migrating with the 25-kDa molecular weight standard (indicated by arrows) and one large diffuse band corresponding to the migration of a 30-kDa protein.

FH-10-12 was expressed in shaker flasks (Figure 3.27-B). The FH-10-12 band co-migrates with the 25-kDa of the molecular marker. The big diffuse band corresponding to a 30-kDa protein could be a FH-10-12 species that is still glycosylated due to incomplete EndoH$_f$-treatment or O-glycosylation. However, aberrant processing of the α-mating secretion signal by *P. pastoris* (as described in the section 3.1.1), cannot be ruled out.

The purification steps for FH-10-12 were almost identical to the ones used for FH-8-15. Following harvest the supernatant was concentrated and submitted to deglycosylation by treatment with endoglycosidase EndoH$_f$. The deglycosylated sample was concentrated and buffer-exchanged into 20 mM potassium phosphate buffer, pH 7.8, and flowed over 1 ml of Con-A lectin resin. The flow-through was applied to the MonoQ anion-exchange column. Bound material was eluted in a linear gradient of 20 column volumes to 1 M NaCl (Figure 3.28-A). Fractions spanning the eluted peak were analysed by SDS-PAGE (Figure 3.28-B).

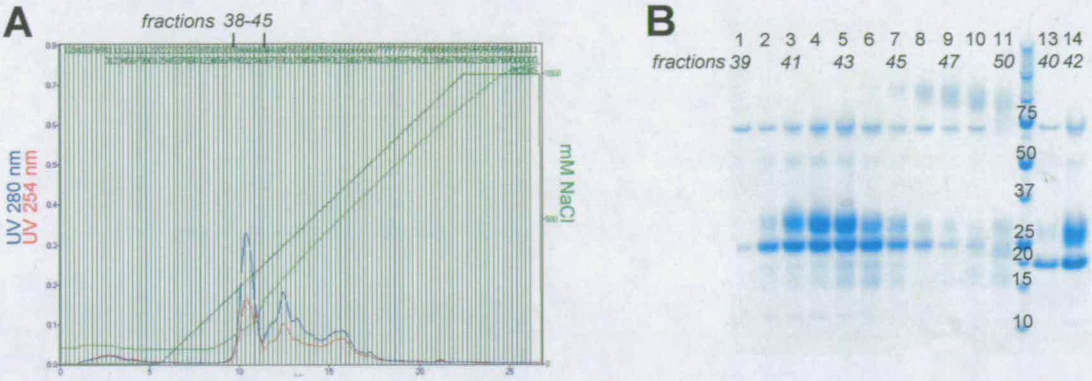


Figure 3.28 First purification step of FH-10-12. (A) Concentrated, deglycosylated and buffer-exchanged FH-10-12 was loaded onto the MonoQ column and then eluted with a linear gradient, over 20 column volumes, to 1 M NaCl. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing gel loading conditions were applied to lanes 1-12, and 13-14, respectively.

Anion-exchange chromatography was able to resolve FH-10-12 from some of the contaminating *P. pastoris* proteins. However, significant amount of impurities remained. Under these conditions the endoglycosidase EndoH$_f$ (~75 kDa) could not be separated from the FH-10-12 material either. However, size-exclusion chromatography should easily separate the added endoglycosidase from FH-10-12. Furthermore, the diffuse higher band (~30 kDa) co-eluted with FH-10-12 (~25 kDa). Neither EndoH$_f$-treatment nor application to Con-A lectin resin were able to resolve the two FH-10-12 species, in the FH-10-12 preparation. Since both approaches (EndoH$_f$ and Con-A lectin resin) work on the basis of recognising mainly high mannose structures (161, 162) it may be that the diffuse band represents N- or O-linked glycosylation of mannose-poor carbohydrates, despite the fact that glycosylation in yeast vectors are know to be rich in high-mannose structures (163).

As a second purification step, size-exclusion chromatography was applied (in 20 mM potassium phosphate buffer, pH 7.4, 500 mM sodium chloride). Several small- to medium-peaks eluted prior to the main peak; the diffuse upper band corresponds to higher molecular weight species of FH-10-12, while a band at the expected place for FH-10-12 is also evident (Figure 3.29).

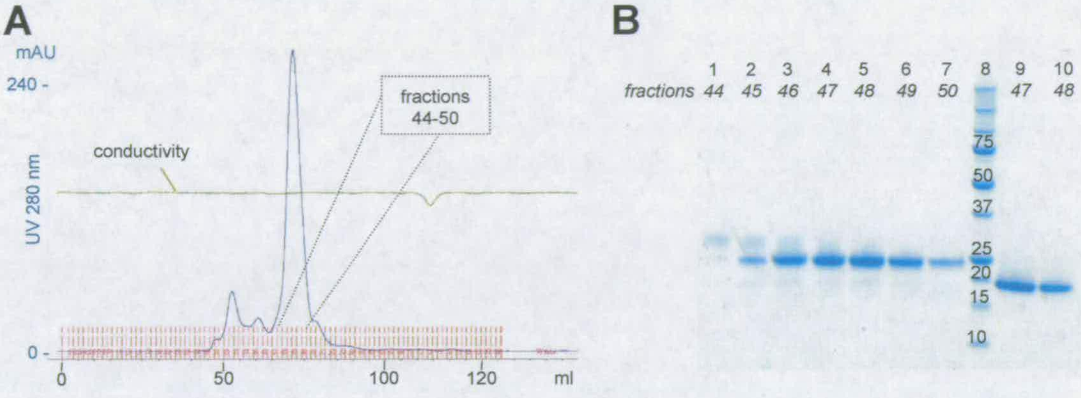
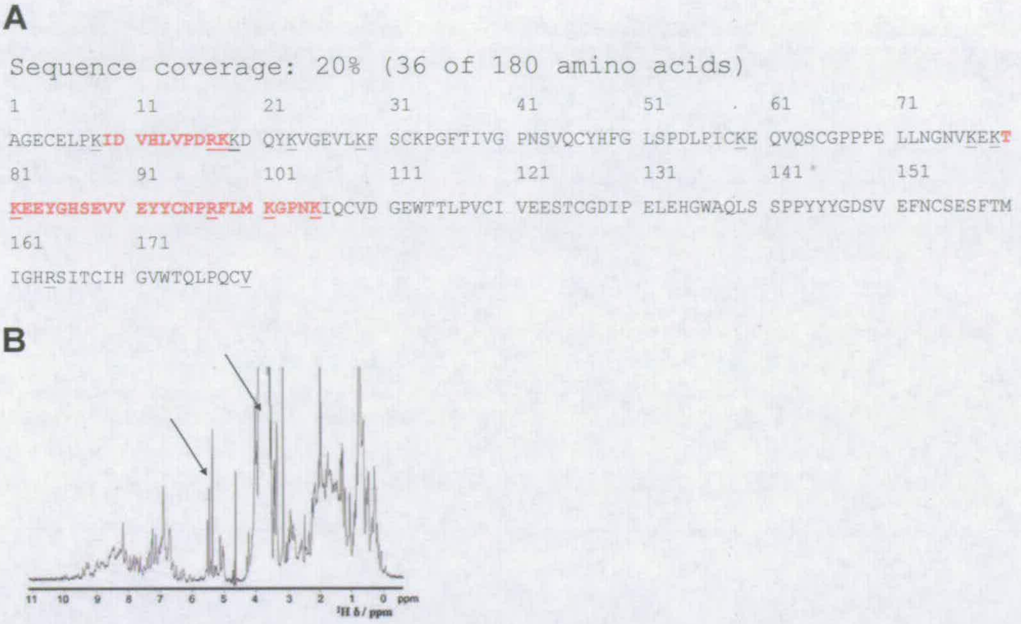


Figure 3.29 Second purification step of FH-10-12 (A) Concentrated and pre-purified FH-10-12 sample was applied to the HighLoad Superdex 75 size-exclusion column. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing conditions applied in lanes 1-7, and 9-10, respectively. The main band is in the right place for FH-10-12.

Fraction 47 to 49 were pooled and concentrated. Partial confirmation of primary sequence was obtained from tryptic digest MS (Figure 3.30-A). Folding was confirmed by acquisition of a 1-D ^{1}H-NMR experiment (Figure 3.30-B).



```
Sequence coverage: 20% (36 of 180 amino acids)
1          11         21         31         41         51       . 61         71
AGECELPKID VHLVPDRKKD QYKVGEVLKF SCKPGFTIVG PNSVQCYHFG LSPDLPICKE QVQSCGPPPE LLNGNVKEKT
81         91         101        111        121        131        141 '      151
KEEYGHSEVV EYYCNPRFLM KGPNKIQCVD GEWTTLPVCI VEESTCGDIP ELEHGWAQLS SPPYYYGDSV EFNCSESFTM
161        171
IGHRSITCIH GVWTQLPQCV
```

Figure 3.30 Validation of FH-10-12 (A) Sequence coverage in peptide mass fingerprinting for FH-10-12. The amino acid sequence of FH-10-12 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points. (for mass spectrum of peptide fingerprint of construct FH-10-12 see Appendices) (B) ^{1}H-NMR spectrum of FH-10-12; Spectrum shows good dispersion with well resolved peaks. The NMR sample was 127 µM in 20 mM potassium phosphate buffer, pH 6.6. Arrows indicate signals arising from contaminating carbohydrate.

The spectrum indicates the presence of some contaminating sugars (indicated with arrows) derived from *P. pastoris*. Lineshape and good peak dispersion in the 1-D ^{1}H-NMR experiment are consistent with a sample containing > 80 % folded material.

3.1.10 Construct FH-10-15

The FH-10-15 coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (Figure 3.2) (in collaboration with Isabell Pechtl and Patience Tetteh-Quarcoo working under the supervision of the author).

A small-scale expression trial (Figure 3.31-A) was performed on clones that grew on YPDS plates containing 300 µg/ml Zeocin. The presence of N-glycosylation sites necessitated removal of the Asn-linked.

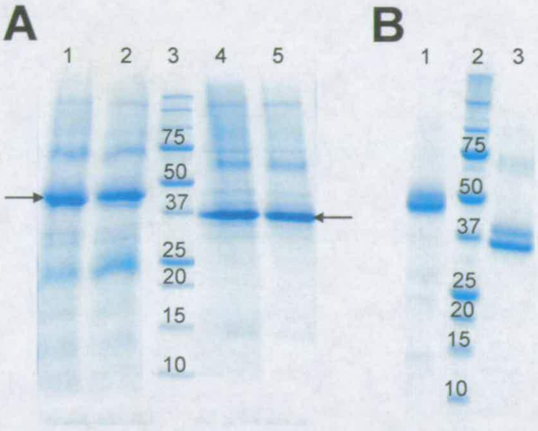


Figure 3.31 Gradient SDS-PAGE of *P. pastoris* FH-10-15 expression. (A) Mini-scale expression trial of two *P. pastoris* clones expressing FH-10-15; samples run under reducing conditions, lanes 1-2, and under non-reducing conditions, lanes 4-5, respectively. (B) EndoH$_f$-treated supernatant from shaker-flask expression. The dominant band shows the mobility expected for FH-10-15. Reducing and non-reducing gel loading conditions applied to lanes 1 and 3, respectively.

FH-10-15 was expressed in shaker flasks or four litre fermentations. Purification followed the same protocol. A small EndoH$_f$-treated sample of the expression medium was treated with EndoH$_f$ and analysed on SDS-PAGE (Figure 3.31-B). The gel shows a protein band of FH-10-15 at a mobility corresponding to 42 kDa (reducing conditions) and at 35 kDa (non-reducing conditions). Concentrated and deglycosylated FH-10-15 was loaded onto the HiLoad Superdex 75 size-exclusion column. Peak fractions of the eluted protein were monitored by SDS-PAGE (Figure 3.32).

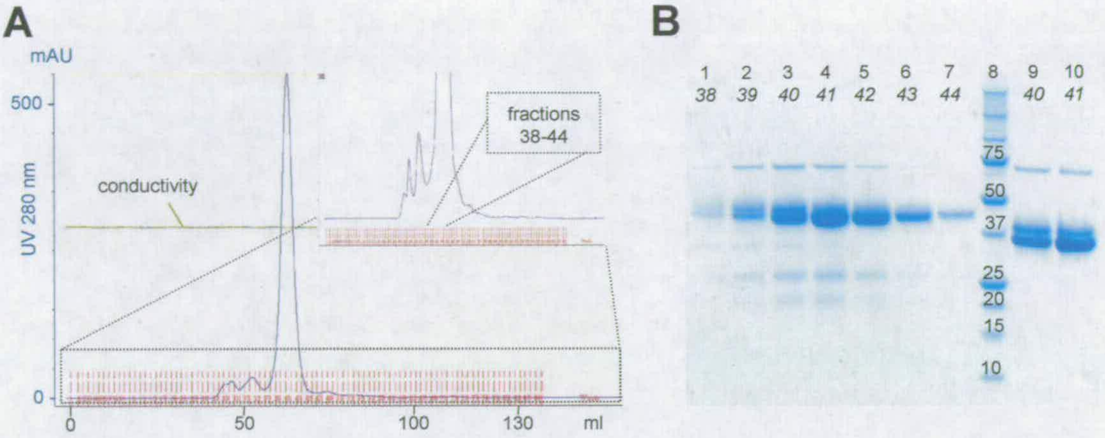


Figure 3.32 First purification step of FH-10-15 (A) Concentrated and deglycosylated FH-10-15 was loaded onto the size-exclusion column equilibrated with 20 mM potassium phosphate at pH 7.4 containing 500 mM NaCl. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing conditions applied to lanes 1-8, and 9-10, respectively.

Fractions 39 to 44, which correspond to the main peak in the size-exclusion chromatogram, show only a little high-molecular-weight impurity and some faint degradation bands. Fractions 40-43 were pooled and buffer-exchanged into 20 mM sodium carbonate buffer, pH 9.0, prior to application onto the MonoQ anion-exchange column. Retarded material was eluted in a linear gradient over 20 column volumes to 1 M NaCl (Figure 3.33). SDS-PAGE analysis shows single bands corresponding to FH-10-15.

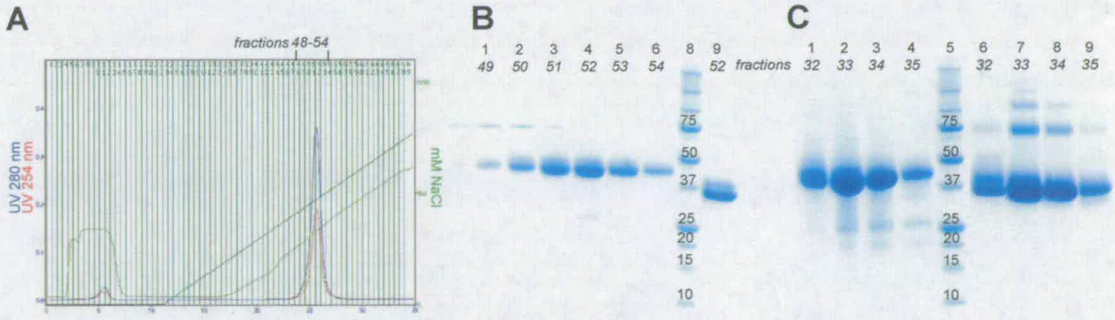


Figure 3.33 Second purification step of FH-10-15 (A) Deglycosylated FH-10-15 was subjected to anion-exchange chromatography. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing conditions applied to lanes 1-8, and 9, respectively. (C) Gradient SDS-PAGE of fractions of anion-exchange chromatography of fermentor-expressed FH-10-15. Reducing and non-reducing conditions applied to lanes 1-5, and 6-9, respectively.

An overloaded gel of fractions from the final anion-exchange purification step (from the fermentor batch) shows that - under non-reducing conditions - a small proportion of FH-10-15 material multimerises to dimers and trimers. It is not clear whether such multimerisation is, in fact present in the other sample but was too weak to be detectable, or whether it only occurs due to the higher protein concentrations. The proportion of multimerised species remained below ~5% (section 5.3). ^{1}H-NMR spectrum of a similarly concentrated protein (see below) showed little sign of aggregation.

A

```
Sequence coverage: 40% (149 of 365 amino acids)
1          11         21         31         41         51         61         71
TAGECELPKI DVHLVPDRKK DQYKVGEVLK FSCKPGFTIV GPNSVQCYHF GLSPDLPICK EQVQSCGPPP ELLNGNVKEK
81         91         101        111        121        131        141        151
TKEEYGHSEV VEYYCNPRFL MKGPNKIQCV DGEWTTLPVC IVEESTCGDI PELEHGWAQL SSPPYYYGDS VEFNCSESFT
161        171        181        191        201        211        221        231
MIGHRSITCI HGVWTQLPQC VAIDKLKKCK SSNLIILEEH LKNKKEFDHN SNIRYRCRGK EGWIHTVCIN GRWDPEVNCS
241        251        261        271        281        291        301        311
MAQIQLCPPP PQIPNSHNMT TTLNYRDGEK VSVLCQENYL IQEGEEITCK DGRWQSIPLC VEKIPCSQPP QIEHGTINSS
321        331        341        351        361        371        381        391
RSSQESYAHG TKLSYTCEGG FRISEENETT CYMGKWSSPP QCESR
```

B

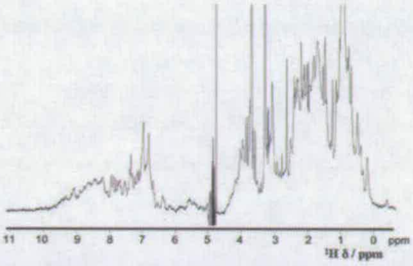


Figure 3.34 Validation of FH-10-15 (A) Sequence coverage in peptide mass fingerprinting for FH-10-15. The amino acid sequence of FH-10-15 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points (for mass spectrum of peptide fingerprint of construct FH-10-15 see Appendices). (B) ^{1}H-NMR spectrum of FH-10-15; spectrum shows good dispersion with well-resolved peaks. The NMR sample is 48 µM in 20 mM potassium phosphate buffer, pH 6.6.

Partial validation of primary structure was obtained from a tryptic digest MS (Figure 3.34-A). Folding was examined through acquisition of a 1-D ^{1}H-NMR experiment (Figure 3.34-B). Lineshape and good peak dispersion in the ^{1}H-NMR experiment are consistent with a sample containing > 80 % folded material.

3.1.11 Construct FH-11

A *P. pastoris* KM71H clone over-expressing FH-11 was prepared using identical steps to those described above (Figure 3.2). Figure 3.35-A shows a small-scale expression trial for clones that grew on YPDS plates containing 150 μg/ml Zeocin. All clones expressed a protein that shows migration corresponding to ~10 kDa under reducing conditions. Under non-reducing conditions, migration of FH-11 was surprisingly decreased. Since the bands corresponding to non-reducing conditions were very faint, the surplus supernatant from lane 3 (Figure 3.35-A) was concentrated in a spin concentrator and analysed again under non-reducing and reducing loading condition (Figure 3.35-B). From this gel it is clear that the single module FH-11 runs at the size of a dimer when loaded under non-reducing conditions. Whether this effect is an artefact of the *P. pastoris* expression system or was artificially introduced due to absence of neighbouring (stabilising) modules, or indeed reflects properties of FH-11 even in the context of full length FH, needs to be investigated further. For other single modules expressed in this study (FH-7$_{Y402}$, FH-7$_{H402}$, FH-12, FH-13) no dimeriasation was observed, neither did FH-10-12 - encompassing module 11 – show dimerisation. On the other hand, FH-10-15 did show a small degree of dimerisation and trimerisation under non-reducing conditions.

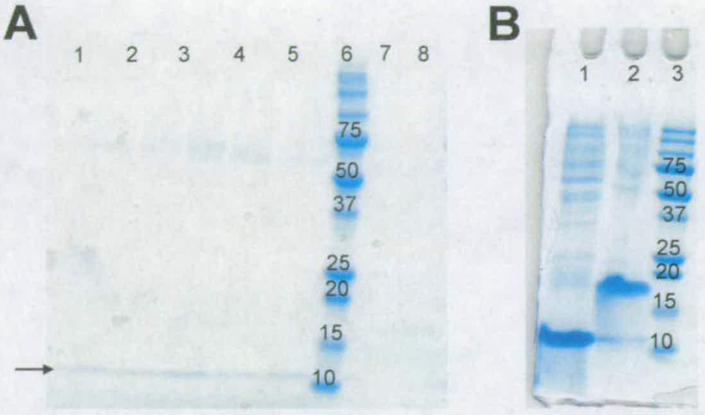

A
1
2
3
4
5
6
7
8
75
50
37
25
20
15
10
B
1
2
3
75
50
37
25
20
15
10

Figure 3.35 Gradient SDS-PAGE of recombinant FH-11. (A) Mini-scale expression study of five *P. pastoris* clones expressing FH-11 (arrow indicating the band corresponding to FH-11); Reducing and non-reducing conditions apply to lanes 1-5, and 7-8, respectively. (B) SDS-PAGE was repeated with a 20-times more concentrated sample from lane 3 of gel A). Reducing and non-reducing conditions apply to lanes 1, and 2, respectively.

3.1.12 Construct FH-11-14

The FH-11-14 coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (Figure 3.2).

A small-scale expression trial was performed on colonies that grew on YPDS plates containing 150 µg/ml Zeocin (Figure 3.36). Asn-linked glycans were trimmed prior to SDS-PAGE analysis.

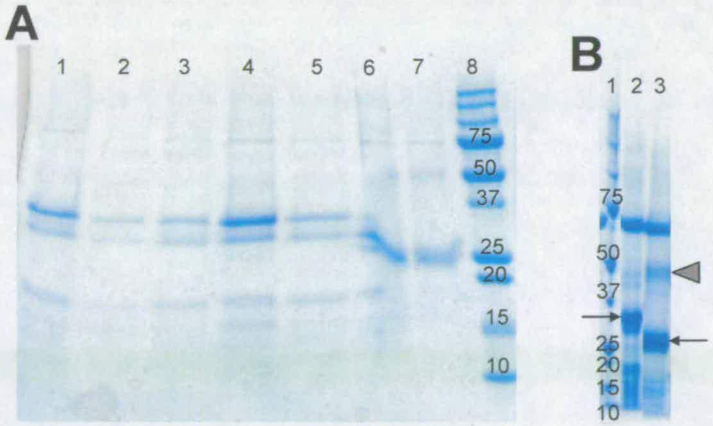

A
1
2
3
4
5
6
7
8
75
50
37
25
20
15
10
B
1
2
3
75
50
37
25
20
15
10

Figure 3.36 Gradient SDS-PAGE of *P. pastoris* FH-11-14 expression. (A) Mini-scale expression trial of five *P. pastoris* clones expressing FH-11-14; reducing and non-reducing gel loading conditions apply to lanes 1-5, and 6-7, respectively. (B) EndoH$_f$ treated supernatant from shaker-flask expression. The dominant band (highlighted by an arrow) shows the mobility expected for FH-11-14. Reducing and non-reducing conditions apply to lanes 2 and 3, respectively. A triangle indicates a band (lane 3) corresponding to the mobility expected for dimerised FH-11-14.

FH-11-14 was expressed in shaker-flasks (Figure 3.36-B) and purified using identical steps as used for FH-10-15. Apart from *P. pastoris* proteins and a band corresponding to EndoH$_f$ (~ 75 kDa) SDS-PAGE shows the presence of putative FH-11-14 degradation under reducing conditions, while under non-reducing conditions the degradation appears to be absent. This effect was also observed for several other constructs (see above) and is thought to reflect a situation in which 'clipped' peptides remain held together by disulfide bridges. As with FH-10-15, a small proportion of dimerised protein is visible on the gel under non-reducing conditions. The harvested supernatant was concentrated, deglycosylated and loaded onto the size-exclusion column equilibrated with 20 mM potassium phosphate buffer, pH 7.4, 500 mM NaCl. Eluted protein was monitored by SDS-PAGE (Figure 3.37).

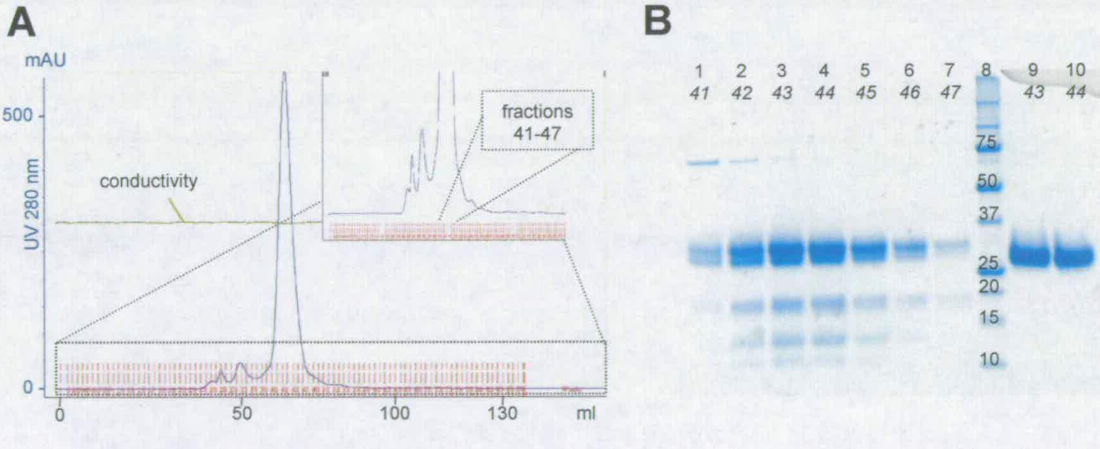


Figure 3.37 First purification step of FH-11-14 (A) FH-11-14 was loaded onto the HiLoad Superdex 75 size-exclusion column. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing conditions apply to lanes 1-8, and 9-10, respectively.

Fractions 42-46 contain putative degradation bands. These fractions were pooled and buffer-exchanged into 20 mM sodium carbonate buffer pH 9.0 prior to application onto the MonoQ anion-exchange column. Retarded material was eluted in a linear gradient over 20 column volumes to 1 M NaCl (Figure 3.38). SDS-PAGE analysis reveals pure FH-11-14. Trace amounts of supposed degradation are only visible under reducing conditions; SDS-PAGE analysis under non-reducing conditions yields a single band.

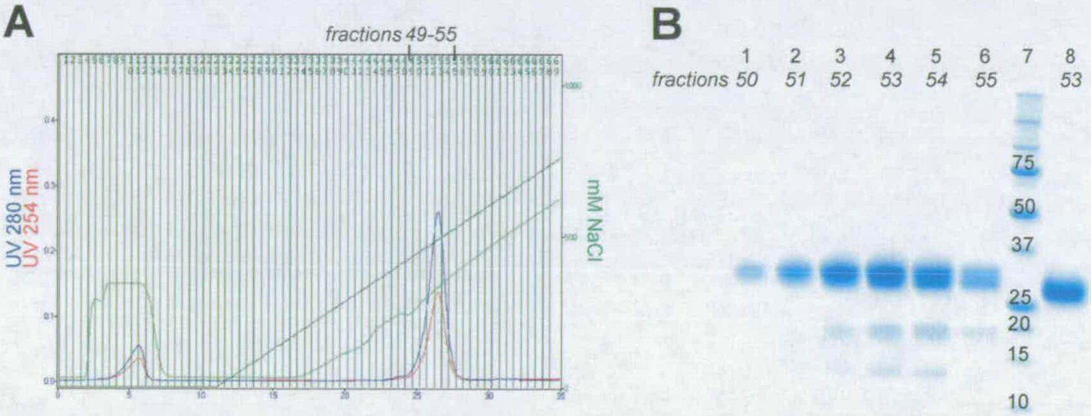
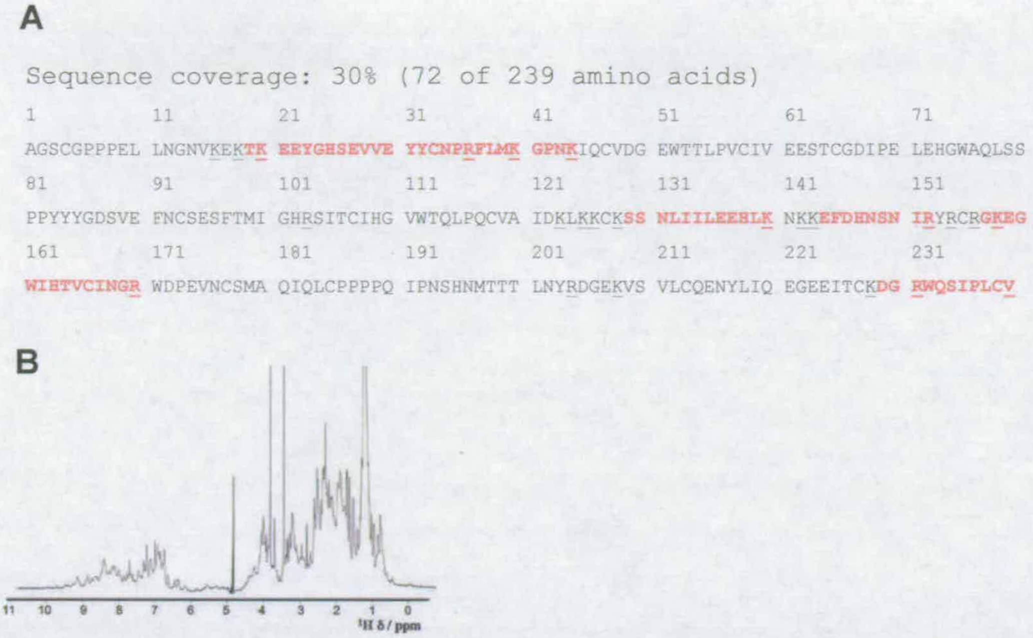


Figure 3.38 Second purification step of FH-11-14 (A) Anion-exchange chromatogram of FH-11-14. (B) Gradient SDS-PAGE of fractions indicated in A). Reducing and non-reducing conditions apply to lanes 1-7, and 8, respectively.

Partial validation of primary structure was obtained from a tryptic digest MS (Figure 3.39-A). Folding was checked through acquisition of a 1-D ^{1}H-NMR experiment (Figure 3.39-B).



```
Sequence coverage: 30% (72 of 239 amino acids)
1          11         21         31         41         51         61         71
AGSCGPPPEL LNGNVKEKTK EEYGHSEVVE YYCNPRFLMK GPNKIQCVDG EWTTLPVCIV EESTCGDIPE LEHGWAQLSS
81         91         101        111        121        131        141        151
PPYYYGDSVE FNCSESFTMI GHRSITCIHG VWTQLPQCVA IDKLKKCKSS NLIILEEHLK NKKEFDHNSN IRYRCRGKEG
161        171        181        191        201        211        221        231
WIETVCINGR WDPEVNCSMA QIQLCPPPPQ IPNSHNMTTT LNYRDGEKVS VLCQENYLIQ EGEEITCKDG RWQSIPLCV
```

Figure 3.39 Validation of FH-11-14 (A) Sequence coverage in peptide mass fingerprinting for FH-11-14. The amino acid sequence of FH-11-14 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points (for mass spectrum of peptide fingerprint of construct FH-11-14 see Appendices). (B) ^{1}H-NMR spectrum of FH-11-14; NMR sample is 82 μM in 20 mM potassium phosphate buffer, pH 6.6.

Lineshape and good peak dispersion in the ^{1}H-NMR spectrum are consistent with a sample containing > 80 % folded material.

3.1.13 Construct FH-12

A *P. pastoris* KM71H clone over-expressing FH-12 was prepared using identical steps to those described in Figure 3.2.

A small-scale expression trial (Figure 3.40) was performed on clones that were selected on YPDS plates containing 150 μg/ml Zeocin. Prior to gel loading, samples were deglycosylated. In both cases, under reducing and non-reducing conditions, migration of FH-12 remains at the same level (~ 10 kDa).

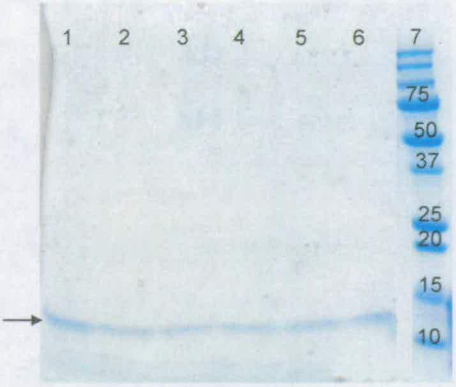


Figure 3.40 Gradient SDS-PAGE of *m*ini-scale expression trial of five *P. pastoris* clones expressing FH-12 (arrow indicating the band corresponding to FH-12). Reducing and non-reducing conditions apply to lanes 1-5, and 6, respectively.

3.1.14 Construct FH-12-13

The FH-12-13 coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (Figure 3.2). A small-scale expression trial (Figure 3.41) was performed on clones that grew on YPDS plates containing 150 µg/ml Zeocin. Samples were deglycosylated prior to gel loading.

Batches of FH-12-13 were expressed (unlabelled) in shaker-flasks or ^{15}N-isotopically and ^{15}N-^{13}C-isotopically enriched in one-litre scale fermentations.

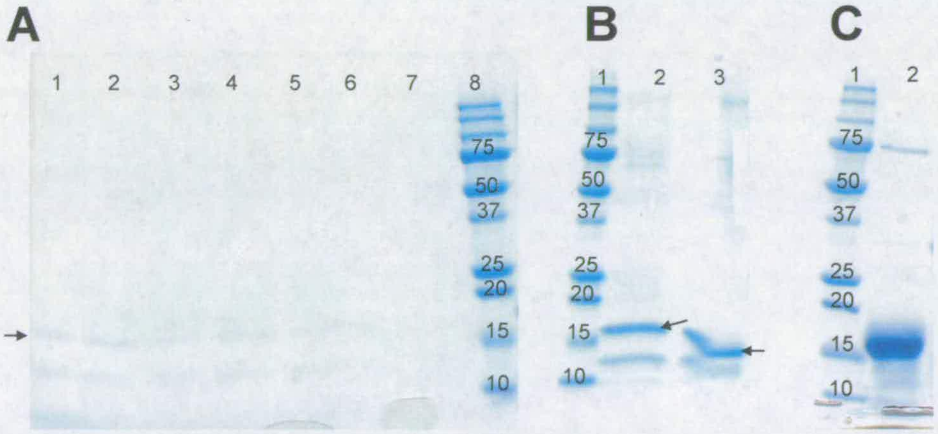


Figure 3.41 Gradient SDS-PAGE of *P. pastoris* FH-12-13 expression. (A) Mini-scale expression trial of five *P. pastoris* clones expressing FH-12-13; faint bands at mobility expected for FH-12-13 are highlighted by an arrow. Reducing and non-reducing conditions apply to lanes 1-5 and 6-

7, respectively). (B) Concentrated supernatant from mini-scale expression trial from lane 2 of A). Reducing and non-reducing conditions apply to lanes 2 and 3 respectively. (C) EndoH$_f$-treated elution from batch-absorbed FH-12-13.

The purification protocol for ^{15}N-/^{15}N-^{13}C-isotopically enriched one-litre scale) fermentations and unlabelled shaker-flask expression was very similar

Supernatant from cell cultures was diluted five-fold and the pH adjusted to 5.0 for batch-absorption to heparin-Sepharose. After elution (1 M NaCl) the protein was mixed with a protease inhibitor cocktail (containing 4-(2-aminoethyl) benzenesulfonyl fluoride, pepstatin A, E-64, and 1,10-phenanthroline – purchased from Sigma) and submitted to EndoH$_f$ treatment (Figure 3.41-C). Deglycosylated sample was buffer-exchange prior to loading onto a Self Pack POROS 20 heparin affinity column (diameter = 4.6 mm, length = 100 mm) equilibrated with 20 mM sodium acetate buffer, pH 5.0. Immobilised material was eluted with a linear gradient over 15 CV to 0.5 M NaCl (Figure 3.42-A). The elution profile shows one dominating peak that slightly overlaps with a small second peak. SDS-PAGE analysis of the fractions from the two peaks show a single protein species at the migration expected for FH-12-13. The fraction corresponding to the second peak exhibits some degraded material (Figure 3.42-B).

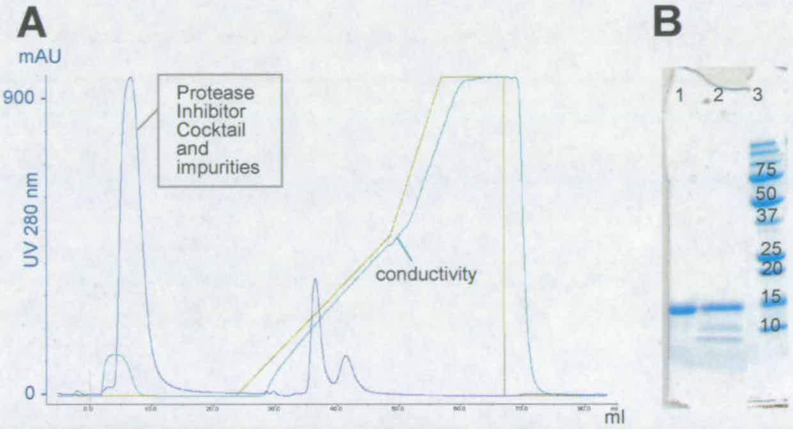


Figure 3.42 Purification of FH-12-13 (A); Pre-purified FH-12-13 was applied to heparin-affinity chromatography. (B) Gradient SDS-PAGE of the two peak fractions from A). Samples in lane 1

and lane 2 correspond to the first and second peak, respectively. The strong peak in the flow through fractions mostly corresponds to constituents in the protease-inhibitor cocktail.

Fractions spanning the first protein peak were pooled, concentrated and submitted to validation by NMR, tryptic-digest MS (Figure 3.43-A) and positive electrospray ionisation mass spectrometry. All analyses by mass spectrometry were performed with unlabelled material. The subsequently determined high resolution structure of FH-12-13 (see CHAPTER 5) validates the integrity and structure of this recombinantly expressed protein. The electrospray-ionisation mass spectrum (Figure 3.43-B) shows four major peaks. Peak B (13565 Da) corresponds to the predicted mass of FH-12-13 having one unit of "EA" (from the *P. pastoris* expression system) at the N-terminus: $M_r = 13563$ Da. Peak D (13767 Da) corresponds to the same species but with a GlucNAc unit (the core sugar unit that does not get trimmed by EndoH$_f$): $M_r = 13766$ Da. Peaks A (13435 Da) and C (13639 Da) are derivatives of peaks B and D, respectively. The mass difference between A and B is 128 Da indicating the loss of the N-terminal Glu (theoretical $M_r = 129$). Similarly, the mass difference between C and D (127 Da) is consistent with the loss of the N-terminal glutamate.

A

```
Sequence coverage: 60% (71 of 117 amino acids)
1          11         21         31         41         51         61         71
AGTCGDIPEL EHGWAQLSSP PYYYGDSVEF NCSESFTMIG HRSITCIHGV WTQLPQCVAI DKLKKCKSSN LIILEEHLKN
81         91         101        111
KKEFDHNSNI RYRCRGKEGW IHTVCINGRW DPEVNCS
```

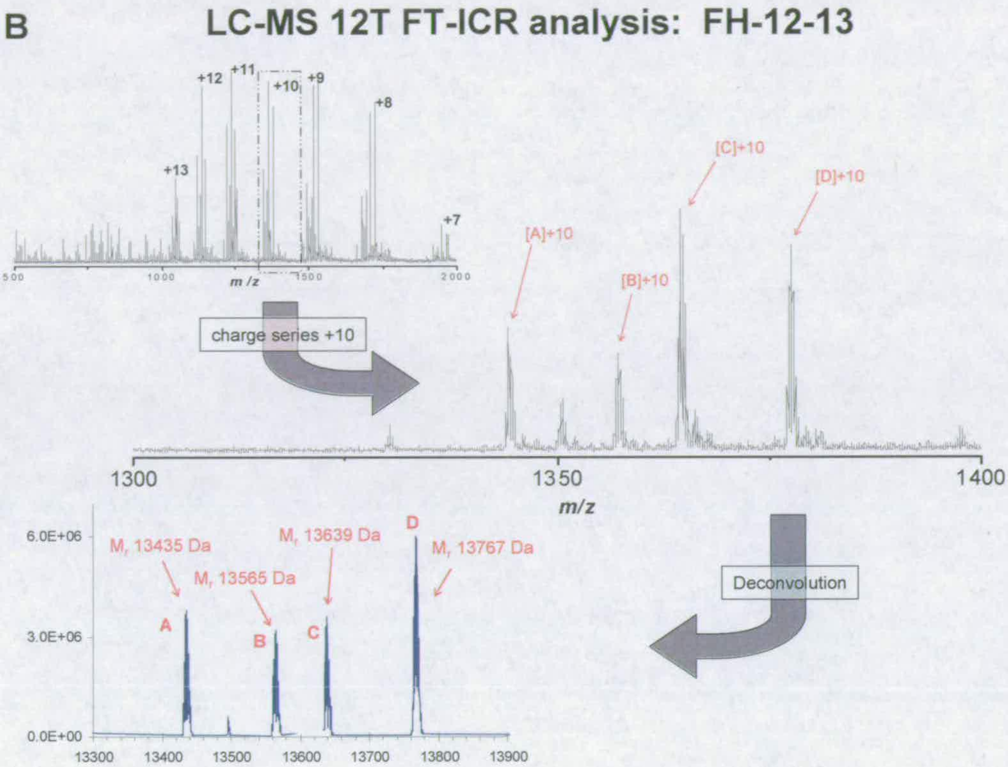

B LC-MS 12T FT-ICR analysis: FH-12-13

Figure 3.43 Validation of FH-12-13 (A) Sequence coverage in peptide mass fingerprinting for FH-12-13. The amino acid sequence of FH-12-13 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points (for mass spectrum of peptide fingerprint of construct FH-12-13 see Appendices). (B) Liquid-chromatography, 12-Tesla, Fourier-Transform, Ion-Cyclotron-Resonance Mass-Spectrometry analysis of FH-12-13.

In summary, mass spectrometry analysis revealed four major species of FH-12-13. It appears that during expression of FH-12-13 one "EA"-unit remained at the N-terminus (peak B: EA-FH-12-13). A proportion of FH-12-13 has, in addition to the leading "EA", one N-glycosylation site occupied with one GlucNAc (peak D: EA-FH-12-13-GlucNAc). Judging from the relative heights of the charge series (+10) signals, roughly 50% of both (EA-FH-12-13 and EA-FH-12-13-GlucNAc species) have lost the N-terminal glutamate yielding A-FH-

12-13 and A-FH-12-13-GlucNAc. A comparison of relative intensities between species having one N-glycosylation site occupied (peaks C and D) and having none occupied (peaks A and B), indicates that about 65% of FH-12-13 is glycosylated.

3.1.15 Construct FH-13-14 and FH-13-14$_{Q802, Q822}$

Two *P. pastoris* KM71H clones over-expressing FH-13-14 or FH-13-14$_{Q802, Q822}$ were prepared using similar steps to those described in Figure 3.2.

Small-scale expression tests were performed for four colonies (FH-13-14$_{Q802, Q822}$) that grew on YPDS-plates containing 300 µg/ml Zeocin (Figure 3.44-A).

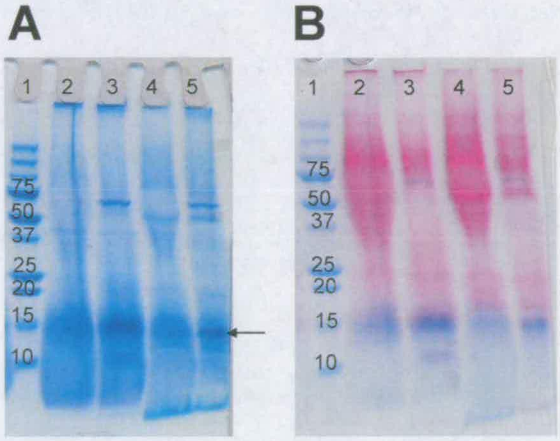


Figure 3.44 Gradient SDS-PAGE of *P. pastoris* FH-13-14$_{Q802\_Q822}$ expression. (A) Mini-scale expression trial of four clones expressing FH-13-14$_{Q802\_Q822}$; bands of concentrated samples at the mobility expected for FH-13-14$_{Q802\_Q82}$ are highlighted by an arrow. Bands are blurred although molecular weight markers (lane 1) show distinct bands. (B) Same gel as in A), but after treatment with the sigma glycosylation detection kit; protein with attached sugars is stained pink.

The resultant SDS-PAGE bands appear quite diffuse Figure 3.44-A). Samples also seemed to have undergone significant proteolysis. One possible reason for the observation of diffuse SDS-PAGE bands might originate from O-glycosylation. Engineering out of two N-glycosylation sites could have created a consensus site for O-glycosylation. To determine if this was the case, gels were treated with Sigma glycosylation detection kit (see section 2.1.18Glycoprotein detection and Figure 3.44-B), which oxidises sugar residues and specifically stains them pink. Some pink colouration is apparent in each band. Thus, the possibility of O-glycosylation in FH-13-14$_{Q802, Q822}$ exists.

Due to these degradation and possible glycosylation problems with FH-13-14$_{Q802\_Q822}$, the wildtype FH-13-14 construct including five of the eight amino acid residues in the 12-13 linker was prepared. The same techniques as described above were utilised to create this constructs. As the construct FH-13-14$_{Q802\_Q822}$ had already been integrated into the pPICZαB vector, two site directed mutagenesis steps were performed to transform the N-glycosylation sites back so as to obtain the wild type sequence. Maxi-prep plasmid extraction of FH-13-14 in pPICZαB vector was followed by linearization, purification and concentration of the cut plasmid prior to transformation into electrocompetent KM71H *P. pastoris* cells. A small-scale expression tests was performed on colonies that were selected on YPDS plates containing 300 μg/ml Zeocin.

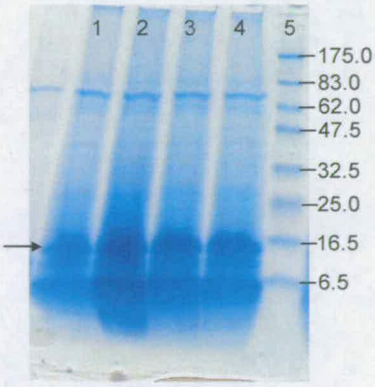


Figure 3.45 Gradient SDS-PAGE of *P. pastoris* FH-13-14 expression. Mini-scale expression trial of four clones expressing FH-13-14; EndoH$_f$–treated, concentrated samples were loaded. An arrow highlights bands at the mobility expected for FH-13-14. A high proportion of degraded material is present.

Expression levels for clones expressing FH-13-14 are high, but substantial degradation is evident. The wild-type form was selected for larger-scale expression and purification studies.

Despite numerous attempts to optimise conditions so as to minimise proteolysis, the presence of degradation products persisted (Figure 3.46-C). Furthermore it became clear that FH-13-14 oligomerises to form dimers and trimers, which can be separated by gel-filtration and yield a single band under non-reducing conditions, but multiple bands under oxidising conditions (Figure 3.46).

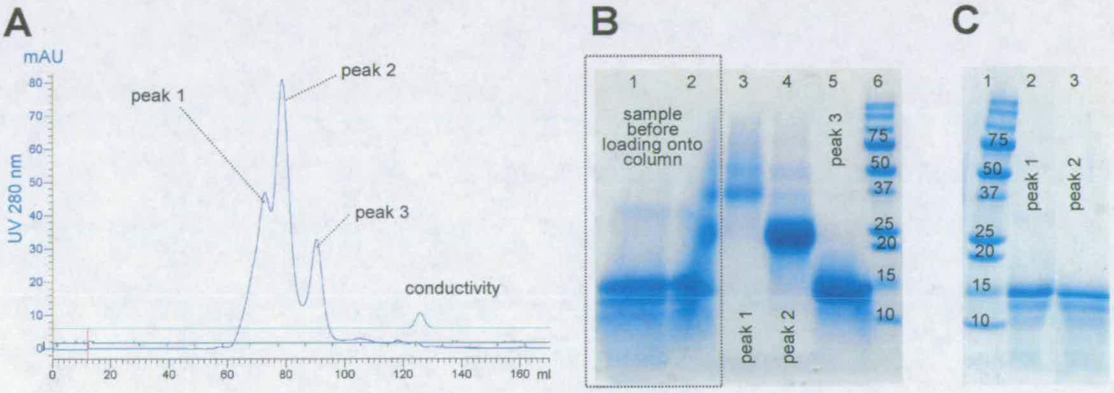
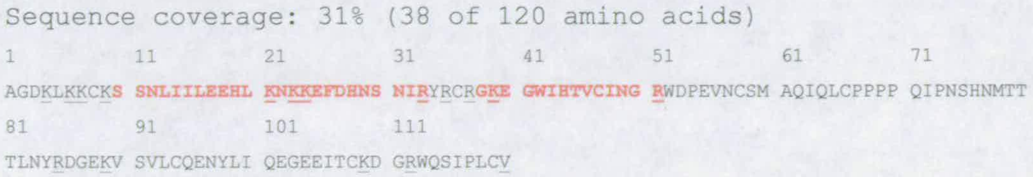


Figure 3.46 Size-exclusion chromatogram of FH-13-14 (A) Pre-purified, deglycosylated and concentrated FH-13-14 sample was loaded onto a Superdex size-exclusion column. (B) Gradient SDS-PAGE of fractions from size-exclusion chromatography shown in A). Reducing and non-reducing gel-loading conditions apply to lanes 1, and 2-5, respectively. Lanes 1 and 2 show condition of the sample prior to loading onto the size-exclusion column. Lane 3, 4 and 5 correspond to peaks 1, 2 and 3, respectively. (C) Samples of peak 1 and peak 2 loaded under reducing conditions.

Size-exclusion chromatography was able to separate three FH-13-14 species corresponding to trimer, dimer and monomer, as judged by their molecular weight from non-reducing SDS-PAGE analysis. FH-13-14 samples prior to gel filtration, and samples corresponding to peaks 2 (dimer) and 3 (monomer) were submitted to primary sequence validation by tryptic digest MS (Figure 3.47). All samples showed very similar to identical peptide fingerprint verifying that oligomers correspond to FH-13-14.



```
Sequence coverage: 31% (38 of 120 amino acids)
1          11         21         31         41         51         61         71
AGDKLKKCKS SNLIILEEHL KNKKEFDHNS NIRYRCRGKE GWIHTVCING RWDPEVNCSM AQIQLCPPPP QIPNSHNMTT
81         91         101        111
TLNYRDGEKV SVLCQENYLI QEGEEITCKD GRWQSIPLCV
```

Figure 3.47 Partial validation of FH-13-14 (A) Sequence coverage in peptide mass fingerprinting for FH-13-14. The amino acid sequence of FH-13-14 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points (for mass spectrum of peptide fingerprint of construct FH-13-14 see Appendices).

Due to the high levels of degradation and oligomerisations of the FH13-14 it was excluded from functional and structural studies.

3.1.16 Construct FH-13

The KM71H *P. pastoris* construct over-expressing FH-13 was prepared by Claire Egan.

FH-13 was expressed in shaker-flasks (Figure 3.48-A). Following batch-absorption (SP-Sepharose) and EndoH$_f$-treatment sample was desalted into 20 mM sodium acetate buffer, pH 5.0. Pre-purified sample was loaded onto the MonoS cation-exchange column. Immobilised material was eluted in a linear gradient over 15 CV to 1 M NaCl (Figure 3.48-B). The elution profile shows two slightly overlapping peaks. SDS-PAGE analysis of the fractions spanning both peaks bands of similar mobility. Fractions corresponding to the second peak exhibit some degradation. The appearance of two baseline-resolved peaks likely reflects the aforementioned "ragged" N-terminal sequence. The fractions exhibiting a single band were pooled and concentrated.

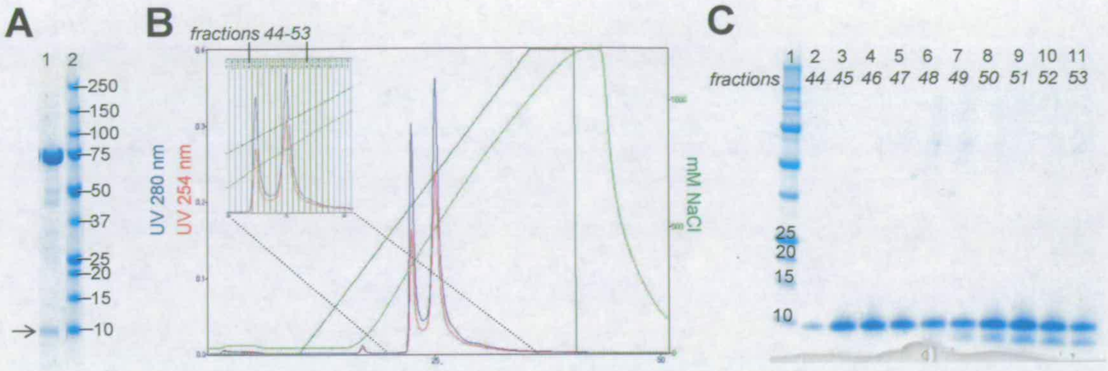


Figure 3.48 Expression and purification of FH-13. (A) Shaker-flask expression of FH-13. The band corresponding to FH-13 is marked by an arrow. (B) Cation-exchange chromatogram of pre-purified FH-13 (C) Gradient SDS-PAGE of fractions indicated in B).

This clone was used to express FH-13 that was used for determination of the high resolution NMR structure for by Chris Fenton. This exercise clearly validates the integrity and structure of recombinantly expressed FH-13.

3.1.17 Construct FH-13-15

The FH-13-15 coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (Figure 3.2).

A small-scale expression trial of clones selected on YPDS plates containing 150 µg/ml Zeocin is shown in Figure 3.49-A. The presence of N-glycosylation sites necessitated enzymatic removal of the Asn-linked glycans.

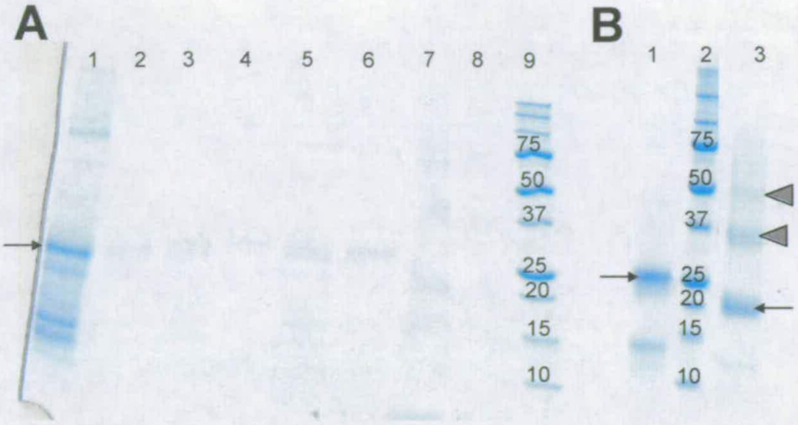


Figure 3.49 Gradient SDS-PAGE of *P. pastoris* FH-13-15 expression trials. (A) Mini-scale expression trial of six clones expressing FH-13-15 (highlighted by an arrow); reducing and non-reducing conditions apply to lanes 1-6, and 7-8, respectively. (B) EndoH$_f$-treated supernatant from shaker-flask expression. The dominant band (highlighted by an arrow) shows the mobility expected for FH-13-15. Reducing and non-reducing conditions apply to lane 1 and 3, respectively. Triangles indicate bands (non-reducing conditions) corresponding to mobilities of dimerised and trimerised FH-13-15.

FH-13-15 was expressed in shaker-flasks. EndoH$_f$-treated supernatant was run on a SDS-PAGE gel (Figure 3.49-B). Some degradation (especially under reducing conditions) is visible. Under non-reducing conditions degradation bands appear to be absent or less pronounced. This effect was observed for some other constructs and is discussed above. Similarly to the observation for FH-10-15, FH-11-14 and FH-13-14 a proportion of dimerised and trimerised FH-13-15 is visible on the gel, under non-reducing conditions. Following harvest the supernatant was diluted 1:1 and passed over two in-line connected HiTrap™ Heparin columns (dimensions of each: 7 mm x 25 mm - Amersham Biosciences®) equilibrated with 20 mM sodium acetate buffer, pH 5.0. Bound material was eluted in a linear gradient over 20 column volumes to 1 M NaCl (Figure 3.50-A). FH-13-15 eluted in a broad peak, likely reflecting the presence of multiple FH-13-15 species (dimers and trimers). For separation of oligomers the eluted protein was concentrated and loaded onto the HiLoad Superdex 75 size-exclusion column (equilibrated with 20 mM potassium phosphate buffer, pH 7.4, 500 mM sodium chloride). Two smaller peaks eluted prior to the main peak (Figure 3.50-B). SDS-PAGE under non-reducing conditions showed that the second highest peak, which elutes before the main peak, corresponds to dimerised FH-13-15, while the main peak shows a band corresponding to monomeric FH-13-15 (Figure 3.50-C).

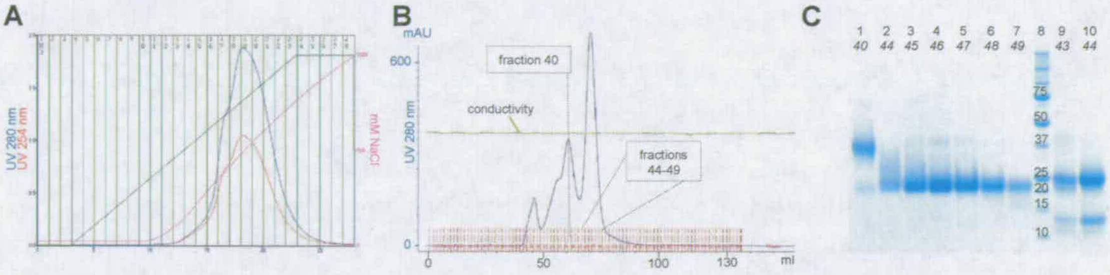
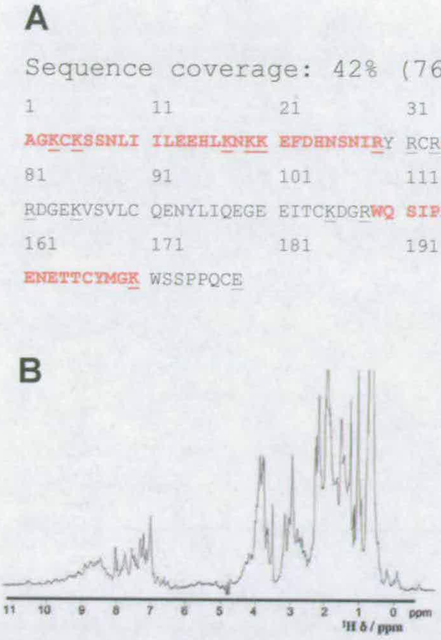


Figure 3.50 Purification of FH-13-15 (A); diluted and deglycosylated FH-13-15 fractionated by heparin-affinity chromatography. (B) Concentrated and pre-purified FH-13-15 was loaded onto the size-exclusion column. (C) Gradient SDS-PAGE of fractions indicated in B). Non-reducing and reducing conditions apply to lanes 1-7, and 9-10, respectively.

Under reducing conditions a distinct degradation band appeared in SDS-PAGE analysis. While a similar phenomenon was observed for other constructs, the relatively high level of degradation shown for FH-13-15 is unusual. Also the tendency to oligomerise, as judged by the proportion of monomeric to dimeric and trimeric species (on a non-reducing gel – sample taken from crude supernatant after harvesting), is observed to be higher for FH-13-15 than for any other construct. Fractions 47 to 49 were pooled and concentrated. Partial validation of primary structure was obtained from a tryptic-digest MS (Figure 3.51). Acquisition of a 1-D ^{1}H NMR experiment revealed upfield-shifted methyl peaks in the ^{1}H-NMR spectrum implying that folded material pre-dominates. However, the "lumpy" appearance of the spectrum is consistent with some degree of aggregation and/or the presence of some improperly folded material. These findings are consistent with the enhanced tendency of oligomerisation and degradation observed for FH-13-15.



A

```
Sequence coverage: 42% (76 of 178 amino acids)
1          11         21         31         41         51         61         71
AGKCKSSNLI ILEEHLKNKK EFDHNSNIRY RCRGKEGWIH TVCINGRWDP EVNCSMAQIQ LCPPPPQIPN SHNMTTTLNY
81         91         101        111        121        131        141        151
RDGEKVSVLC QENYLIQEGE EITCKDGRWQ SIPLCVEKIP CSQPPQIEHG TINSSRSSQE SYAHGTKLSY TCEGGFRISE
161        171        181        191        201        211        221        231
ENETTCYMGK WSSPPQCE
```

B

Figure 3.51 Validation of FH-13-15 (A) Sequence coverage in peptide mass fingerprinting for FH-13-15. The amino acid sequence of FH-13-15 is shown; highlighted in red are matched peptide masses, underlined are trypsin cleavage points (for mass spectrum of peptide fingerprint of construct FH-13-15 see Appendices). (B) ^{1}H-NMR spectrum of FH-13-15; Spectrum shows some degree of peak broadening and generally has less sharp and well resolved peaks. The NMR sample is 100 μM in 20 mM potassium phosphate buffer, pH 6.6.

3.1.18 Construct FH-14

The FH-14 coding sequence was cloned into the pPICZαB vector and transformed into *P. pastoris* (Figure 3.2). A small-scale expression trial (Figure 3.52) was performed on five colonies that were selected on YPDS plates containing 150 μg/ml Zeocin. Deglycosylated samples were analysed by SDS-PAGE. No protein band was visible from any of the five selected clones. To compare expression of FH-14 to other FH deletion-module constructs a mini-scale expression trial of different constructs was performed in parallel. The clones of all constructs subjected to this mini-scale expression (FH-11, FH-11-14, FH-12, FH-12-13 and FH-14) have been selected on agar plates that contain 150 μg/ml Zeocin. Supernatant from FH-12, FH-14, FH-11-14 and FH-12-13 cultures were deglycosylated prior to spin-concentration and SDS-PAGE. Only the supernatant of construct FH-14 was concentrated by 50-times; the supernatant from all other constructs was concentrated 10-fold.

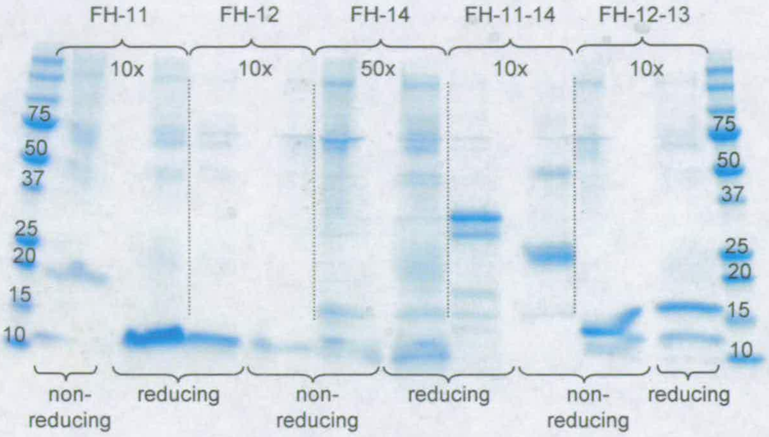

FH-11
FH-12
FH-14
FH-11-14
FH-12-13
10x
10x
50x
10x
10x
75
50
37
25
20
15
10
75
50
37
25
20
15
10
non-reducing
reducing
non-reducing
reducing
non-reducing
reducing

Figure 3.52 Expression-level comparison on gradient SDS-PAGE. Constructs and loading conditions (reducing or non-reducing) are indicated.

In spite of a five-fold higher concentration level (also reflected in the denser background of *P. pastoris* proteins) no clear band corresponding to the expected size of a single CCP module was visible in the lanes of FH-14. For a conclusive assessment as to why FH-14 could not be expressed in this study, more detailed analyses and higher sample numbers would be needed. It is hypothesised, however, that the lack of expression of FH-14 could reflect problems in folding when expressed as a single module. Improperly folded material would not be directed to the secretion-pathway, but would be sent for recycling to the proteasome.

3.2 Conclusions – protein production, purification and characterisation

Structural and biophysical studies of proteins are critically dependent on the provision of 10^{-4}-10^{-2} g-quantities of pure proteins. When the target protein is large and flexible with multiple binding sites, the ability to produce more experimentally tractable truncated versions, via genetic engineering, is paramount. For NMR purposes, isotopic labelling is a virtual requirement mandating the use of organisms with simple nutritional requirements. In the case of extracellular proteins with disulfide linkages, bacterial protein production is notoriously problematic. These considerations formed the background to the work of this chapter that has illustrated the utility of *P. pastoris* as an expression vector for production of multiple-module segments of FH.

Mammalian, insect, yeast and bacterial vectors have all been applied to the production of proteins and protein-segments from the RCA family (reviewed in (2)). Mammalian and insect systems suffer from the drawback that isotopic enrichment is expensive; disulfide-containing proteins expressed in bacterial systems often require re-folding.. Yeast has proved the method of choice in this laboratory, and the methylotropic yeast *P. pastoris* – with the option of fusing the target protein to a cleavable secretory signal peptide - has proved to be particularly advantageous. The double-CCP protein FH-19-20 had previously been expressed in large quantities in *P. pastoris* (51), and shown to be properly folded and functional; more recently the *P. pastoris*-produced proteins FH-1-2, FH-2-3 and FH-1-3 were used in a detailed structural analysis of the N-terminal region of FH (148). The current work greatly extends these protein production efforts and demonstrates that this vector is also suitable for producing proteins containing many more modules – as many as eight in the case of FH-8-15. Moreover, these recombinant proteins are folded as judged by 1-D and 2-D NMR spectra.

The results described here also illustrate the potential difficulties that arise from the use of *P. pastoris* to express multiple-CCP module products. Some logistical issues such as the relative complexity of cloning (compared to bacteria), the multiple days required for cell culture, and the large volumes of supernatant that require processing have to be factored into work-schedules but do not constitute insurmountable barriers to progress. The tendency of *P. pastoris* to attach high-mannose sugars to N-glycosylation sites is a disadvantage but this can be dealt with by mutagenesis or by post-harvest removal of glycans (leaving a GlcNAc stub) with EndoH$_f$ (although there is an associated risk of proteolysis). There are few reports of selective isotopic labelling in *P. pastoris* of the sort used in recent years for isotopic enrichment of larger proteins in *E coli*, so this is a limitation at least until similar labelling strategies are devised for the higher organism. Finally, several of the protein products exhibited a tendency to form dimers, trimers and tetramers some of which appeared, puzzlingly, to survive the process of SDS-PAGE under non-reducing conditions. Thus it proved impossible to produce the target segments FH-13-14 and FH-14.

Despite these hurdles, the current work resulted in a panel of validated proteins corresponding to segments of FH. Together with segments produced in previous work in this lab, these provide "coverage" of some 85% of the factor H molecule. They thus form a uniquely valuable resource for the functional and structural work described in the next two chapters.

CHAPTER 4

FUNCTIONAL STUDIES[13]

---

[13] Parts of this chapter have been published in (1), (3).

4.1 Perspective

A key biological role of FH is to ensure that the alternative pathway of complement activation operates selectively on pathogenic surfaces (as opposed to self-surfaces). The principal ligands involved in this process of self versus non-self discrimination are surface-bound C3b molecules, and polyanionic carbohydrate structures (glycosaminoglycans (GAGs) and sialic acid). This work set out to establish structure-function relationships for FH; the previous chapter described the production of a set of recombinant proteins representing various regions of the FH molecule. In this chapter, the functional characterisation of these proteins is described. In order to investigate the carbohydrate ligand of FH, heparin was employed for most studies. Heparin itself is a soluble (rather than a cell-surface) GAG and is heterogeneous, but it is widely used as a model compound in protein: GAG interaction studies due to its ready availability. It consists predominantly of a repeating disaccharide unit consisting of 2-O-sulfated iduronic acid and 6-O-sulfated, N-sulfated glucosamine. Levels of sulfation are variable but, overall, heparin is one of the most negatively charged molecules in nature. It resembles the abundant surface-bound GAG, heparin sulphate. In the current study, either heparin-affinity columns were purchased and used in chromatographic assays, or enzymatically digested and size-fractionated preparations of heparin were employed in gel mobility shift assays (see METHODS). For studies of C3b binding, commercially available C3b protein molecules were immobilised on sensor chips and SPR (see METHODS) was utilised to measure affinities. Although the chip surface is only an approximate representation of the cell surface it nonetheless allows a reproducible and quantitative assessment of protein-ligand interactions.

4.1.1 Analysis of Heparin binding sites in FH

First, it was attempted to confirm a previous report that FH CCP 9 harbours a GAG-binding site (61).

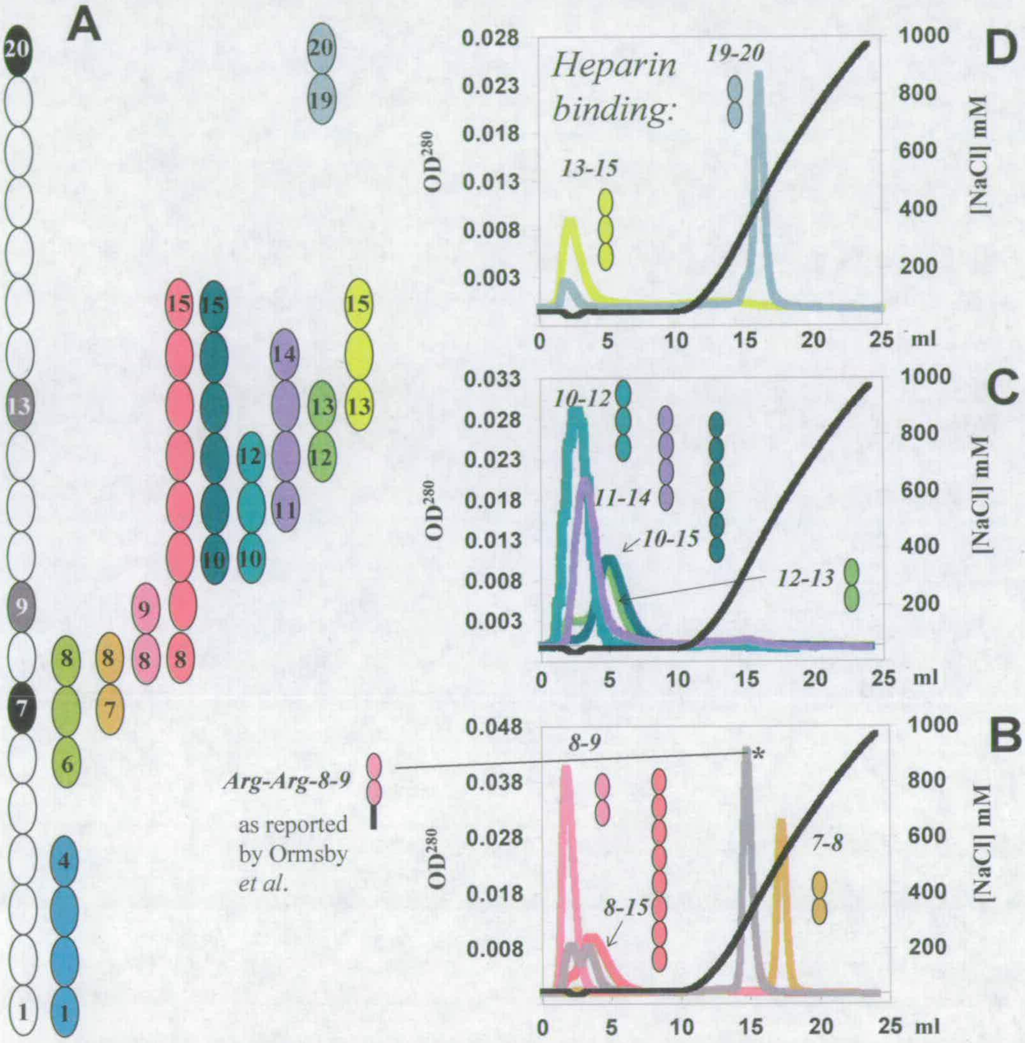


Figure 4.1 Factor H segments employed in this study and results of heparin-affinity chromatography. (A) Schematic of FH with CCPs implicated in heparin binding shaded: black, well-established sites in CCPs 7 and 20; grey, putative sites in CCPs 9 and 13 under investigation here. Recombinant protein constructs employed in the present study are also drawn. Profiles of FH segments on a HiTrap heparin-affinity column are shown in (B-D). Ten proteins were chromatographed individually but plotted here on one of three frames, each with a representative trace to show the salt gradient applied: (B) FH-7-8, FH-8-9, FH-8-15, and FH-$_{RR}$8-9 (indicated by *); (C) FH-10-12, FH-10-15, FH-11-14, and FH-12-13; and (D) FH-13-15 and FH-19-20.

Fully characterized and authenticated FH-8-9 eluted from a HiTrap column (Figure 4.1-B) in 20 mM phosphate buffer (pH 7.4) with no additional salt; on the other hand, >150 mM NaCl was required to elute the positive controls, FH-7-8 and FH-19-20 (Figure 4.1-B and -D). In a consistent result, FH-8-9, unlike the positive controls, failed to retard the mobility of a range of purified, defined-length heparin-derived oligosaccharides in a GMSA (Figure 4.1-A and -C).

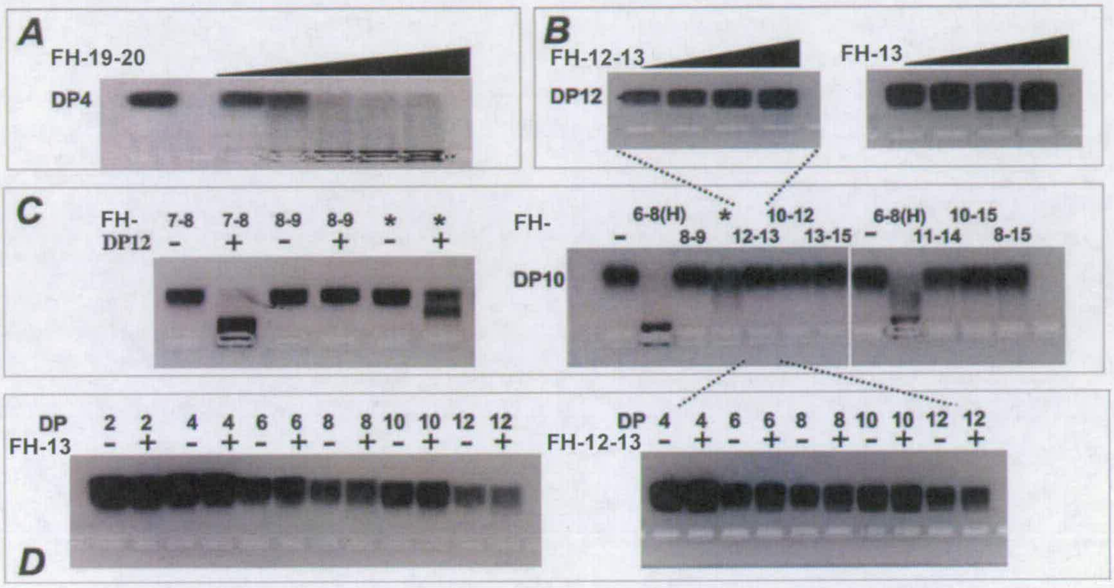

A
FH-19-20
DP4
B
FH-12-13
FH-13
DP12
C
FH- 7-8 7-8 8-9 8-9 * *
DP12 - + - + - +
FH- 6-8(H) * 10-12 6-8(H) 10-15
- 8-9 12-13 13-15 - 11-14 8-15
DP10
DP 2 2 4 4 6 6 8 8 10 10 12 12
FH-13 - + - + - + - + - + - +
DP 4 4 6 6 8 8 10 10 12 12
FH-12-13 - + - + - + - + - +
D

Figure 4.2 Gel-mobility shift assays. In the GMSA, electrophoretic migration toward the anode (upward in this figure) of fluorescently labelled heparin oligosaccharides (1 μg), of defined degree of polymerization (DP, *i.e.* number of sugar units), can be retarded by binding to equimolar amounts (unless otherwise stated) of the indicated FH segments. Note that the resulting fluorescent protein-heparin complex often stays in, or close to, the well, resulting frequently in loss of fluorescent intensity (relative to the free oligosaccharide) upon subsequent gel handling. (A) Increasing ratios of FH-19-20 to DP4 demonstrate retardation of sugar migration by this positive control (lanes contain, from left to right, no protein, blank, 0.7:1, 1.4:1, 2.8:1, 7.0:1, and 13.8:1 molar ratios of protein-DP4). (B) Even at a 4:1 ratio of protein to sugar, neither FH-13 nor FH-12-13 bind to DP12 (lanes contain 0:1, 1:1, 2:1, and 4:1 ratios, from left to right). (C) The GMSA shows clearly that FH-7-8 binds to sulfated heparin fragments DP12 while FH-8-9 does not. The exact construct reported by Ormsby *et al.* (61), signified by an asterisk, retards migration, although less markedly than does FH-7–8. Unlike positive control FH-6-8$_{H402}$, none of the segments between CCPs 8 and 15 retards migration of more than trace amounts of DP10. (D) Neither FH-13 nor FH-12-13 binds to longer fragments of heparin (up to DP12).

This lack of detectable affinity for heparin by FH-8-9 is in apparent contradiction to the previously published report of significant binding of heparin by several constructs that encompass CCP 9 (61). Explanations based on mis-folding of FH-8-9 may be eliminated on the basis of its NMR-authenticated structure (see CHAPTER 3). To investigate further, the same sequence as was used in the key experiment of the previous published study was analysed: *i.e.* FH-$_{RR}$8-9, which incorporates an N-terminal sequence artefact containing two Arg residues. This construct does indeed bind to a HiTrap heparin-affinity column significantly better than does FH-8-9 (Figure 4.1-B) and almost as tightly as the positive controls. It also binds to sulfated heparin fragments according to GMSA (Figure 4.2-C). The clear implication is that the extraneous di-basic sequence contributes non-specifically to the heparin-binding affinity of FH-$_{RR}$8-9. To investigate further, a synthetic peptide of sequence EFTWPSRPSRIGTKT[14] was tested for binding to heparin-affinity resin. This sequence matches the non-native sequence at the N terminus of FH-$_{RR}$8-9 plus two native residues (Lys and Thr). The peptide was found not to have a strong affinity for heparin (data not shown). It is therefore concluded that the heparin binding site in FH-$_{RR}$8-9 is a composite of non-native N terminus and native sequence.

Based on a previous report of heparin binding to CCP 13 or CCP 14 of FH (67), this potential GAG-interacting site was investigated further. Of note is that CCP 13 has more positively charged residues than do other CCPs in FH. Surprisingly, FH-12-13 failed to adsorb to either of the two heparin-affinity columns (see METHODS) used in this study at physiological salt concentration and pH (Figure 4.2-C). Nor did FH-12-13 bind to defined-length (from 4 to 12 sugar units) heparin-derived oligosaccharides in the GMSA (Figure 4.2-B), even at a 4:1 ratio of protein to sugar. It was conceivable (prior to the determination of the 12-13 structure, see CHAPTER 5) that their long (eight-residue) intermodular linker allows CCPs 12 and 13 to arrange themselves side-by-side in the context of the isolated pair such that the putative GAG-binding site on CCP 13 is inaccessible. This possibility was eliminated by our observation that, when expressed as a single module, CCP 13 (FH-13) was unable to bind GAGs (Figure 4.2-B and D). It was also considered a possibility that residues from module 14, or from the 13-14 linking sequence, are required to complete a GAG-binding sub-site in CCP 13. This option was excluded by a study in which it was observed (Figure 4.1-C and Figure 4.2-C) that FH-11-14 is neither retained to a significant extent on a heparin column nor does it bind to heparin in a GMSA. Next, the possibility was considered that CCP 14 rather than CCP 13 might be central to a longer putative binding site. With this in mind, FH-13-15 was produced, but it was also found not to bind the heparin-affinity column or to produce more than a hint of binding according to GMSA (Figure 4.1-D and Figure 4.2-C). The presence (as judged by NMR) of some unfolded or aggregated material in the FH-13-15 sample would be most unlikely to explain this lack of binding by the majority of folded FH-13-15 material that is also present. Thus, these results show clearly that modules 13 and 14 of FH and their immediate neighbours do not constitute a discrete GAG-binding site comparable to the ones present in CCP 7 and CCP 20.

---

[14] Peptide EFTWPSRPSRIGTKT was purchased from Sigma.

Having failed to detect heparin binding within CCPs 8 or 9 or CCPs 11-15, intervening modules were chosen for examination. FH-10-12 was not retained on a heparin column nor was it positive by GMSA (Figure 4.1-C and Figure 4.1-C). It was reasoned that CCPs 12-13 (small modules joined by a long linker) might allow the FH molecule to fold back upon its self, allowing non-neighbouring modules to form a composite GAG-binding site (this was subsequently shown to be the case – see CHAPTER 5). To investigate this further, a construct that includes two CCPs on either side of FH-12-13 was analysed next. This construct – FH-10-15 - did not bind heparin with significant affinity (Figure 4.1-C and Figure 4.2-C). Finally, FH-8-15 was tested in order to find out whether CCPs 9 and 13 might individually be relatively weak GAG binders but nonetheless contribute to a common, higher affinity GAG binding site. Despite encompassing two previously reported GAG-binding modules, this longer construct does not have significant heparin affinity (Figure 4.1-B and Figure 4.2-C).

4.1.2 Comparison of heparin binding between the AMD-protective 402Tyr and the at-risk 402H allotypic variants in the contexts of CCP 7 and CCPs 6-8[15]

Since FH contains two strong GAG-binding sites (*i.e.* CCPs 7 and 20) that likely compete or cooperate in biochemical binding assays, the heparin binding properties of the binding site centred in CCP 7 were investigated in the context of the single-module variants FH-7_{Y402}/FH-7_{H402} and triple module variants FH-6-8_{Y402}/FH-6-8_{H402}. (A detailed study of FH-19-20:heparin interactions was undertaken previously in this lab by A. Herbert) (51).)

A mixture of the His and Tyr variants of triple module proteins (CCP6-8) could not be resolved on a heparin-affinity column (Figure 4.3).

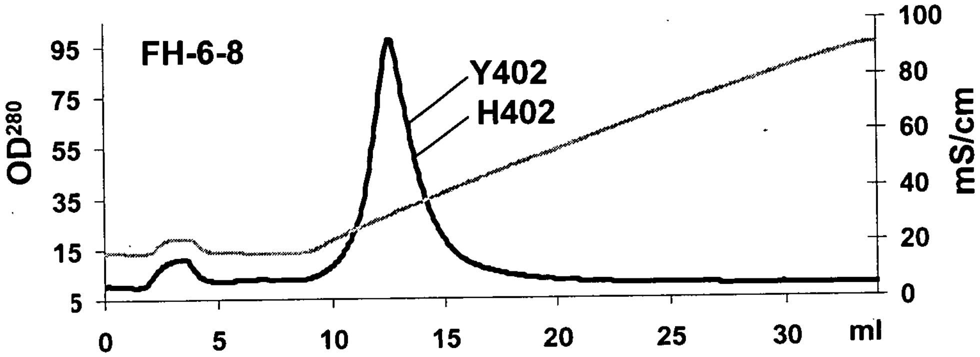


Figure 4.3 A mixture of His402 and Tyr402 CCP 6-8 (triple modules) applied on a Poros heparin-affinity column.

They also bound equally well to a panel of chemically pure, defined-size, fully sulfated heparin fragments (Figure 4.4-A). A similar observation was reported for these variants in the context of full-length FH (1). However, when the His402 and Tyr402 variants were tested in the context of isolated CCP 7, base-line resolution on a heparin-affinity column (pH 7.4) (Figure 4.4-B) could be achieved. Moreover, irrespective of oligosaccharide length, Tyr402 CCP 7 retards a higher proportion of heparin than His402 CCP 7 according to a GMSA (Figure 4.4-A).

---

[15] This work was undertaken together with and under guidance of Dr Andrew Herbert The GMSAs were performed by in the lab of Malcolm Lyon, Paterson Institute for Cancer Research, University of Manchester, UK.

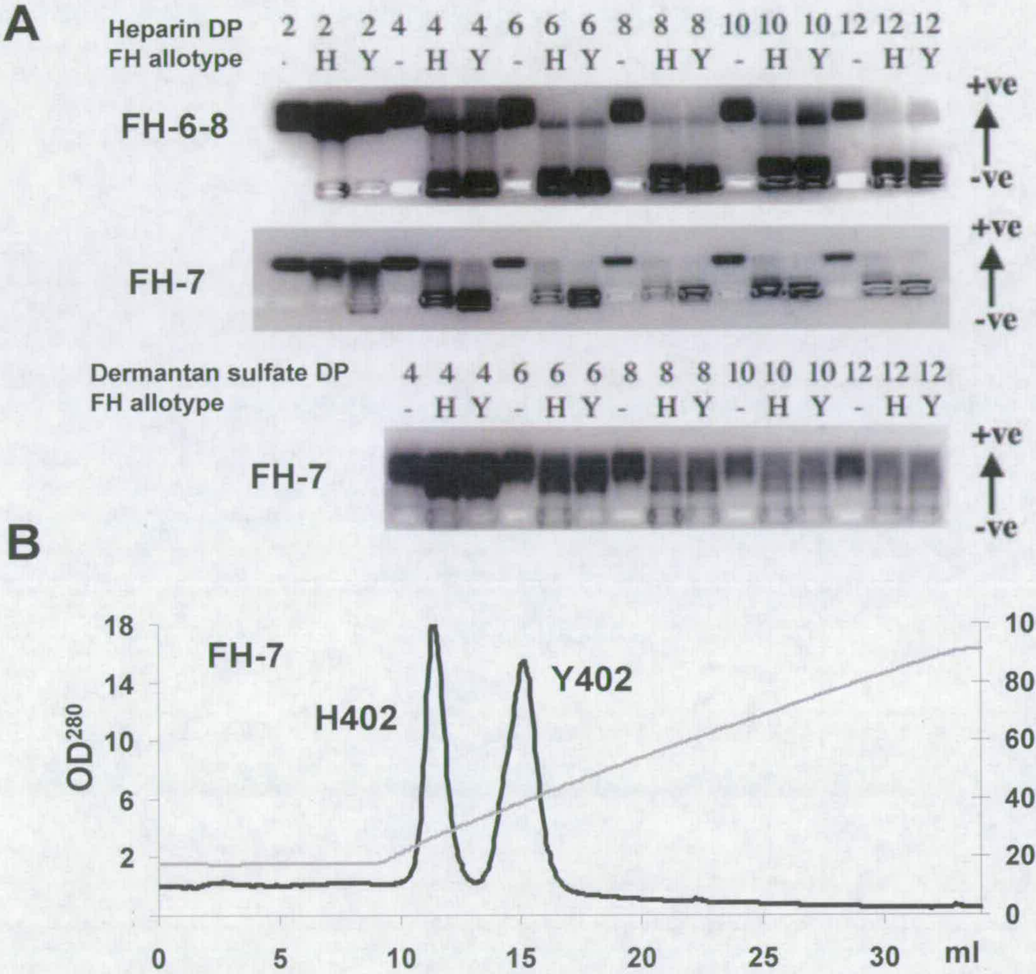


Figure 4.4 Affinity of FH-7 and FH-6-8 His402 and Tyr402 allotypes for GAGs (A) Comparison of GMSA data for His402 and Tyr402 allotypes of CCP 6–8 (*top*) or CCP 7 (*middle*) with defined-length heparin oligosaccharides (degree of polymerization (*DP*) values refer to the number of monosaccharide units). A four-fold molar excess of FH-7 was added to 0.5 g of oligosaccharide, or equimolar amounts of FH-6-8 were added to 0.5 g of oligosaccharide. *Bottom*, a comparison of GMSA data for His402 and Tyr402 allotypes of FH-7 with defined length dermatan sulfate oligosaccharide; equimolar amounts of FH-7 were added to 0.5 g of

oligosaccharide. (B) A mixture of His402 and Tyr402 CCP 7 (single modules) separated on a Poros heparin-affinity column.

This effect is most pronounced for shorter oligosaccharides (di- and tetrasaccharide). A simple non-specific charge effect seems unlikely since the Tyr402 allotype of CCP 7 binds more strongly than the potentially positive His402 variant. Although CCP 7_{Y402} binds heparin more tightly than does CCP 7_{H402}, such discrimination is not manifested towards a panel of size-fractionated dermatan sulfates (in a GMSA experiment (Fig. 77-A)). This observation indicates that there are GAG-specific differences between the binding properties between the two allotypes. These results are discussed further in CHAPTER 6.

4.2 Analysis of C3b-binding sites

1.2.1 Mapping of C3b-binding sites of FH

When acting as a complement regulator on self surfaces, FH presumably binds simultaneously to both GAGs and C3b. The affinities for C3b of various segments of FH, employing constructs FH-1-4 and FH-19-20 as positive controls, were measured by SPR. For these exploratory experiments K_d values were not determined, but affinities were compared based on the number of response units (RU) measured (that correlates with the amount of analyte that is bound to the sensor chip) when a 10 μM solution of the fragment was passed over the chip surface. Table 4.1 summarises the C3b-coated sensor-chips prepared for this work.

Chip (type)	**C3b loading (RU)$^\$$**			
	Flow cell 1	**Flow cell 2**	**Flow cell 3**	**Flow cell 4**
A (CM5)	0	4109	0	4104
B (CM5)	0	384	1593	3002
C (C1)	0	140	499	752

$^\$$RU = response units; Reference surfaces (0 RU) were prepared as described in Materials and Methods.

Table 4.1 Chip-surface summary. C3b was amine-coupled to three flow-cells of two CM5-chips and one C1-chip. The CM5-chip surface is characterised by a carboxymethylated dextran matrix that is covalently attached to the sensor-chip surface. The C1-chip surface is directly carboxymethylated and therefore matrix-free.

Figure 4.6-A illustrates the quality of the sensorgrams; Figure 4.6-B summarises the size of the response (normalized to the binding of FH-19-20) obtained both before and after adjustment to take account of the molecular weight of the analyte.

The construct FH-11-14 incorporates the previously inferred C3b-binding modules 12-14, yet its affinity for C3b was found to be insignificant compared with that of FH-1-4 or FH-19-20 (Figure 4.6). The presence of a putative C3b-binding site in modules 12-14 had been deduced from module-deletion experiments, the interpretation of which may have been complicated by neighbouring-module effects. To investigate this, FH-10-12 and FH-13-15 were also tested, but were found to have no affinity for C3b either. Nonetheless, the possibility of a composite binding site for C3b, requiring cooperation of non-contiguous stretches of modules, could not be eliminated on the basis of results obtained with shorter constructs. Therefore, the six-module construct FH-10-15 was tested for C3b binding. With this construct a response (at a target loading of 4200 RU) of < 4 RU was obtained, which was interpreted as reflecting negligible affinity (Figure 4.6). Finally, FH-8-15, the longest FH construct that has been expressed so far, was tested for C3b-binding affinity; some evidence of binding was detected when a 10 μM sample was passed over the sensor chip (Figure 4.6), although the response obtained was significantly smaller than that obtained from injection of even 1 μM positive controls. This implies that CCPs 8 or 9 contribute wholly or partly to a weak C3b binding site. On the other hand, the double module FH-8-9 was found not to have a measurable affinity for C3b according to this assay. It was therefore concluded that there is likely some degree of cooperativity between low-affinity sites in module 8 or 9 and another low-affinity site within the 10-15 region.

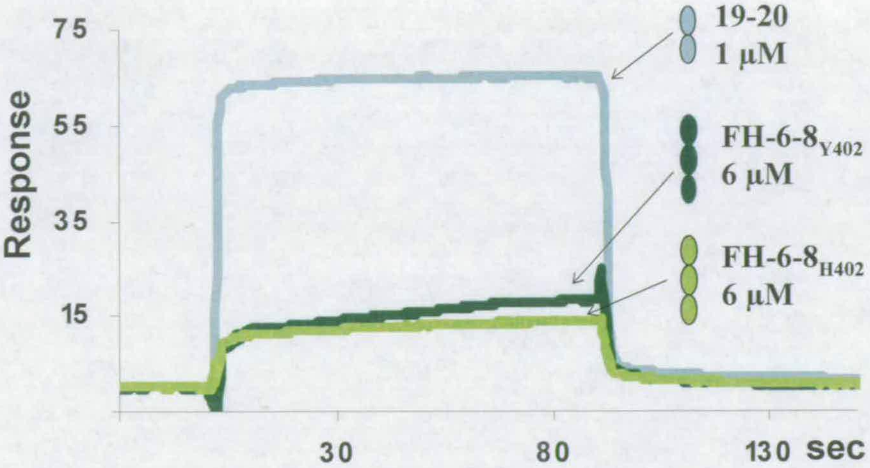


Figure 4.5 Surface plasmon resonance experiments to compare affinities for C3b binding of both allotypes of CCPs 6-8, 402Y and 402H. Duplicate traces recorded on flow cell 4 of chip A are shown (see Table 4.1).

A further weak C3b binding site was detected in CCPs 6-8. The relative binding of H402 and Y402 FH-6-8 was compared; these were found to bind with approximately equal strengths (Figure 4.5).

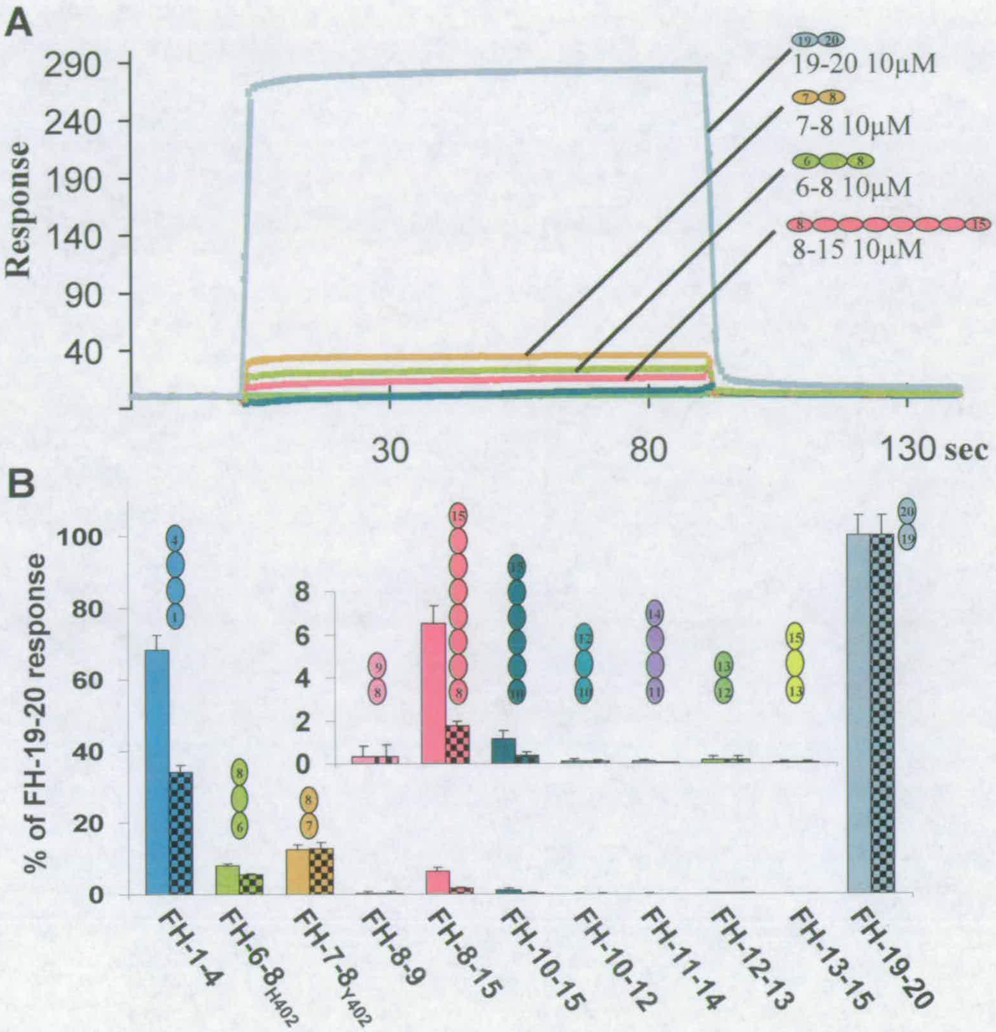


Figure 4.6 Surface plasmon resonance experiments to identify C3b-binding segments of FH. (A) The duplicate traces recorded during one experiment in which a series of 10 μM solutions of the segments indicated were flowed across flow cell 4 of chip A (see Table 4.1). These serve to illustrate data quality and reproducibility but also to highlight differences between the sizes of response for the various segments. Note that the traces for FH-10-15, FH-8-9, FH-12-13, FH-13-15, FH-10-12, and FH-11-14 are all very close to the baseline. (B) Bar charts to illustrate the strength of the response obtained for each FH segment on CM5 chips bearing immobilized C3b (*inset*: a 12-fold *y*-axis expansion of the responses from segments covering the CCPs 8-15 region). Coloured bars indicate averages of multiple readings, normalized to reflect differences in the target density (and expressed as a percentage of the response obtained for FH-19-20). Coloured bars with black diamonds show the same data following adjustment to take into account the direct correlation between a response and the M_r of an analyte. Error bars indicate SEs from the mean and are for four measurements (duplicate runs on two flow cells on chip A) with the exception of FH-1-4 (6 measurements on three flow cells of chip B) and FH-19-20 (10 measurements in total on five flow cells of chips A and B).

To investigate the role of module 8 more thoroughly, the C3b binding behaviour of FH-7-8 was assayed. Surprisingly, since the possibility has not received serious attention before, FH-7-8 (Y402) binds C3b (the H402 allotype of 7-8 was not examined). Taking into account its smaller mass (two modules in FH-7-8 compared with eight in 8-15), the sensorgrams indicate that significantly more FH-7-8 molecules bind to the C3b-coated chip compared with FH-8-15 under comparable conditions, although FH-7-8 binding is still much weaker compared with FH-19-20. This implies that module 7 is a more important contributor to C3b binding.

1.2.2 Further characterisation of major C3b-binding sites in FH

To further investigate the extent to which the binding to C3b of the major sites, CCPs 1-4 and 19-20 contribute to the FH-C3b interaction, K_d values were measured for FH-1-4 and FH-19-20 and compared to that of full-length FH. These experiments were repeated at several C3b densities and on two different types of sensor chip (Figure 4.7, Table 4.1 and Table 4.2 and APPENDICES). The averaged K_d values of FH-1-4 and FH-19-20 (as measured on the CM5 dextran-coated chip surface) are approximately eight- and two-fold, respectively, weaker than that of full-length FH.

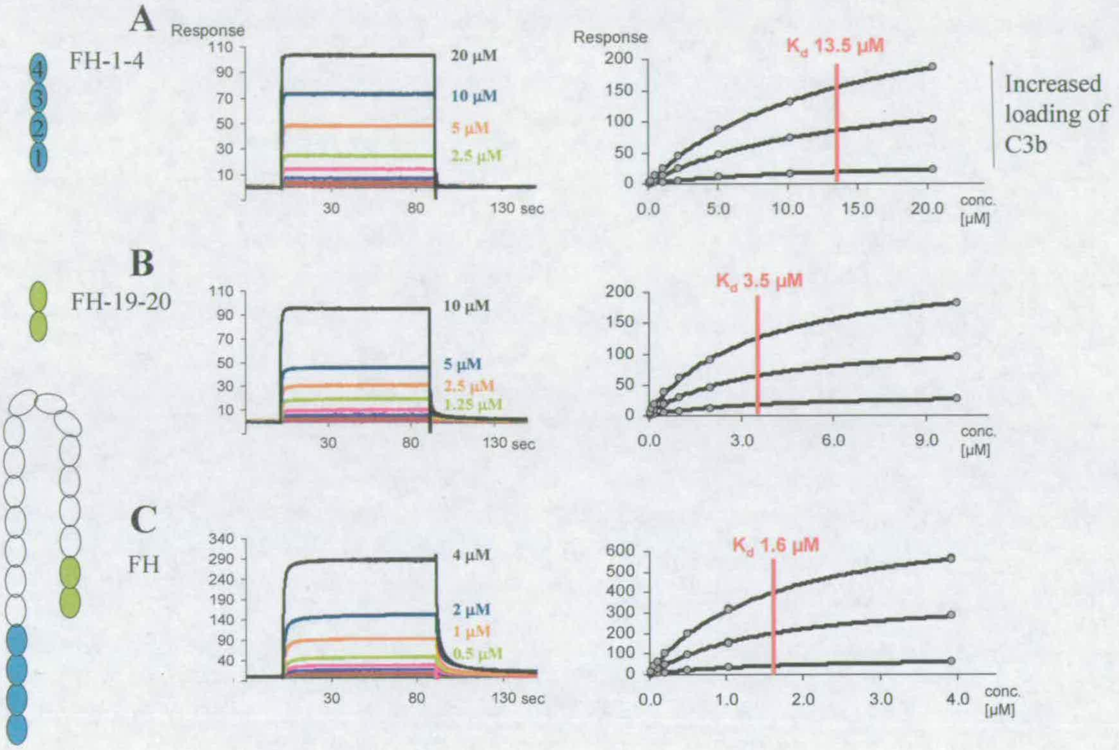


Figure 4.7 Use of SPR to measure dissociation constants for major C3b-binding sites in FH. Duplicate sensorgrams are shown for (A) FH-1-4, (B) FH-19-20, and (C) full-length FH at a range of analyte concentrations (FH-1-4, 0.05-20.4 μM; FH-19-20, 0.05-10 μM; FH, 0.01-3.93 μM) (left panels). These data are illustrative and show results obtained using flow cell 3 of chip B (CM5) (see Table 4.1). Right panels; plots of the response obtained vs. analyte concentration at each of three C3b densities (as indicated) on chip B (CM5). The equivalent data (FH-1-4, 0.05-102 μM; FH-19-20, 0.05-50 μM; FH, 0.01-3.93 μM) obtained on chip C (C1) are not shown here (see APPENDICES), but all K_d measurements for C3b are summarized in Table 4.2.

These results are consistent with a modest avidity effect arising from the simultaneous occupancy of the two major C3b-binding sites. To investigate further, these measurements were repeated on a C1 chip, and the loading on the C1 chip of C3b was minimized in an attempt to achieve a situation where adjacent molecules of C3b in appropriate orientations are unlikely to be available for binding to a single FH molecule. The K_d values obtained on the C1 chip are comparable to those measured on the CM5 chip, but display a slightly larger avidity effect (Table 4.2).

Chip (type)	**Protein**	**/ K_d ± S.E.M (μM)**[$]			
		Flow cell 2	**Flow cell 3**	**Flow cell 4**	**Combined**[&]
B (CM5)	**FH-1-4**	14.5 ± 0.5	13.7 ± 0.3	13.5 ± 0.3	13.5 ± 0.2
	FH-19-20	4.7 ± 0.3	3.7 ± 0.2	3.4 ± 0.2	3.5 ± 0.1
	FH	2.2 ± 0.1	1.8 ± 0.1	1.5 ± 0.1	1.6 ± 0.1
C (C1)	**FH-1-4**	10.0 ± 0.3	9.3 ± 0.5	10.0 ± 0.5	9.8 ± 0.3
	FH-19-20	7.8 ± 1.7	4.6 ± 0.8	4.4 ± 0.8	4.5 ± 0.5
	FH	0.70 ± 0.06	0.63 ± 0.07	0.56 ± 0.07	0.59 ± 0.04

[$] Derived K_d values are plus or minus the standard error of the mean (S.E.M.) calculated from the data exemplified in Figure 4.7, as described in Methods.

[&] Obtained by combining data from all three of the flow cells, *i.e.* at three different C3b loadings, on the sensor chip.

Table 4.2 Derived K_d values for interaction of C3b with FH, and with FH fragments that correspond to its two major binding sites.

4.3 Analysis of the binding sites for C3b-cleavage products, C3c and C3d

In a preceding section it was shown that the central modules of FH (embracing its CCPs 9-15) do not harbour any discrete binding sites for C3b according to SPR. *In vivo*, C3b gets cleaved by factor I (in the presence of FH and other cofactors) creating the C3d/C3dg fragment (that corresponds to the thioester domain of C3/C3b and remains bound to the surface) and the larger C3c fragment representing the bulk of the original C3 molecule (see CHAPTER 1). The cleavage process reveals previously buried regions of these molecules that serve as nascent interaction sites for a range of receptors. In a previous paper that reported the "third" C3b binding site in FH, a case was made for the C3c-portion of C3b being the binding site of CCPs 12-14 region of FH (60). Thus it remained possible that the central modules of FH bind better to C3c than intact C3b.

To investigate this possibility further an experiment was designed based on SPR. To ensure that a direct comparison could be made between C3b- and C3c-binding affinities, these molecules were coated (by amine coupling) onto different flow-cells (flow-cell 2 = C3c, flow-cell 4 = C3b) of the same CM5 chip. As a further control, C3d was immobilised on the surface of the third flow-cell. An effort was made to achieve approximately equal densities of protein molecules on all three chip-surfaces resulting in loadings of 2984 RU for C3b, 2371 RU for C3c and 509 RU for C3d. These translate to mole-ratios of 1.04:1 and 0.88:1 for C3c and C3d, respectively, versus C3b. That the recombinant FH segments and the amine-coupled C3b were interacting in the expected way was confirmed by the results for flow-cell 4 (Figure 4.8) - FH-19-20 gives a larger response than FH-1-4, while FH-6-8 gives a weak response and FH-8-15 a still weaker one. The results for C3d (immobilised on flow-cell 3) are also consistent with all the previous work - only FH-19-20 binds appreciably to this fragment. These positive and negative controls increased confidence in the reliability of the results obtained for C3c. In the case of this fragment, however, no appreciable binding was recorded for any of the FH segments, including FH-8-15 that incorporates the previously reported C3c-binding site.

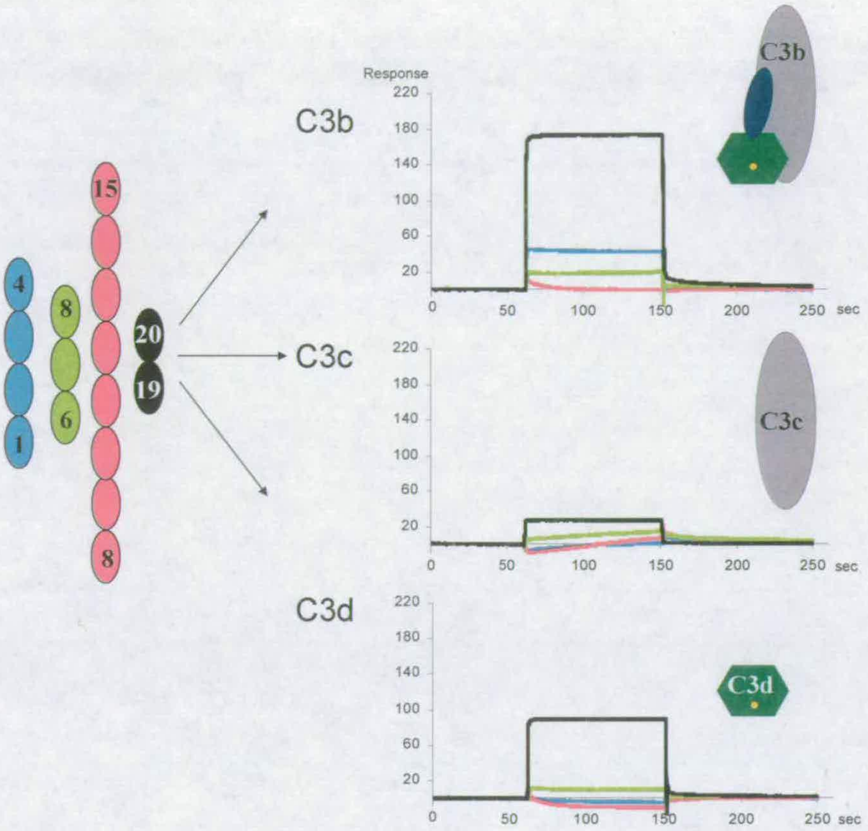


Figure 4.8 Surface plasmon resonance experiments to identify C3b-, C3c- and C3d-binding segments of FH. (A) The duplicate traces recorded during one experiment in which a series of 10 μM solutions of the segments (colour coded) were flowed across three flow cells coated with 2984 RU of C3b, 2371 RU of C3c and 509 RU of C3d. These not only confirm the findings for C3b, but also highlight differences between the sizes of response of the various FH segments for different fragments of C3.

In an additional experiment, an attempt was made to measure the affinity of full-length FH for immobilised C3d. As before, C3b, C3c and C3d were coated on separate flow-cells of the same CM5 sensor chip. The positive control (FH *versus* C3b) demonstrated that these molecules were behaving as expected (Figure 4.9) *i.e.* that the FH employed in the study had functional integrity, and that the SPR technique was working. Interestingly though, only a weak interaction between FH and C3d could be detected under these conditions.

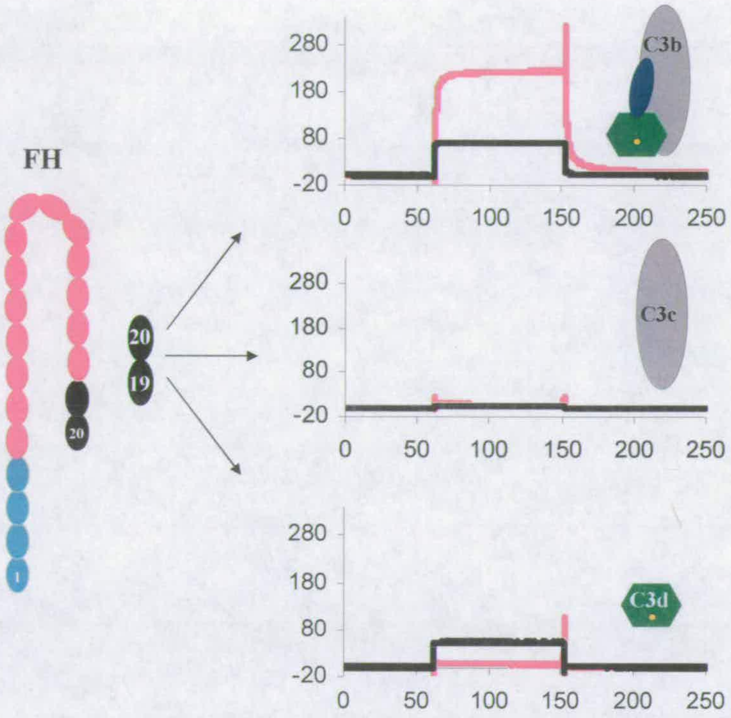


Figure 4.9 Surface plasmon resonance experiments with FH and FH-19-20 to identify affinities for C3b, C3c and C3d. The duplicate traces recorded during one experiment in which a series of 5 μM solutions were flowed across three flow cells coated with 1494 RU of C3b, 1191 RU of C3c and 310 RU of C3d. These translate to ratios of 1.04:1 and 1.07:1 for C3c and C3d, respectively, *versus* C3b.

Given the consequent inability to measure a K_d for C3d of intact FH, it was decided to perform this measurement using FH-19-20[16]. For this experiment, a C1 sensor chip was coated with three densities of C3d molecules (33 RU, 102 RU and 153 RU). A range of concentrations of FH-19-20 was passed over the chip allowing the derivation of the $K_d = 6.9$ μM. This may be compared to the K_ds previously measured for the FH-19-20:C3b interaction (3.5 μM and 4.5 μM for CM5 and C1 chips). The comparison implies that a significant portion of the binding contacts in the C3b-FH complex are provided by the C3b(TED)-FH-19-20 interactions (discussed further below).

---

[16] This experiment was performed together with, and under the guidance of Dr. Andrew Herbert. FH-19-20 was provided by Dr Andrew Herbert.

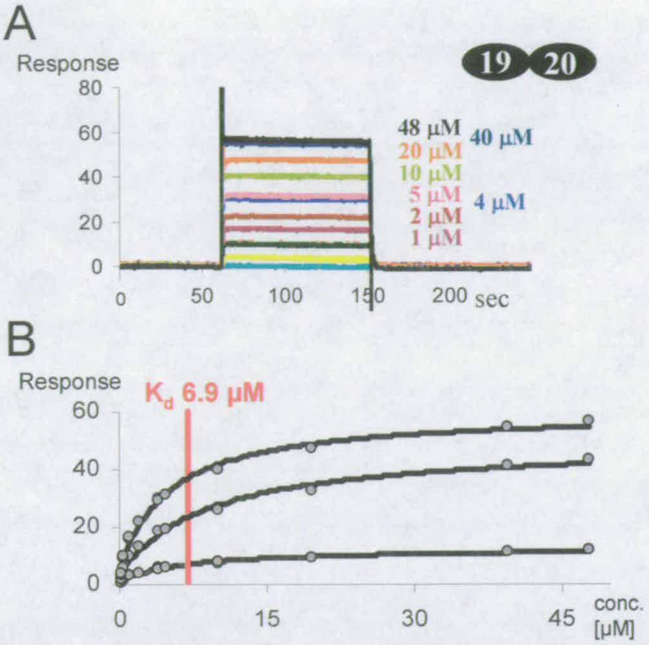


Figure 4.10 Use of SPR to measure FH-19-20 dissociation constant for C3d. (A) Duplicate injections of FH-19-20 at a range of analyte concentrations (0.1-48 µM). These data are illustrative and show results obtained using flow cell 4 (C1-chip-surface coated with 154 RU of C3d). (B) Plots of the response obtained *vs.* analyte concentration at each of three C3d densities (33 RU, 102 RU and 154 RU of C3d) on a C1-chip surface.

Taken together, these results provide a self-consistent picture of FH engaging through its N- and C-terminal regions with two distinct sites on C3b, at least one of which (the FH-19-20 target) is in the TED region. Both of these sites for full-length FH are apparently lost upon cleavage of C3b (despite the fact that affinity for the cleaved-off C3d is detectable in FH-19-20 when examined in isolation from the rest of the FH molecule).

4.4 Conclusions

Factor H has an apparently simple structure – being composed entirely from CCP modules – but has a sophisticated function that involves engagement with both protein and carbohydrate ligands. Previous work had suggested a multitude of binding sites on FH for its two principal binding partners – GAGs and C3b - that were difficult to reconcile with a plausible mechanistic model. This chapter described a systematic attempt to map binding sites by examining the binding abilities of recombinantly expressed segments, covering 85% of the FH molecule, which have been carefully checked for purity and structural integrity (see CHAPTER 3).

The current work is consistent with multiple literature reports that identified modules 6-8 and modules 19-20 of FH as being important for binding to GAGs (the potential differences between GAG-binding by the Y402 and H402 allotypes are considered in the DISCUSSION). Indeed these well-established sites served as useful positive controls for the two methods employed here to measure affinity for GAGs – heparin chromatography and GMSA. Reassuringly, both methods gave very similar results. Thus the lack of GAG-binding, inferred from the present study, by CCPs in the central portion of FH is strong evidence that this region of the regulator interacts weakly, if at all, with the polyanionic carbohydrates in the vicinity of surface-bound C3b. This is despite the contrary claims of previous studies in the literature and the highly positively charged nature of CCP 13. Indeed, in the case of FH-8-9 the reported heparin-binding site (61) was clearly shown to be an artefact. There remains the possibility (considered in the DISCUSSION) that, by focussing mainly on heparin as a model GAG, important interactions have been missed. Nonetheless, the simplest interpretation of the current results is that there are only two GAG binding sites in FH.

The current work also confirms the C3b-binding sites at the N- and C-terminal modules of FH; it furnishes K_d values for these sites with respect to C3b that is immobilised by amine-coupling to the surface of sensor chips, and provides evidence for cooperative binding of these two sites to a single C3b molecule (Figure 4.11).

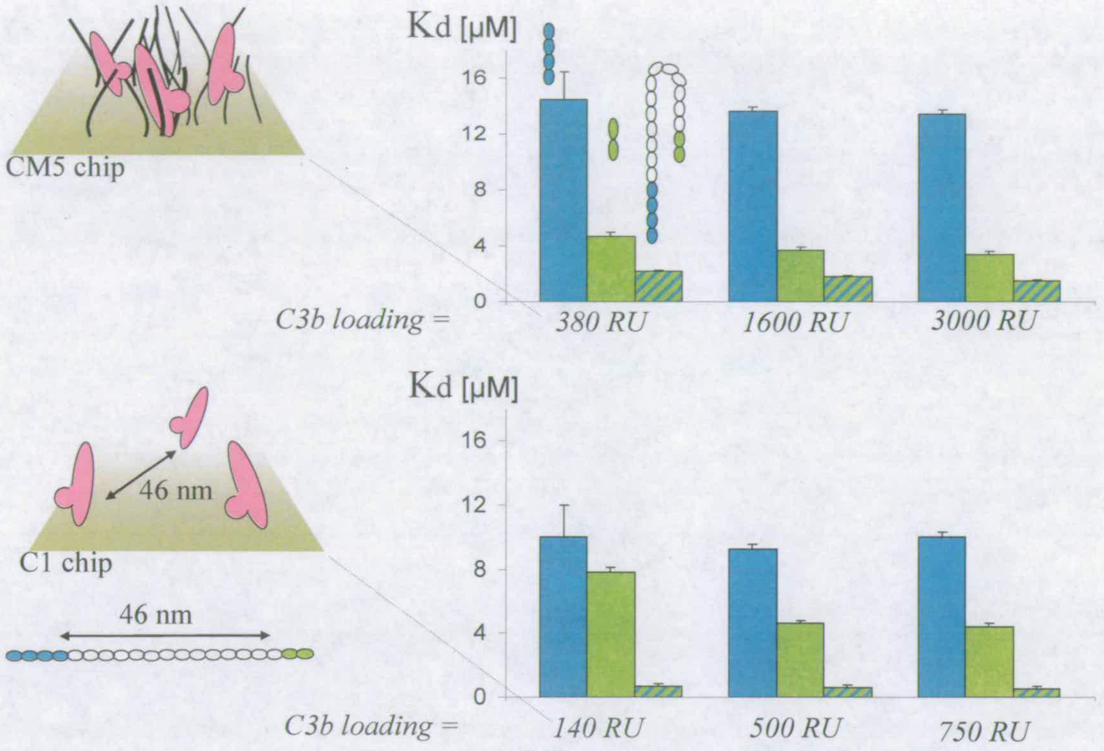


Figure 4.11 Bar-chart diagram summarising K_d values determined for C3b binding of FH (green and blue stripes), FH-1-4 (light blue) and FH-19-20 (green). On the left: Sketch representation of C3b molecules (pink) immobilised on the dextran matrix of a CM5-chip surface (top) or on the carboxylated matrix of a C1-chip surface (below). For the lowest C3b loading of 140 RU on the C1-chip, C3b molecules, on average, space out by 46 nm, a distance that nearly spans 15 CCP modules. The fact that avidity is exhibited in binding of FH to the 140-RU C3b-CM1 chip is evidence for engagement of both the 1-4 and 19-20 sites by the *same* molecule of C3b.

These findings are in line with the literature and engender confidence in the "negative evidence" for strong binding sites elsewhere in FH. Weak binding of C3b by FH-6-8 is a novel observation; it may have physiological significance by contributing to the stability of the complex once FH has engaged with C3b via its principal recognition sites. The even weaker C3b (and C3c)-binding inferred for modules from 9 through 15 is of very questionable relevance. Thus the simplest interpretation of the current results is that there are two predominant, discrete, bindings sites in FH for C3b. A puzzling result was obtained in the case of the FH-C3d interaction since this does not match to the literature or any of the current dogma. If “true”, this observation has mechanistic implications (see DISCUSSION), but it requires further investigation.

CHAPTER 5

STRUCTURAL STUDIES

5.1 Overview

In the previous chapter, evidence was presented in strong support of the contention that modules representing the central region of FH (CCPs 9 to 15) do not make a major, direct, contribution to binding the primary ligands of FH, *i.e.* C3b and GAGs. Although not involved directly in binding, these central modules could nonetheless be crucial to the process whereby FH engages its ligands. Indeed the longer-than-average length of linkers in this region, which also includes the smallest CCP module of FH (CCP 13, spanning only 51 amino acid residues), likely reflects an important role of this region in the overall architecture of the full-length, 20-CCP module, molecule. To investigate this possibility, a structural approach was adopted with the aim of underpinning future mutagenesis and functional studies of the region, in addition to providing direct insights into the spatial organisation of these central modules. The CCP module-pair, FH-12-13, was selected for initial high-resolution structure determination by NMR: intriguingly, it compasses both the longest linker and smallest CCP in FH; and it is a convenient size for rapid NMR-based structural determination. (Note that triple-module and longer segments of RCAs produce NMR spectra with unfavourable relaxation properties leading to inefficient magnetization transfer and the poor performance of many triple-resonance experiments.). The involvement of the flanking inter-modular interfaces creating a bent back structure was examined by including FH-11-14 and FH-10-15 in the study. In the cases of these larger segments, AUC was the primary technique for examining structure. The strategy was to combine the high- and low-resolution structural information obtained from this study to build up an understanding of the architecture of central-FH.

5.2 NMR structure of FH-12-13

The determination of a high-resolution 3D structure by NMR requires a near-to-complete resonance assignment of all atoms. To facilitate such an exercise, a double-labelled sample of FH-12-13 was prepared and subjected to a suite of NMR experiments as described previously (METHODS). Utilising information from through-bond connectivity experiments, such as HN(CO)CACB and HNCCACB, the amide protons (except 783G) and H_α, H_β, C_α and C_β of all residues were assigned. It was subsequently possible to assign nearly all side-chain atoms utilising the ^{13}C-HSQC, HCCH-TOCSY and (for aromatics) (HB)CB(CGCD)HD, and (HB)CB(CGCDCE)HE experiments. The final assignment for [^{13}C,^{15}N] of FH-12-13 (see METHODS) includes 98.3% of all non-exchangeable hydrogen, nitrogen and carbon atoms (*e.g.* Figure 5.1).

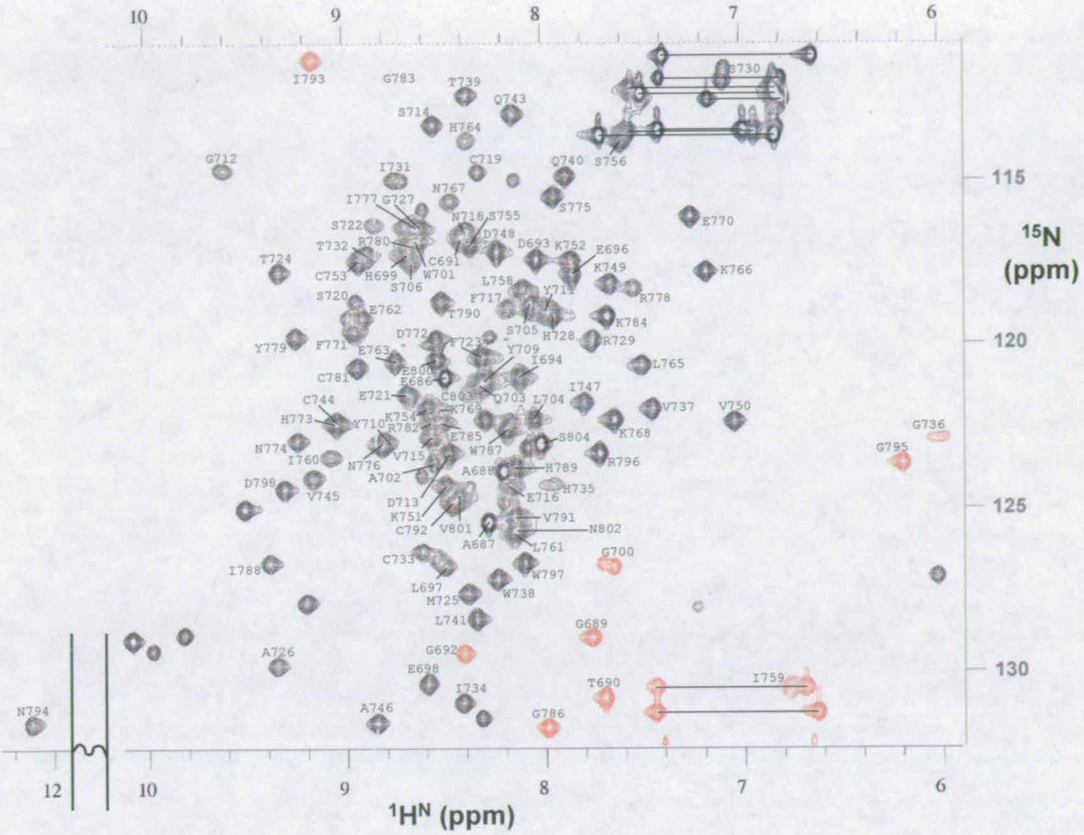


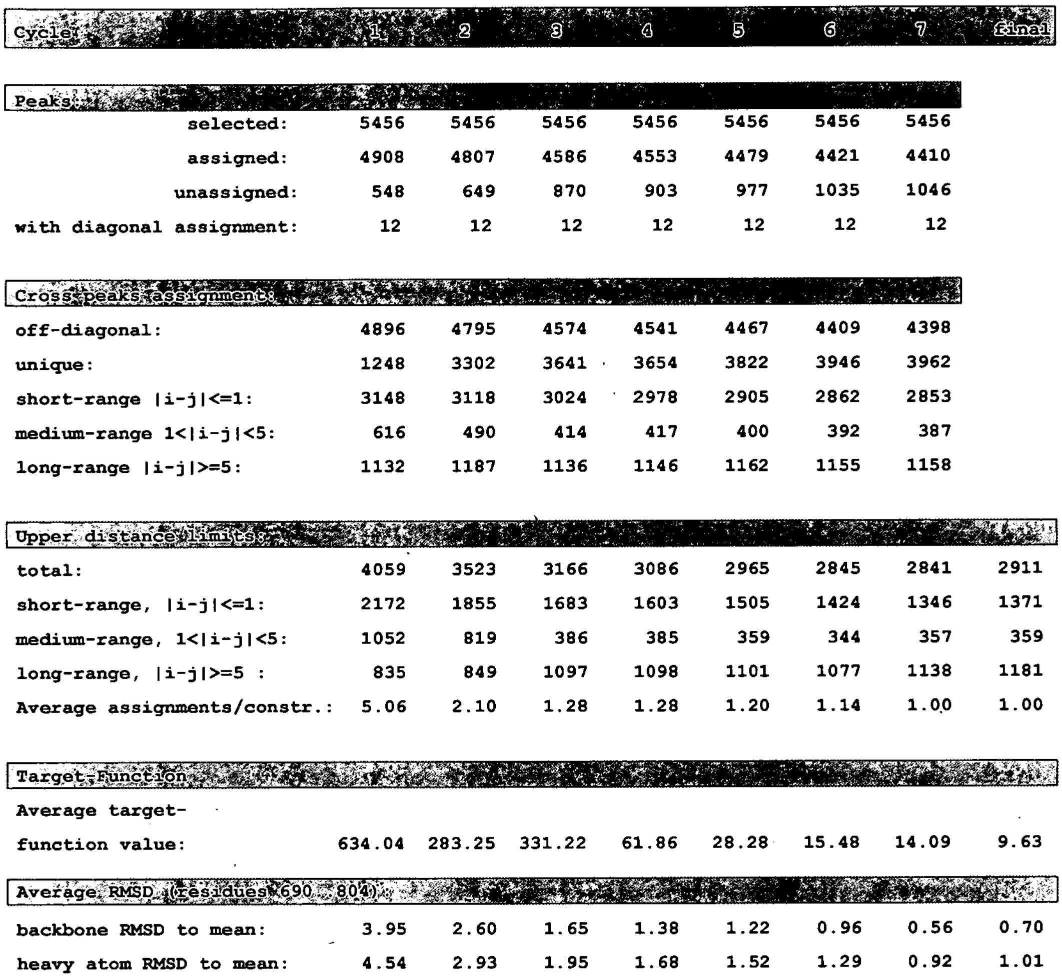
Figure 5.1 Assigned ^{15}N-HSQC spectrum of FH-12-13. Only backbone NH resonances are marked. NH resonances from Asn and Gln side-chains are connected, in each case, by a horizontal line.

5.2.1 Structure calculation

The CANDID automated NOE-assignment algorithm embedded within the CYANA software (132) was employed in the present study. This time-saving procedure works on the well-established principle that ambiguous NOEs can be automatically assigned in a reliable fashion on the basis of calculated structures as they emerge from sequential rounds of structure calculations. It is essential that a robust starting structure is generated in the initial round, and this is judged on the basis of several input and output criteria.

Since more than 90% of the (non-labile) backbone amide protons were present in the list of assigned chemical shifts, these data satisfied one of the key input criteria for the use of CANDID (132). To ensure satisfaction of the second criterion, a total of 5456 NOE cross-peaks were picked, thus providing a "faithful representation" of the available NOESY spectra. Seven successive rounds of CYANA-based structure calculations were then initiated. After the last cycle 80.8% of all picked NOESY cross-peaks had been assigned (compared to a threshold of 80% that is considered satisfactory for this stage of the CANDID algorithm). The target function after cycle one was 634·Å^2; this is poorer than the 250-Å^2 threshold recommended by the authors of CANDID. On the other hand, the target function had dropped to a reassuring 9.63 Å^2 after the seventh, final cycle, inside the 10 Å^2 value regarded as a satisfactory CANDID output criterion. The backbone heavy atom-RMSDs for the ensemble of first-cycle structures were found (in MolMol) to be 2.4 Å for the residues of CCP 12 and 1.9 Å for CCP 13 - these values are well inside the recommended maximum value of 3.0 Å. Moreover the "RMSD drift" between the first-cycle ensemble and the seventh-cycle ensemble (a critical measure of the extent to which the first cycle captures the "true" polypeptide fold) was less than 25 % of these values.

After seven cycles of CYANA-based structure calculation, a total number of 2911 unique upper distance restraints had been generated. These may be subdivided into 1371, 359 and 1181 restraints corresponding to intra-residue, medium-range (between residue i and residues i+1 to i+4), and long-range (between residue i and residues i+(≥5)) restraints, respectively (Figure 5.2).



```
Cycle:                             1       2       3       4       5       6       7     final

Peaks:
                 selected:      5456    5456    5456    5456    5456    5456    5456
                 assigned:      4908    4807    4586    4553    4479    4421    4410
               unassigned:       548     649     870     903     977    1035    1046
with diagonal assignment:         12      12      12      12      12      12      12

Cross-peaks assignments:
off-diagonal:                   4896    4795    4574    4541    4467    4409    4398
unique:                         1248    3302    3641    3654    3822    3946    3962
short-range |i-j|<=1:           3148    3118    3024    2978    2905    2862    2853
medium-range 1<|i-j|<5:          616     490     414     417     400     392     387
long-range |i-j|>=5:            1132    1187    1136    1146    1162    1155    1158

Upper distance limits:
total:                          4059    3523    3166    3086    2965    2845    2841    2911
short-range, |i-j|<=1:          2172    1855    1683    1603    1505    1424    1346    1371
medium-range, 1<|i-j|<5:        1052     819     386     385     359     344     357     359
long-range, |i-j|>=5 :           835     849    1097    1098    1101    1077    1138    1181
Average assignments/constr.:    5.06    2.10    1.28    1.28    1.20    1.14    1.00    1.00

Target Function
Average target-
function value:               634.04  283.25  331.22   61.86   28.28   15.48   14.09    9.63

Average RMSD (residues 690..804):
backbone RMSD to mean:          3.95    2.60    1.65    1.38    1.22    0.96    0.56    0.70
heavy atom RMSD to mean:        4.54    2.93    1.95    1.68    1.52    1.29    0.92    1.01
```

Figure 5.2 Report from CYANA structure calculation cycles for FH-12-13. Statistical tracking of NOE cross-peaks assignment, distant-restraint deduction, target function values and root mean square deviation values for the entire module FH-12-13 is shown (over all CANDID cycles).

Using Format Converter of the CCPNMR software (131), the upper distance restraints from the final CYANA cycle were converted into CNS-style distance restraints (because, as explained in METHODS, the use of well-established scripts within the CNS software allows for easier water refinement in the final step of the process).

A total of 100 structures of FH-12-13 were generated using appropriate scripts within the CNS structure calculation program. Figure 5.3 shows a plot of rank number (after ordering structures according to energy, lowest first) against overall energy and NOE energy of the structures generated in the final cycle. As may be judged from the positions of the major inflection points, good convergence was achieved for the lowest-energy 43 structures. On the other hand, smaller inflections points were observed in the structure-numbers 25-26 region of the curve. It was decided to select the 20 lowest-energy structures and submit them to a final refinement step in water solvent.

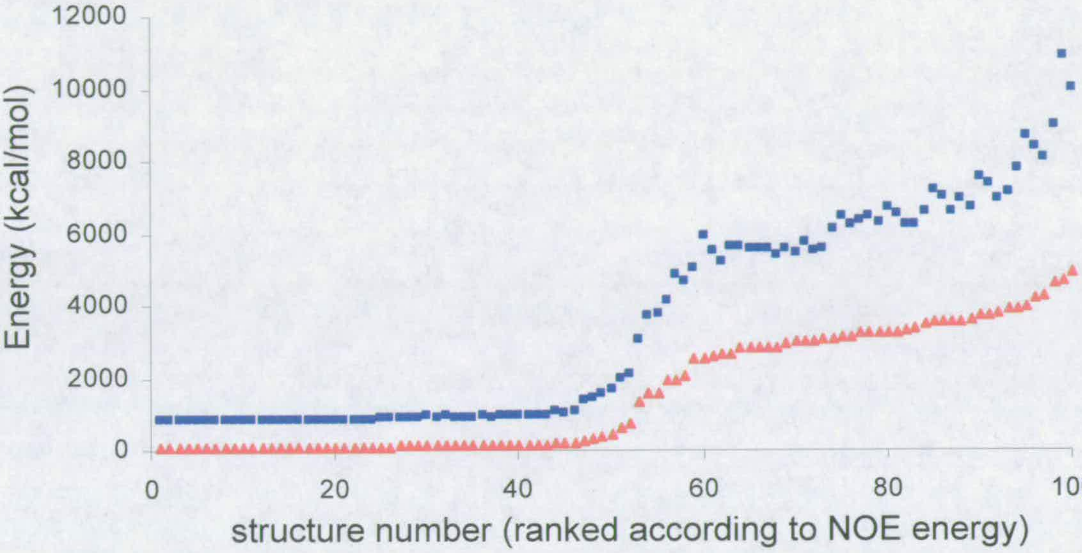


Figure 5.3 Energy plot of 100 structures from the final round of CNS structure calculations of FH-12-13. NOE energy is shown in red triangles. Total energy is shown in blue squares.

The resulting ensemble of 20 representative, water-refined, NMR structures is shown in Figure 5.4; backbone RMSD values for overlays on CCP 12, CCP 13 and CCPs 12-13 are indicated. Totals of 10, 75 and 75 distance constraints were observed between module 12 and 13[17], between module 12 and the linking amino acid residues, and between module 13 and the linker residues, respectively. Within CCP 12, CCP 13 and the linker residues, 1243, 1360 and 147 distance constraints were observed, respectively.

---

[17] Module boundaries are defined to span from the first consensus Cys to the last consensus Cys.

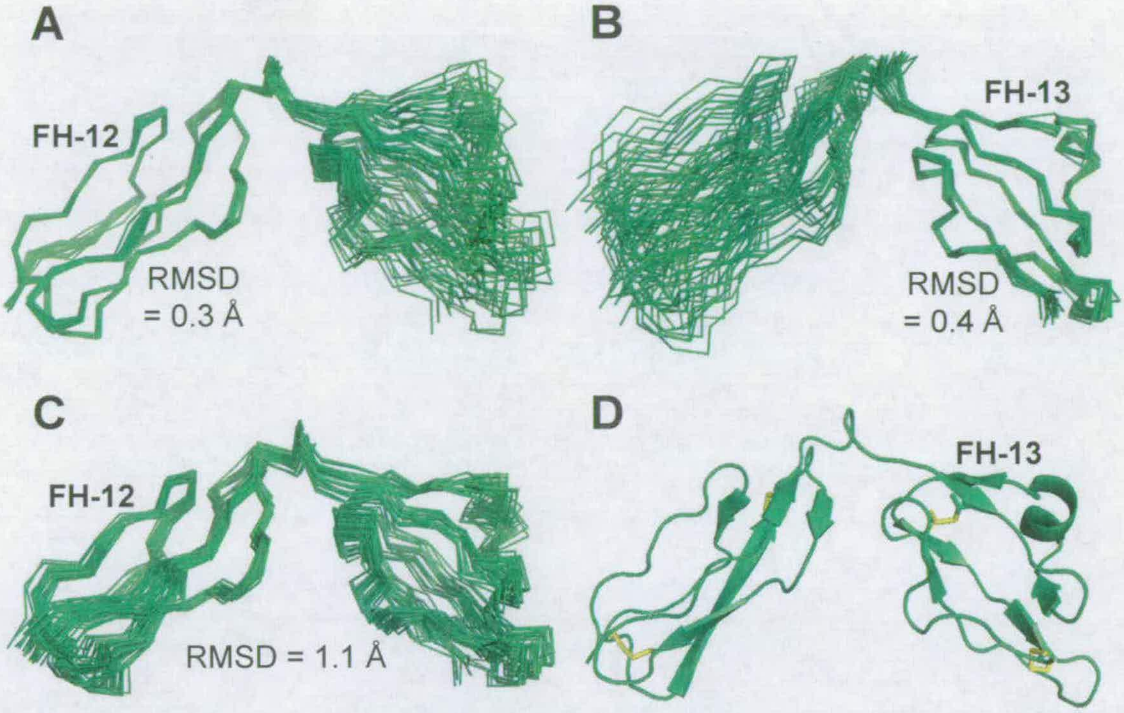


Figure 5.4: Ensemble of NMR-derived structures. Backbone overlay of 20 lowest-energy structures selected from a total of 100 calculated (RMSD over backbone atoms as indicated). (A) Overlaid on module 12. (B) Overlaid on module 13. (C) Overlaid on both modules - the good convergence in this overlay indicates the existence of a well-defined module-module orientation. (D) View of one of the 20 structures (the nearest to the mean) with secondary structure shown by a cartoon (PyMol (141)). Disulfide bonds are highlighted in yellow.

5.2.2 Description of structure and analysis

The structures of FH-12-13 are in good agreement with spectra collected in D_2O, which reveal slow-exchanging amides that are likely participating in H-bonds – see below. The structure reveals an approximate 100-degree bend between CCP 12 and CCP 13 (Figure 5.5).

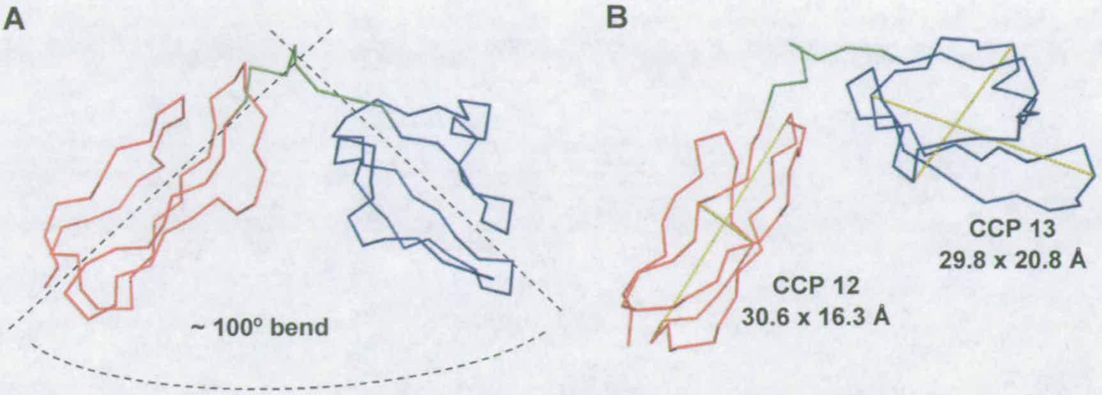
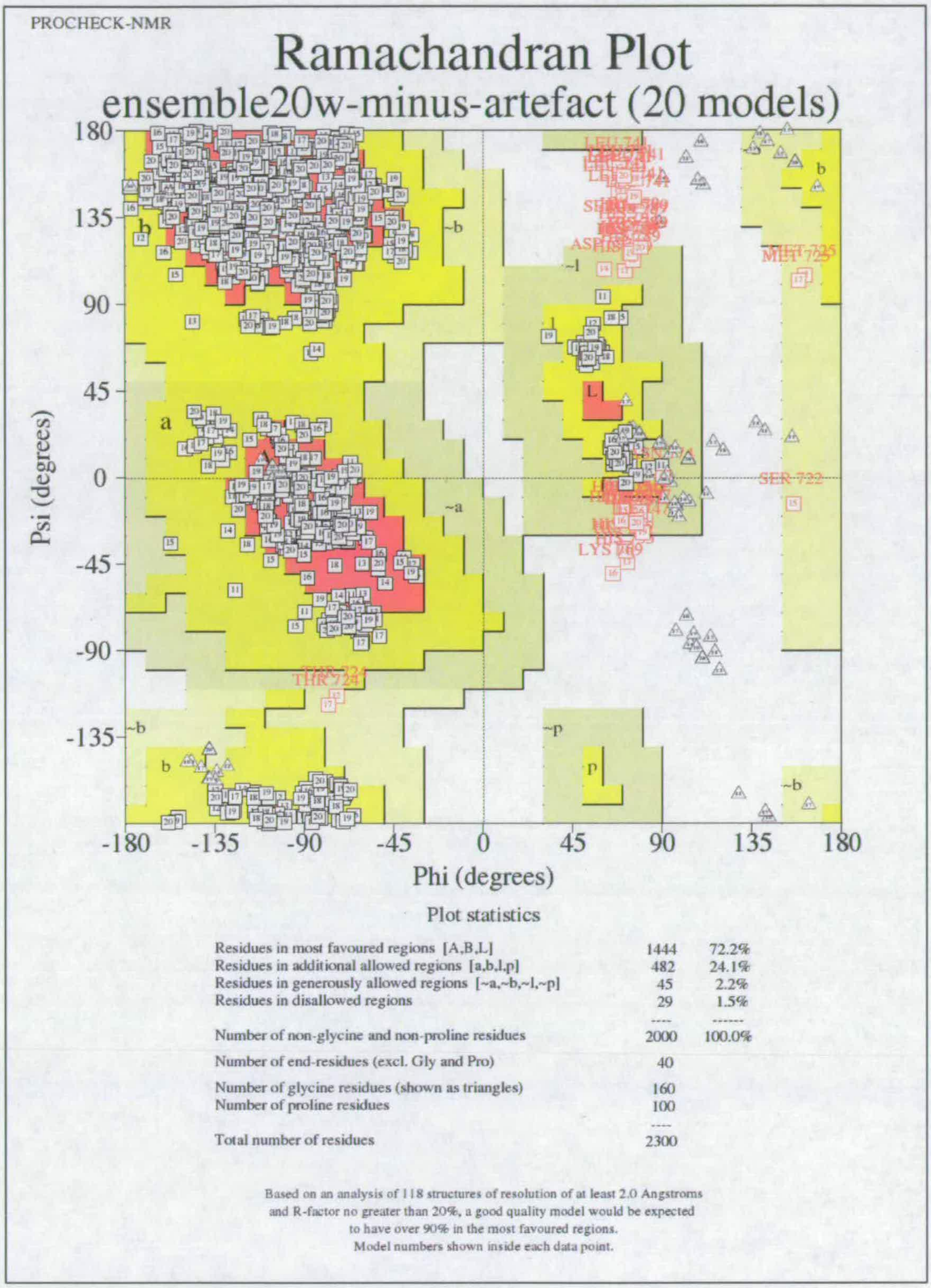


Figure 5.5 Characteristics of CCP in the structure of FH-12-13 (backbone traces are shown; CCP 12 and CCP 13 are coloured red and blue, respectively). (A) The 100°-bend between CCPs 12 and 13 is indicated. (B) Dimensions of each CCP are highlighted on the backbone traces by a dashed, yellow line.

Both CCP 12 and CCP 13 have structures that are fairly typical for modules of this type: the disulfide bonds are Cys(I)-Cys(III) and Cys(II)-Cys(IV) and occur towards either end of the ovoid CCP modules. Anti-parallel β-strands are the prevailing secondary structural element. The shape of CCP 13 appears to be less ovoid and more spherical than that of CCP 12 (Figure 5.4-D) and in comparison with other CCPs. The hypervariable loop of CCP 13 shows helical elements in all 20 lowest-energy structures, and in five structures out of the 20, this region complies with the criteria of an α-helix structure (according to MOLMOL).

The water-refined ensemble of 20 lowest energy structures of FH-12-13 was submitted to the "Coarse-packing quality control" check within the programme WHATIF (164)). This routine assesses the local environment of individual amino acids residues in terms of a "directional atomic contact analysis" which compares the distribution of atom types around amino fragments to the average distribution observed in the PDB. A WHATIF course-packing score lower than -5.0 indicates improper packing. The average score for the submitted ensemble is comfortably inside this value at -1.371.

PROCHECK is a programme that evaluates the stereochemical quality of protein structures in terms of the extent to which its *phi* and *psi* angles occupy energetically favoured regions of the Ramachandran plot (165, 166). A PROCHECK analysis of ensemble yielded acceptable scores. Overall, 72.2 % and 24.1 % of amino acid residues are in the most-favoured and additionally-allowed regions of the Ramachandran plot (Figure 5.6), respectively. These scores are of comparable quality to other, published , NMR-derived structures of CCP modules (1, 51, 148, 167).



PROCHECK-NMR

Ramachandran Plot

ensemble20w-minus-artefact (20 models)

Plot statistics

```
Residues in most favoured regions  [A,B,L]                  1444     72.2%
Residues in additional allowed regions  [a,b,l,p]            482     24.1%
Residues in generously allowed regions  [~a,~b,~l,~p]         45      2.2%
Residues in disallowed regions                                29      1.5%
                                                            ----    ------
Number of non-glycine and non-proline residues              2000    100.0%

Number of end-residues (excl. Gly and Pro)                    40

Number of glycine residues (shown as triangles)              160
Number of proline residues                                   100
                                                            ----
Total number of residues                                    2300
```

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.
Model numbers shown inside each data point.

Figure 5.6 Ramachandran statistics (PROCHECK) of water refined ensemble of the 20 lowest energy structures of FH-12-13.

In a hydrogen/deuterium exchange experiment, amide protons that are exposed to solvent exchange rapidly with deuterium atoms. The substitution of N-H with N-D results in a loss of the relevant NMR signal. Hence ^{15}N-HSQC cross-peaks of exposed amide groups disappear within a period of seconds-to-minutes when the sample is dissolved in D_2O. On the other hand, protons that are buried and inaccessible to solvent, or are involved in the formation of hydrogen-bonds, are less susceptible to exchange and their signals remain detectable in a ^{15}N-HSQC spectrum for a period of hours-to-days. These slowly exchanging amides generally correspond to regions of hydrogen-bond-stabilised secondary structure. In the current work, a double-labelled sample of [^{15}N-^{13}C]-FH-12-13 was lyophilised and re-suspended in 99.96% D_2O. A series of ^{15}N-HSQC spectra were recorded, initially at intervals of 30-40 min and finally after 24 hours.

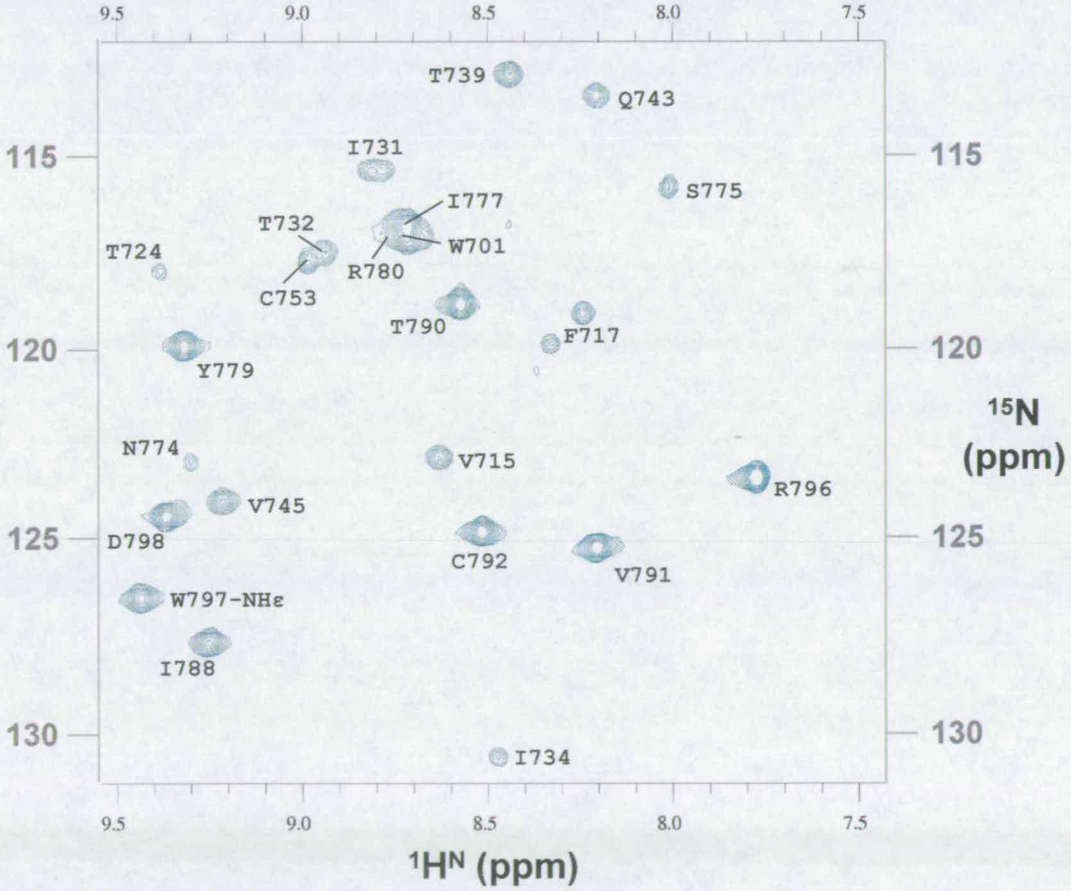


Figure 5.7 ^{15}N-HSQC spectrum of FH-12-13 30 minutes after re-suspension in D_2O. Sample is 600 μM in a volume of 300 μl.

After 30 minutes of exposure to D_2O, 23 H^N signals are still clearly visible in the spectrum. These are highlighted in red on the structure of FH-12-13 (Figure 5.7). Of these 23, 17 (*i.e.* 77%) correspond to H-bonded amides within beta strands (Figure 5.8).

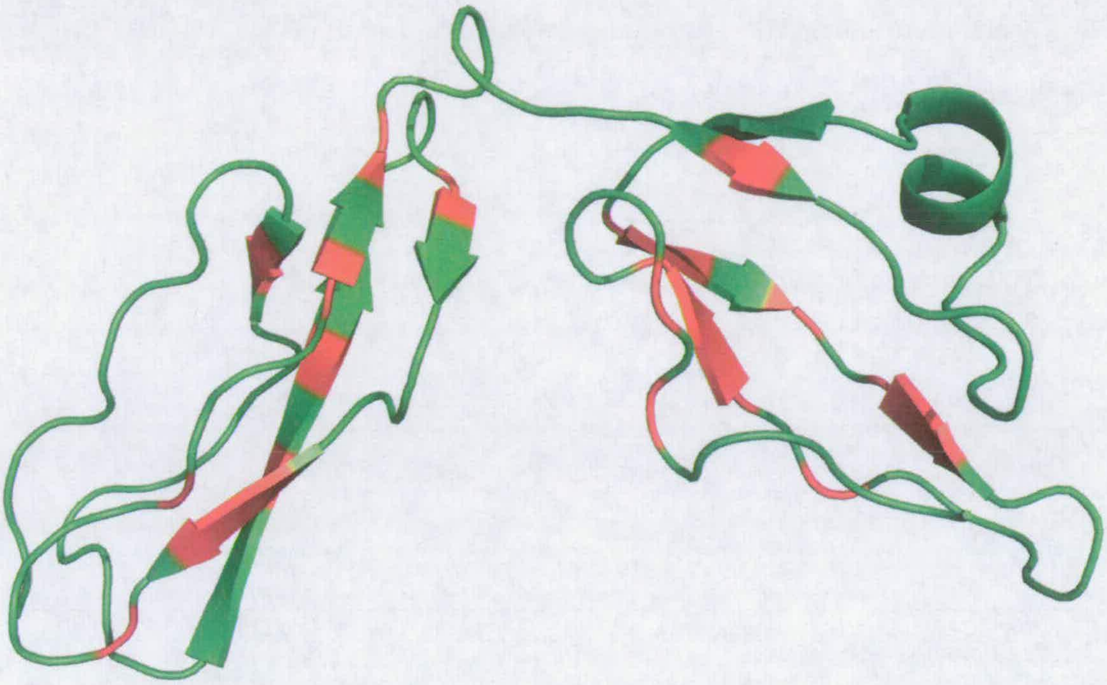


Figure 5.8 Amino acids corresponding to resonances unaffected by the hydrogen/deuterium exchange are highlighted in red on the structure of FH-12-13.

Five of the remaining six signals also correspond to amide protons that are engaged in formation of hydrogen bonds: D798 H^N with the H_α and hydroxyl-oxygen of T790 is (T790 is part of a beta-strand); R796 H^N shows main-chain hydrogen-bonding with carbonyl of I793 (which lies just after a beta-strand); I788 H^N hydrogen bonds with carbonyl of beta-strand residue Y779; T739 H^N hydrogen bonds with the carbonyl of β-strand residue T732; and I734 H^N engages with the carbonyl of V737 to form a β-turn (Figure 5.9). Typically for a β-turn, the position i+2 is occupied by a glycine (168).

The remaining signal corresponds to the H_ε of the consensus tryptophan in CCP 12.

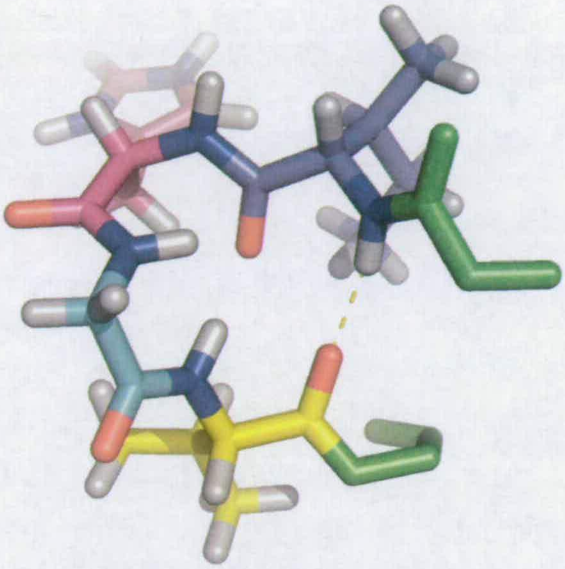


Figure 5.9 Type IV β-turn involving I734 (blue), H735 (magenta), G736 (cyan) and V737. Yellow dotted line indicates the H-bond between H^N of I734 and carbonyl oxygen of V737.

In summary, 22 of 23 slowly exchanging amide resonances were demonstrated to be a part of or linked to secondary structure elements.

5.2.3 Analysis of the intermodular interface

There are two unusual features in the primary structure of the module-pair FH-12-13. CCP 13 is the smallest domain in FH and it connects to CCP 12 by eight linking amino residues, thereby constituting the longest linking sequence in FH. Prior to structure determination of FH-12-13, a range of possibilities regarding the structure of this linker were considered plausible: for example the linker might form a 180° U-bend and positions CCPs 12 and 13 in an anti- parallel side-by-side orientation; or the linker might not maintain a stable (rigid) conformation, but rather be flexible and allow a large degree of movement between the two CCP modules; or the eight amino acid residues could form a specific, stable structure fixing the orientation of CCPs 12 and 13 with respect to each other. The experimentally derived structure proves the third of these to be the case and reveals a 100° bent between modules 12 and 13 (Figure 5.5). In this respect it is noteworthy that six of the eight linking residues are characterised by long or bulky side-chains (Figure 5.10), providing many methylene- and methyl-groups for van der Waals contacts not only within the linker residues, but with side-chains from both flanking CCP modules.

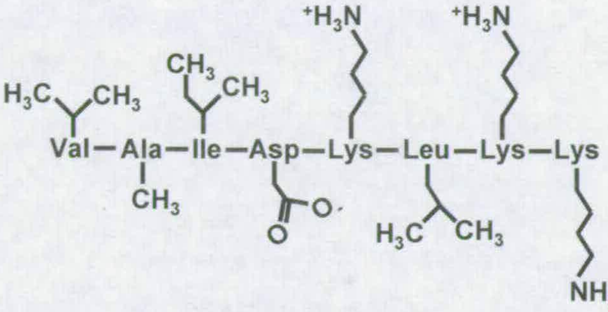


Figure 5.10 Sketch representation of the eight linker residues between module CCPs 12 and 13. Side-chain atoms are shown.

The total of 75 NOE-derived distance restraints between the linker residues and each of the two modules reflects the large intimacy contacts with both CCPs. The CCP module-to-linker contacts would only stabilise the relative orientation of the two modules if the linker itself folds into a stable structure. This does indeed appear to be the case with 147 distance constraints identified within the linker. The result is a "mini" domain with its own small core that acts to "glue together" the two modules (Figure 5.11-D). The compactness of the linking "mini-domain" is such that the modules make direct contacts, as reflected in the ten NOE-derived distance restraints. Figure 5.11 (A-C) highlights direct intermodular contacts at the interface of CCPs 12 and 13.

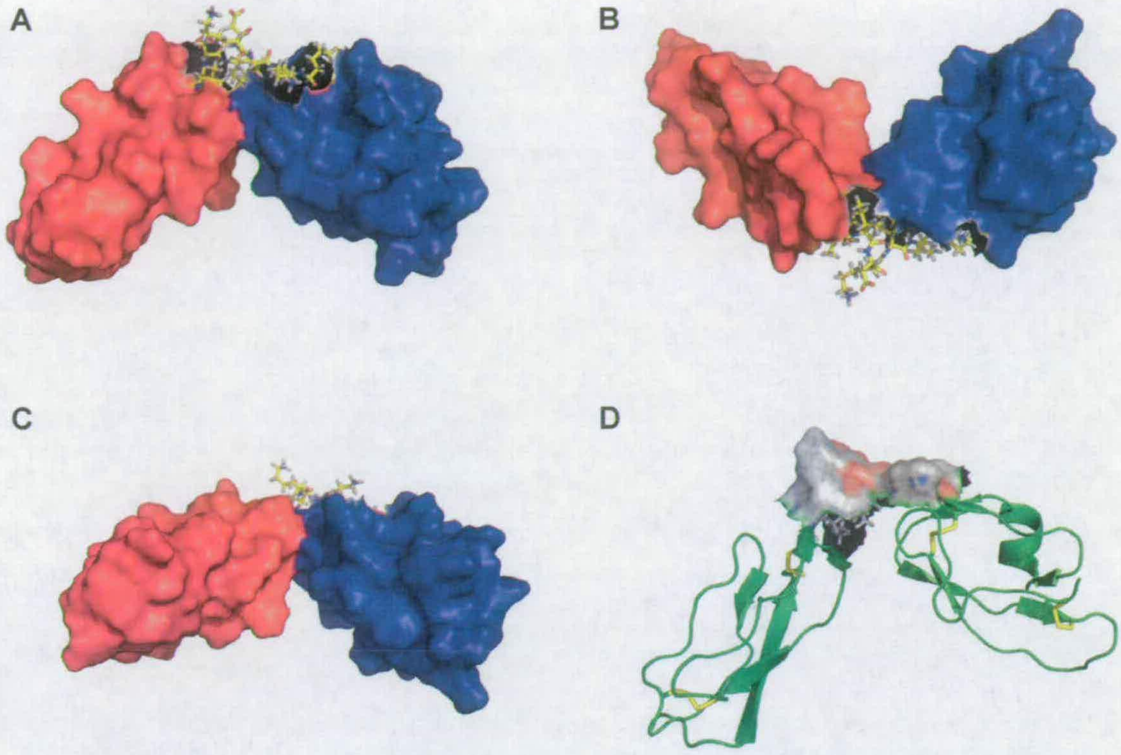


Figure 5.11 Surface and cartoon representations of FH-12-13. (A-C) The eight linker amino acids are shown as stick-representation. The surface of the FH-12-13 module is shown, but surface corresponding to the linker residues has been made invisible. CCPs 12 and 13 are coloured red, and blue, respectively. (D) The eight linker residues (shown in surface representation – the remaining surface of FH-12-13 has been made invisible) are characterised by bulky side-chains that form a compact core ("mini-domain") which fixes the 100°-bend between CCPs 12 and 13.

5.2.4 Analysis of FH-12-13 surface

Surface bound polyanionic carbohydrates are a primary ligand of FH. Recognition of these surface markers is thought to provide the basis for self *versus* non-self discrimination (INTRODUCTION). The present study mapped GAG-binding sites in FH to two regions, CCPs 6-8 and 19-20 (CHAPTER 4). In a previous study, however, CCP 13 was implicated in heparin binding (67). Binding to the negatively charged sulphate- and carboxyl groups of surface GAGs (or heparin as a model GAG) is assumed to require a complementary, positively, charged area on the protein surface.

A surface-charge analysis of FH-12-13 reveals a negative and positive surface patches that lie on opposite faces of the structure and are localised in CCP 12, and CCP 13, respectively (Figure 5.12).

Three glutamate residues (E696, E698, E721) and one aspartate (D693) are the main contributors to the electronegative surface region of CCP 12. The positive patch in the centre of CCP 13 extends to the C-terminus. Two lysines from the linker (K751 and K752) and five basic residues from CCP 13 (K754, K766, K768, K769 and R796) form the main positive patch on the molecule. The amino acids K780, R782, K784 extend this positive area towards the C-terminus. It is striking, that despites this extended positive surface on one face of module 13, no affinity for heparin could be detected at physiological pH (METHODS). This will be discussed further below (DISCUSSION).

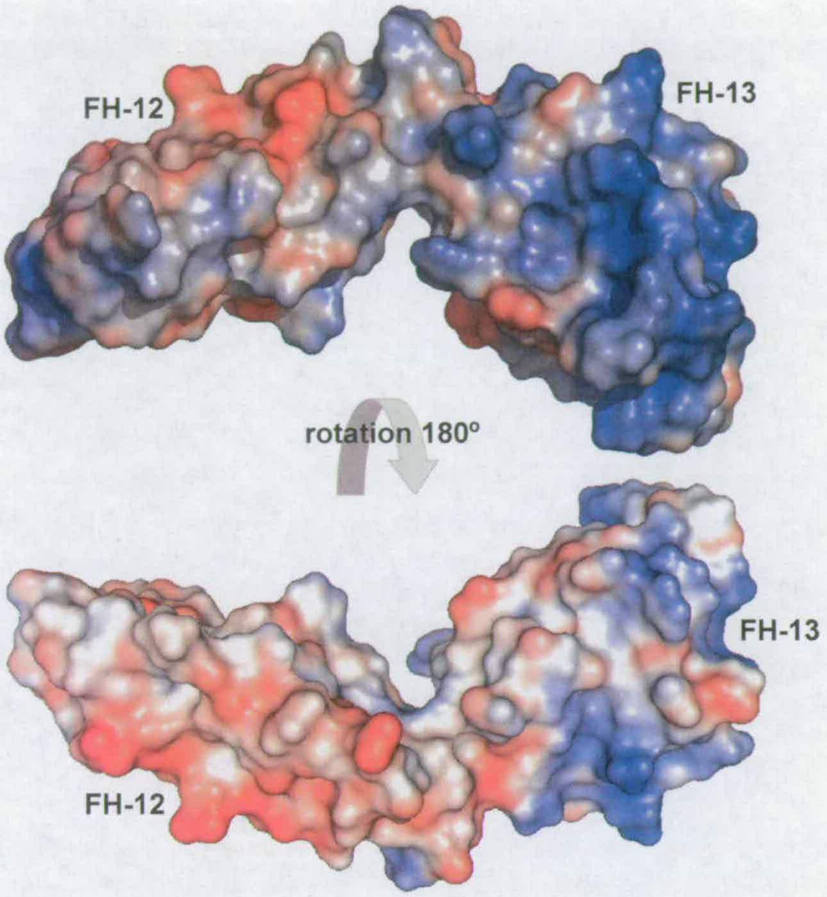


Figure 5.12 GRASP (142) Electrostatic surface representations of FH-12-13. Electrostatic surface potential: red = negative charge, blue = positive charge, range = -5/+5 kT. While acidic amino acids in CCP-12 cluster into a negative surface patch, basic amino acids in CCP-13 cluster into a positive surface patch.

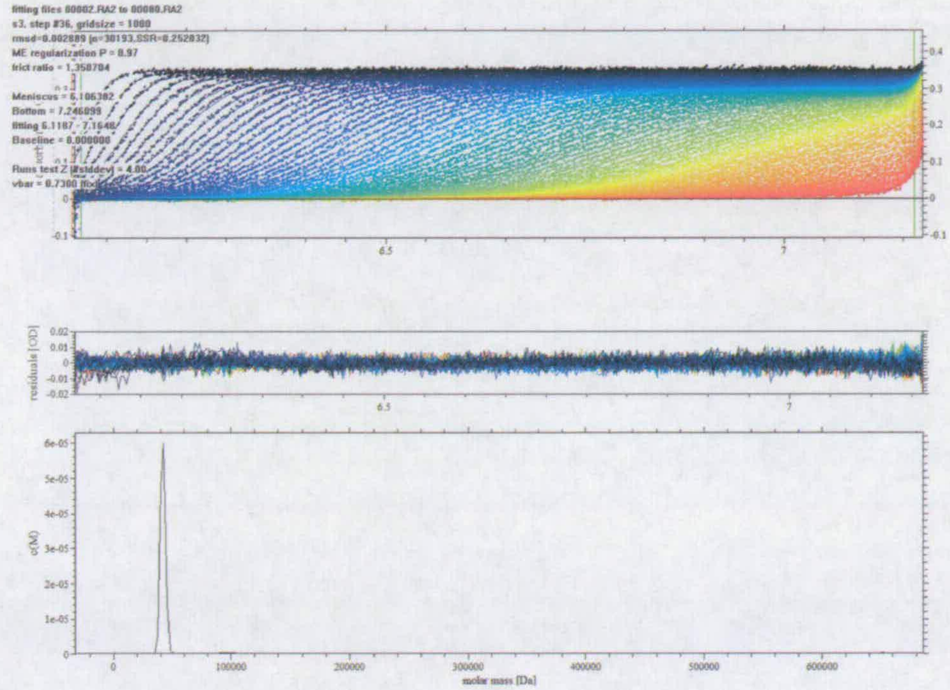
5.3 Analytical ultracentrifugation[18]

The results of the preceding section imply that despite the long linking sequence, modules 12 and 13 do not form a U-turn nor do they comprise a flexible hinge. Assuming that the 100-degree bend between 12-13 observed in the structure does indeed persist in the intact FH, it is intriguing to speculate on whether the 13-14 junction for example, characterised by another long linker, is also bent and if so whether it is bent in the same direction. To address these issues, the technique of analytical ultracentrifugation was employed.

---

[18] Ultracentrifugational analysis was carried out by Dr. Arthur Rowe, School of Biosciences, University of Nottingham, UK

Sedimentation velocity analytical ultracentrifugation is a well-established technique for characterising a protein sample (144, 169). Not only does it allow verification of quaternary structure and detection (and quantification) of aggregates it also provides information on the overall shape of a protein at the level of discrimination between roughly spherical entities and extended or rod-like molecules. The method involves monitoring the protein concentration distribution across a cell that is being spun at very high speed (typically 45,000 rpm) in a rotor within a centrifuge. Under the influence of strong centrifugal force, the protein molecules migrate away from the centre of the rotor such that an expanding region that is depleted of molecules forms behind the air-water interface (meniscus). The boundary between populated and unpopulated regions of the cell moves towards the outside of the rotor over time (eventually forming a pellet). The concentration distribution is measured by detection of absorbance or refractive index at intervals over a time-course. Runs are typically three-four hours in length and the sample consists of roughly 0.45 ml at a concentration of 0.02-0.1 mg.ml^{-1}. The results of a sedimentation velocity run on FH-10-15 are shown in Figure 5.13. The upper frame shows a series of scans across the centrifuge cell (absorbance at 273 nm plotted against distance from the centre of the rotor). The first scan (black) was performed at 4 minutes after starting the 45,000-rpm (the black data set in the graph), and subsequent scans (rainbow colours) were recorded at two-minute intervals thereafter. The first data set shows that the area behind the meniscus is already largely depleted of protein molecules. Over time the boundary moves - sediments - outwards. Both the molecular weight and the shape of the protein determine its rate of sedimentation; high-molecular weight, spherical proteins sediment faster than low-molecular weight, elongated ones. The shape-dependency is a consequence of the greater friction experienced by a less compact protein.



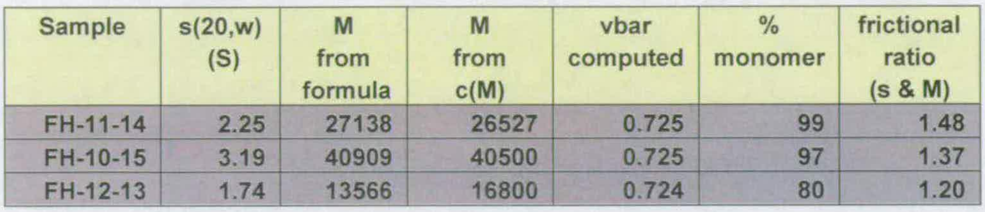
```
fitting files 00002.RA2 to 00080.RA2
s3, step #36, gridsize = 1000
rmsd=0.002889 (n=30193,SSR=0.252832)
ME regularization P = 0.97
frict ratio = 1.358704

Meniscus = 6.106382
Bottom = 7.246099
fitting 6.1187 - 7.1648
Baseline = 0.000000

Runs test Z (#stddev) = 4.00
vbar = 0.7300 (fix)
```

Figure 5.13 A screenshot showing the outcome of processing the data from a sedimentation velocity AUC run (performed and analysed in Nottingham) on FH-10-15. The top frame shows multiple traces of the distribution (with respect to distance (cm) from the axis of rotation, x-axis) of OD^{273} values; black is the initial trace, and subsequent traces recorded at time intervals during the course of the run are shown rainbow-fashion, violet-to-red. The bottom frame shows a fitted (Sedfit) plot of the distribution of molecular weights within the sample (mostly a monomer of the expected molecular weight, but with a minor population of putative dimer). The middle frame shows the residuals and therefore indicates the goodness of fit.

The lower frame of Figure 5.13 shows the distribution of sedimentation rates (coefficients) for the population of protein molecules in the cell, as extracted from the scans in the upper frame using a fitting-routine (the goodness of fit is summarized in the plot of residuals in the middle frame). The area under the peak is proportional to the amount of protein, and the plot resembles an inverse size-exclusion chromatogram. The presence of a single (Gaussian) peak is good evidence for the existence of a single species in this sample. The diffusion coefficient can be obtained from the width of the peak and used in combination with the position of the peak to calculate the molecular weight of the protein. The fitting routine also provides a frictional ratio that relates to the "axial ratio" of the molecule and pertains to its overall shape (Table 5.1).



Sample	s(20,w) (S)	M from formula	M from c(M)	vbar computed	% monomer	frictional ratio (s & M)
FH-11-14	2.25	27138	26527	0.725	99	1.48
FH-10-15	3.19	40909	40500	0.725	97	1.37
FH-12-13	1.74	13566	16800	0.724	80	1.20

Table 5.1 Summary of results from sedimentation velocity AUC. S(20,w) is the sedimentation coefficient measured in S = Svedberg units; M = relative molecular mass (either calculated from the sequence (c(M)) or estimated from the fitted sedimentation data); vbar is the partial specific volume in units of ml per g ;the frictional ratio is calculated from vbar, c(M) and s(20,w) and is an indicator of shape – the higher its value, the higher the axial ratio (i.e. the more elongated the molecule).

It was hypothesised that CCPs 12-13 form part of a 180-degree curve in FH that is created by a succession of same-direction bends between adjacent modules. If this is true, then the addition of modules 11 and 14 will result in a structure with a smaller axial ratio (a less extended structure) as summarised in Figure 5.14. Moreover, if these four modules achieve a U-turn in the structure then the further addition of CCPs 10 and 15 would result in a still smaller axial ratio (Figure 5.14). Thus the hypothesis predicts that axial ratios will not tend to increase as more CCP modules are added to FH-12-13 and this is a prediction that is testable by AUC. Therefore, the set of constructs, FH-12-13, FH-11-14 and FH-10-15 were submitted to the University of Nottingham (Professor Arthur Rowe) for sedimentation velocity AUC.

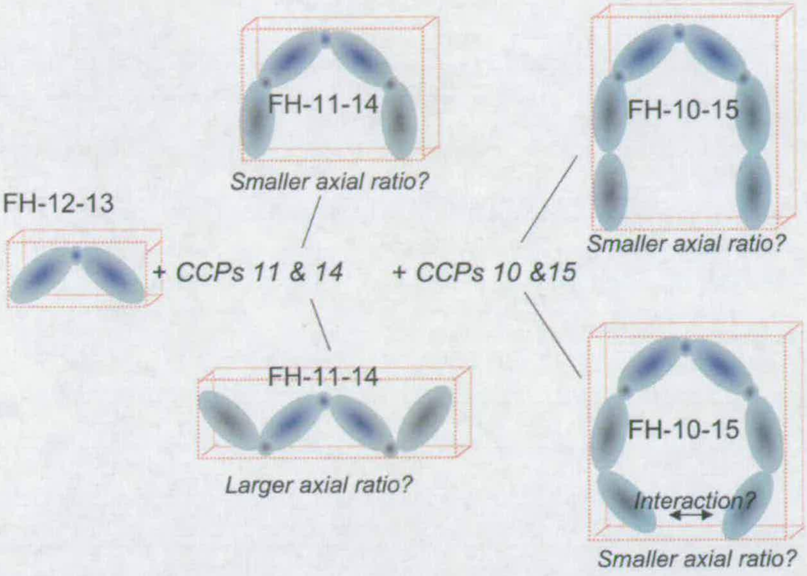


Figure 5.14 Schematic to illustrate the thinking behind the AUC-based experiments and their interpretation. Each CCP module is indicated by an oval shapes; "long" linkers (six or more residues) are drawn as small circles; the cubes (dotted lines) are drawn to help visualise the axial ratios of the structures as drawn. The initial hypothesis (upper cartoons) proposed a decreasing trend in axial ratios with increasing numbers of modules on either side of CCPs 12-13. This was not supported by the data, which was consistent with the speculative suggestions drawn in the lower cartoons (see text).

The first thing to notice, is that the results of this analysis demonstrate all of the samples to be predominantly monomeric. Although oligomers are present as a minority component in some samples, the simplest interpretation is that self-association is not an important property of these segments of FH. As may be seen from Table 5.1, the fitted frictional ratios are *inconsistent* with the initial hypothesis in that FH-11-14 is considerably more extended in shape that FH-12-13. This strongly argues against a "smooth" bend as envisaged in the previous paragraph. There are numerous plausible explanations for this result, including the possibility that the 11-12 and 13-14 junctions are flexible resulting in multiple conformations that average to a rod-like shape (see Figure 5.14). The most striking feature of the AUC results, however, is that FH-10-15 is more compact in overall shape than FH-11-14. Given the uncertainties in explaining the sedimentation data for FH-11-14, it is obviously impossible to provide more than speculative suggestions regarding the structure of FH-10-15. On the other hand, it is tempting to hypothesize that the 12-13-induced 100-degree bend, in combination with flexibility at the 11-12 and 13-14 junctions, allows modules 10 and 15 to interact, stabilising the compact conformation that is implied by the AUC data. Clearly, more investigation is required to investigate this and other possibilities.

5.4 Conclusions

Previous structural work (1, 50, 51, 53, 148) on FH focussed on modules towards the N- and C-termini, either because of their functional relevance (CCP modules 1-3, 6-8 and 19-20) or because (in the cases of CCP modules 5, 15 and 16) of their ease of expression. This left a dearth of structural information for CCP modules 9-14 – the central portion of FH. The work of the previous chapter seems to rule out the involvement of these modules in important direct interactions with the principal ligands of FH – GAGs and C3b. On the other hand, these functional results pointed to an architectural role of the smaller-than-average CCP modules joined by longer than average linkers in this portion of FH. The current chapter describes results that shed the first light on the structural biology of these modules and helps to formulate hypotheses for the mechanism of action of FH (see DISCUSSION).

This work reveals that the long linker between modules 12 and 13 does not promote flexibility nor does it allow the two modules to form a side-to-side interaction that would promote a 180 degree-bend in FH. Rather, the bulky residues of the linker form a compact "mini-module" that acts like "glue" to hold the flanking modules in a rigid 100-degree bent conformation. This is the first time such an arrangement has been observed between CCP modules (In the only other example of a CCP-module pair joined by an eight-residue linker, the modules lie side-by-side (170)). There is no reason to suspect that the mere presence of neighbouring modules (11 and 14) in FH will disturb this arrangement. The 12-13 structure solved in this work provides a basis for modelling other intermodular "junctions" in this region and for speculating that they continue the bend in the same direction. Experimental work is required to confirm whether this is the case.

This chapter also describes an attempt to obtain less detailed structural information on the central region of FH using AUC. The principal result is that FH-10-15 is more compact in its overall shape than FH-11-14. Such a result could have several explanations but the most obvious one is that modules 10 and 15 contact one another to stabilise a bent-back conformation (that is induced in part by the bend between CCPs 12 and 13). This possibility is considered further in the DISCUSSION.

CHAPTER 6

DISCUSSION

In this brief final chapter an attempt is made to draw together the various strands of the work, and to suggest a model for FH engagement with C3b

6.1 Factor H has just two principal binding sites each for GAGs and C3b

The deployment of complement regulatory proteins as a means of selectively protecting self tissue from the potentially detrimental consequences of indiscriminate amplification of surface-deposited C3b (via the alternative pathway) is a simple but apparently effective strategy for the provision of immune surveillance (37). As demonstrated in the current work, and in confirmation of numerous existing reports in the literature, the first four CCP modules of FH are necessary and probably sufficient for the ability of this molecule to function as a cofactor for factor I in the fluid phase; these N-terminal four modules are also reported to have some ability to accelerate the decay of the convertase-complexes (58, 59, 171). This then leaves the remaining 16 modules with the responsibility for ensuring that the regulatory potential of FH is delivered selectively on self surfaces. The current work (again, backed by copious literature) strongly suggests that this delivery is achieved by GAG-recognition through modules 7 and 20 (and their flanking modules), and, by the presence of an additional C3b-binding site at the FH C-terminus. In contradiction to several previous papers, the current data demonstrate that modules 9-15 lack discrete, strong, binding sites for C3b (or C3c or C3d) or heparin. The new results suggest that any contribution these modules make to binding is relatively small.

It is quite striking that the single module CCP 13, which bears a prominent positively charged patch on one face (STRUCTURAL STUDIES), is not a good heparin binder. It suggests that the specific positioning or orientation of side-chains, as opposed to complementary charge alone, is necessary for the binding of GAG at physiologically relevant salt concentrations. The additional observation that CCP 13 in the context of FH-12-13, and, in the context of four longer segments, does not bind heparin appears to rule out the possibility of flanking modules serving to complete a partial GAG-recognition site in CCP 13. It does remain possible, though, that species of GAG other than the heparin/heparan sulfate family, or GAGs with different sulfation densities for example (68), do interact with CCP 13; it should be borne in mind, however, that heparin is very rarely a weaker ligand than heparan and dermatan sulfates for GAG-binding proteins. It is also conceivable that a very weak GAG-binding site in CCP 13 cooperates with the GAG-binding sites in modules 7 and 20 once FH has become anchored at the cell surface via these latter modules. What is clear from the current work, however, is that CCP 13 ought not to be discussed in the same vein as modules 7 and 20 in terms of GAG binding by FH. Unlike CCP 13, CCP module 9 appears as an unlikely candidate for a heparin binder on the basis of its pI. Indeed, a protein consisting of CCP modules 1–6 followed by 8 and 9 (*i.e.,* the delta-CCP 7 version of FH-1-9) has been reported previously not to bind heparin (62). The more recent results of Ormsby *et al.* identifying CCP module 9 as a GAG-binding module (61) were therefore surprising. The current work shows that FH-8-9 could be converted from a non-GAG binding construct to a GAG binding one by addition of non-native cationic N-terminal residues present in the *Ormsby et al.* constructs. This strongly suggests that the non-native Arg residues were critical contributors to the previously reported interaction with GAGs. It nonetheless remains possible that CCP module 8 contributes to the well-explored GAG binding site centred on neighbouring CCP 7 (53).

In the current study C3b was immobilized via amine coupling to a carboxymethylated dextran SPR sensor chip, enabling direct measurements of C3b-FH interactions. It is noteworthy that no significant differences in the SPR-derived K_d for FH-19-20 – when comparing amine-coupled C3b with C3b immobilized via a biotinylated thioester linkage to an avidin chip – was detectable (binding data for the biotinylated C3b was acquired in the lab of Dr. John D. Lambris, University of Pennsylvania). Previously reported SPR studies (60) indicated that FH-8-20 and FH-19-20 bind C3b, but FH-8-11 and FH-15-18 do not (60). An earlier published study observed a perturbation in binding to C3b, immobilized via its thioester to an erythrocyte surface, by an FH deletion mutant that lacks CCPs 6–10 (*i.e.*, FH-delta6–10) (65). It was not entirely clear if this observation reflected a loss of direct interaction with C3b, as there was a concurrent loss in this mutant of the GAG-binding site in CCP 7 that could, presumably, contribute to FH association with C3b in the context of a GAG-bearing cell surface. None of these results conflict with the ones reported in the present study.

In previous work (60), however, a "third" C3b binding site in CCPs 12-14 was inferred by Jokiranta et al. (60) on the basis that FH-8-20 binds both C3d and C3c while FH-19-20 binds C3d but not C3c (ergo, there is a C3c-binding sites somewhere between modules 8 and 18 – the authors pinpointed CCPs 12-14 based on some further deductions). It should be remembered that C3c and C3d are distinct fragments of C3b; when C3b is cleaved into the fragments C3c and C3d new surfaces are exposed on both fragments (13, 14) and it is conceivable that these sites act as binding sites for CCP domains from the centre of FH. In the current study, however, no affinity for either C3c or C3d (amine coupled to a CM5 sensor chip) of the FH-8-15 construct was detectable. Thus this finding contradicts the previous report of a third C3b/C3c binding site. It is of note that the experiments reported by Jokiranta were carried out in one-third of physiological salt concentration (1/3 VBS buffer). In contrast to this, all SPR experiments undertaken for this study have been carried out in physiological salt concentration (HBS-EP+).

On the other hand, our observation that FH-7-8 and FH-6-8 have weak but possibly significant affinities for C3b immobilized on a chip is consistent with the original studies on FH-delta6–10 (65); as mentioned above, the reduced affinity of FHΔ6–10 to C3b coated surfaces could have arisen from the loss of the second polyanion binding site in FH and therefore the possibility of CCPs 6-8 to exhibit affinity for C3b never attracted much attention. However, in the light of the AMD-linked polymorphism in CCP 7 this observation needs to be explored further in future studies.

6.2 A model for engagement of FH with C3b on the cell surface (Figure 6.1)

The structural work described in CHAPTER 5 (along with SAXS-based evidence from the Perkins lab (48)) is consistent with a bent-back conformation of FH *that allows spatial proximity of its N- and C-terminal portions*. Based on the functional studies in CHAPTER 4 (and the majority of the previous literature in this field) it is clear that between them *these regions of FH contain all the experimentally proven, discrete binding sites for GAG and C3b*. The previous observation that an antibody to CCP 20 can block the cofactor activity of FH (70) is also consistent with the relevant sites being close in space, at least when FH is engaged with the surface-associated convertase (but also, speculatively, in non-liganded FH as discussed below). That FH-19-20 can overcome the protective effects of FH against complement-mediated lysis of erythrocytes (69) also points to a key role for this region in anchoring FH to the surface, perhaps via a composite binding site consisting of C3b and GAGs (64) with specific patterns or densities of sulfation. The aforementioned bend in the central region of FH (and/or a direct or carbohydrate-mediated association between these two GAG-binding regions) would allow CCP 7 to bind nearby, perhaps acting in a "proofreading" role in that it could recognize a second composite C3b-GAG binding site, again containing a particular distribution of sulfates.

Binding of this nature would then place the N-terminal four CCPs at a specific position relative to the C3b(TED) binding site in the C-terminal module and hence to the surface-bound C3b (Figure 6.1). Such positioning might be critical for efficient operation of the N-terminal functional unit in its cofactor and decay-accelerating roles. Involvement of two (or more) sites in binding of FH to a common C3b molecule as discussed in CHAPTER 4 (and see, in particular Figure 4.11 page 166) could explain the stronger K_d of the full-length FH protein compared with that of the tightest binding fragment, FH-19-20 (measured on C1 and CM5 sensor chips). That FH-1-4 binds well to C3b (K_d of 10 or 14 μM when C3b was amine-coupled to a C1- or a CM5 chip, respectively), but lacks affinity for the cleavage fragments C3c and C3d (CHAPTER 4), implies that the binding site for CCPs 1-4 on C3b includes the domain that gets cleaved by factor I (*i.e.* the CUB domain). This inference has been borne out by a recently presented (International Complement Workshop, Basel, 2008) crystal structure of a C3b:FH-1-4 complex.

Thus, in this working model (Figure 6.1), the principal role of some of the central CCP modules is to act as a set of spacers, projecting the functional regions away from a bent-back region (composed from the remaining central CCP modules). This configuration allows the key binding sites to approach each other (or even interact) so as to act cooperatively in selectively engaging and destabilizing the self surface-associated convertase complexes. Such an arrangement is reminiscent of the mechanism employed by the RCA C4b-binding protein, in which functional sites at the tips of seven arms (each consisting of eight CCPs) cooperate in recognizing a composite surface of GAGs and C4b (172). The complement receptor type-1 also employs cooperation between sites that are remote in its sequence, for example to accelerate decay of the C5 convertase (173).

The C3c-binding data (in CHAPTER 4) suggest that once FH has accomplished its cofactor role and the CUB domain is cleaved by Factor I, CCPs 1-4 lose (largely) their affinity for C3c. This, together with the more tentative data that shows a lack of C3d binding by FH, implies that FH is able to dissociate from the products of the reactions it helps to catalyse. More details of the proposed model are presented in the legend to Figure 6.1.

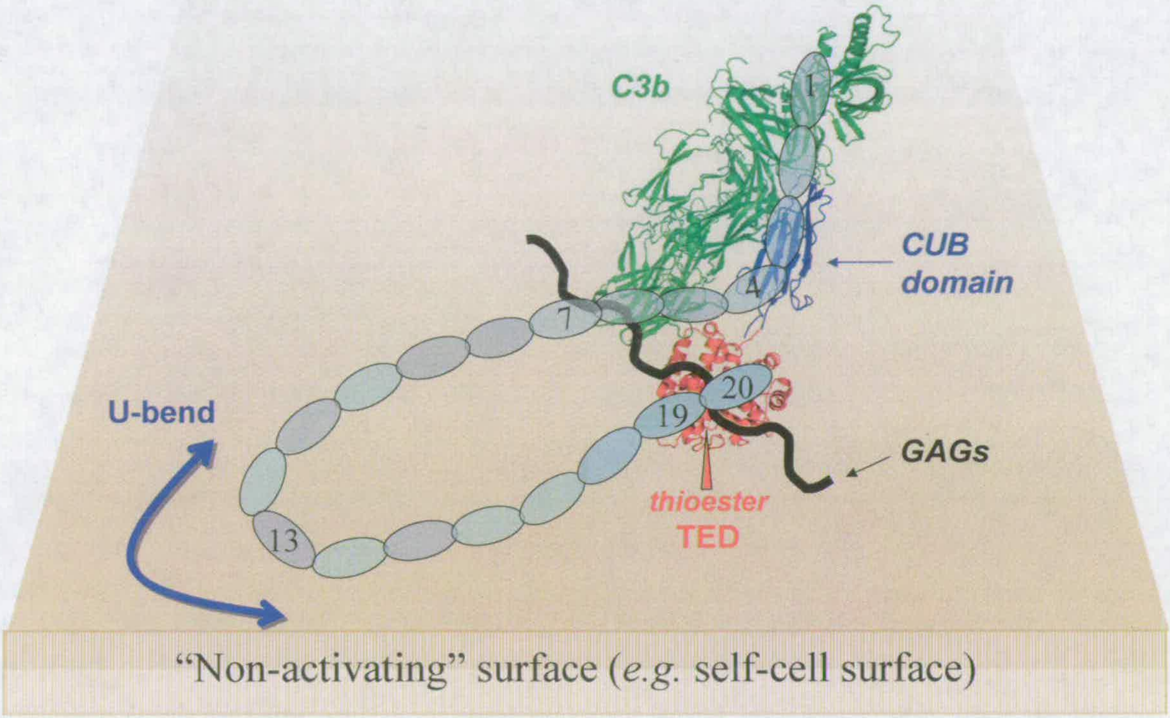


Figure 6.1 In this model FH binds to surface-bound C3b (anchored via its TED domain) co-localised with specific polyanionic GAGs that provides the self-signal for specific complement down-regulation. While CCPs 1-4 of FH associate with the "C3c-like" body of C3b plus the CUB-domain (that is specific to C3b), CCPs 19-20 mainly bind to the TED (*i.e.* the "C3d-like" part of C3b). Additionally, the 19-20 and the 6-8 regions associate with cell-surface GAGs thereby strengthening the affinity of FH for the C3b-bearing surface, and possibly stabilising a functionally active orientation of CCP modules. Subsequent factor I-mediated cleavage of the CUB-domain of C3b leaves opsonic iC3b at the cell surface (iC3b may be further cleaved yielding C3d). The loss of the CUB destroys the N-terminal C3b-binding site within FH and therefore the avidity effect resulting from engagement of both FH-termini with C3b is lost. Whether FH can remain bound to iC3b is untested; likewise its affinity for C3d is (following our tentative results described in CHAPTER 4) a matter for further investigation. The possibility of a *self*-association of the C-terminal part (CCPs 19-20) of FH with CCPs 1-7 – once CCPs 1-4 have lost their affinity for surface immobilised C3 - might present an additional rationale explaining how CCPs 19-20 could be disengaged from the TED of iC3b. Such a self-association of CCPs 19-20 with a region in CCPs 1-7 within full-length FH would also explain why the double module 19-20 binds to C3d, while (in the current work) full-length FH shows comparatively little association with C3d.

6.3 Implications of the working model for the link between mutations/SNPs and disease

The potential for cooperation between CCPs 7 and 20 suggested in CHAPTER 4 is interesting in that such an arrangement has an enhanced capacity for combinatorial recognition of specific types of GAG molecules among diverse possibilities. This supports a model in which complement regulation in tissues displaying different GAGs could be differentially susceptible to mutations and polymorphisms within modules 7 and 20; it is surely no coincidence that between them, these two modules account for nearly all the disease-related sequence variations in FH.

A dominant GAG-binding role of CCP 20 is indeed consistent with the fact that mutations in this region have been linked to aHUS and a failure to prevent complement-mediated damage to endothelial cells and basement membrane in the walls of glomerular capillaries. In recent, unpublished, work (D. Kavanagh, personal communication and V. Ferreira (University of Texas) presentation at the 2008 International Complement Workshop) it was shown that aHUS-linked mutations can result in enhanced binding both to GAGs and C3b when each ligand is studied separately; yet the same mutants have a depleted ability to compete with FH for binding to a C3b-coated cell surface. This intriguing observation is in line with our model that envisages a complex, multivalent interaction of FH with its target that could be upset by an imbalance of affinities. For example if module 20 binds more tightly to C3b it might also bind more tightly to C3d and fail to be released after cleavage of C3b. Alternatively, if a mutant binds too strongly to GAGs it may lose specificity for a putative composite GAG-C3b binding site and bind indiscriminately to GAGs that are remote from C3b.

DISCUSSION

A "proof-reading" role of the heparin-binding site in CCP 7 as proposed in the model is consistent with the link between the H/Y402 SNP and a predisposition to develop AMD. Moreover, a distinct difference in heparin binding between 402 Y and H allotypes, within isolated module 7, was reported in CHAPTER 4. Although such differential GAG binding was only observed when analysed within the single module 7, and was not apparent in the context of FH or FH-6-8, this is an intriguing observation. Since differences (in the context of single modules) were observed for heparin but not for dermatan sulphate, GAG-specific effects might play a role in AMD onset and progression. Even a very low level of differential GAG binding of FH, not detectable in biochemical assays, could gain significance over decades; AMD visual impairment (associated with the 402 polymorphism in FH) manifests predominantly in the elderly.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Herbert, A.P., J.A. Deakin, C.Q. Schmidt, B.S. Blaum, C. Egan, V.P. Ferreira, M.K. Pangburn, M. Lyon, D. Uhrin, and P.N. Barlow. 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.
2. Schmidt, C.Q., A.P. Herbert, H.G. Hocking, D. Uhrin, and P.N. Barlow. 2008. Translational mini-review series on complement factor H: structural and functional correlations for factor H. *Clin Exp Immunol* 151:14-24.
3. Schmidt, C.Q., A.P. Herbert, D. Kavanagh, C. Gandy, C.J. Fenton, B.S. Blaum, M. Lyon, D. Uhrin, and P.N. Barlow. 2008. A new map of glycosaminoglycan and C3b binding sites on factor H. *J Immunol* 181:2610-2619.
4. Sontheimer, R.D., E. Racila, and D.M. Racila. 2005. C1q: its functions within the innate and adaptive immune responses and its role in lupus autoimmunity. *J Invest Dermatol* 125:14-23.
5. Favoreel, H.W., G.R. Van de Walle, H.J. Nauwynck, and M.B. Pensaert. 2003. Virus complement evasion strategies. *J Gen Virol* 84:1-15.
6. Carroll, M.C. 2004. The complement system in regulation of adaptive immunity. *Nat Immunol* 5:981-986.
7. Kohl, J. 2006. The role of complement in danger sensing and transmission. *Immunol Res* 34:157-176.
8. Nonaka, M., and A. Kimura. 2006. Genomic view of the evolution of the complement system. *Immunogenetics* 58:701-713.
9. Sahu, A., and J.D. Lambris. 2000. Complement inhibitors: a resurgent concept in anti-inflammatory therapeutics. *Immunopharmacology* 49:133-148.
10. Seelen, M.A., A. Roos, and M.R. Daha. 2005. Role of complement in innate and autoimmunity. *J Nephrol* 18:642-653.
11. Roitt I., B.J.a.M.D. 1998. Textbook "Immunology". Mosby International Limited,
12. Morgan, P.B., and H.C. L. 1999. Textbook "Complement Regulatory Proteins". Academic Press,
13. Janssen, B.J., E.G. Huizinga, H.C. Raaijmakers, A. Roos, M.R. Daha, K. Nilsson-Ekdahl, B. Nilsson, and P. Gros. 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* 437:505-511.
14. Janssen, B.J., A. Christodoulidou, A. McCarthy, J.D. Lambris, and P. Gros. 2006. Structure of C3b reveals conformational changes that underlie complement activity. *Nature* 444:213-216.
15. Milder, F.J., L. Gomes, A. Schouten, B.J. Janssen, E.G. Huizinga, R.A. Romijn, W. Hemrika, A. Roos, M.R. Daha, and P. Gros. 2007. Factor B structure provides insights into activation of the central protease of the complement system. *Nat Struct Mol Biol* 14:224-228.
16. Gros, P., F.J. Milder, and B.J. Janssen. 2008. Complement driven by conformational changes. *Nat Rev Immunol* 8:48-58.
17. Rooijakkers, S.H.M., and J.A.G. van Strijp. 2006. Bacterial complement evasion. *Molecular Immunology* 44:23-32.
18. DiScipio, R.G. 1992. Formation and structure of the C5b-7 complex of the lytic pathway of complement. *J Biol Chem* 267:17087-17094.
19. Law S. K. A., and Reid K.B.M. 1995. Textbook "Complement". Oxford University Press,
20. Meri, S. 2007. Loss of self-control in the complement system and innate autoreactivity. *Ann N Y Acad Sci* 1109:93-105.
21. Gershov, D., S. Kim, N. Brot, and K.B. Elkon. 2000. C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement

components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity. *J Exp Med* 192:1353-1364.

22. Kemper, C., L.M. Mitchell, L. Zhang, and D.E. Hourcade. 2008. The complement protein properdin binds apoptotic T cells and promotes complement activation and phagocytosis. *Proc Natl Acad Sci U S A* 105:9023-9028.
23. Sokoloff, M.H., A. Nardin, M.D. Solga, M.A. Lindorfer, W.M. Sutherland, A.J. Bankovich, H.E. Zhau, L.W. Chung, and R.P. Taylor. 2000. Targeting of cancer cells with monoclonal antibodies specific for C3b(i). *Cancer Immunol Immunother* 49:551-562.
24. Morgan, B.P., K.J. Marchbank, M.P. Longhi, C.L. Harris, and A.M. Gallimore. 2005. Complement: central to innate immunity and bridging to adaptive responses. *Immunol Lett* 97:171-179.
25. Longhi, M.P., C.L. Harris, B.P. Morgan, and A. Gallimore. 2006. Holding T cells in check--a new role for complement regulators? *Trends Immunol* 27:102-108.
26. Jha, P., and G.J. Kotwal. 2003. Vaccinia complement control protein: multi-functional protein and a potential wonder drug. *J Biosci* 28:265-271.
27. Rooijakkers, S.H., F.J. Milder, B.W. Bardoel, M. Ruyken, J.A. van Strijp, and P. Gros. 2007. Staphylococcal complement inhibitor: structure and active sites. *J Immunol* 179:2989-2998.
28. Kirkitadze, M.D., and P.N. Barlow. 2001. Structure and flexibility of the multiple domain proteins that regulate complement activation. *Immunol Rev* 180:146-161.
29. Blein, S., R. Ginham, D. Uhrin, B.O. Smith, D.C. Soares, S. Veltel, R.A. McIlhinney, J.H. White, and P.N. Barlow. 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* 279:48292-48306.
30. Law S.K.A., and Reid K.B.M. 1995. Textbook "Complement". Oxford University Press,
31. Blom, A.M., B.O. Villoutreix, and B. Dahlback. 2004. Complement inhibitor C4b-binding protein-friend or foe in the innate immune system? *Mol Immunol* 40:1333-1346.
32. URL. 07.08.2008. http://www.ebi.ac.uk/Tools/clustalw2/index.html. In.
33. URL. 07.08.2008. http://www.ch.embnet.org/software/BOX_form.html. In.
34. Soames, C.J., A.J. Day, and R.B. Sim. 1996. Prediction from sequence comparisons of residues of factor H involved in the interaction with complement component C3b. *Biochem J* 315 (Pt 2):523-531.
35. Norman, D.G., P.N. Barlow, M. Baron, A.J. Day, R.B. Sim, and I.D. Campbell. 1991. Three-dimensional structure of a complement control protein module in solution. *J Mol Biol* 219:717-725.
36. Sim, R.B., and R.G. DiScipio. 1982. Purification and structural studies on the complement-system control protein beta 1H (Factor H). *Biochem J* 205:285-293.
37. Meri, S., and M.K. Pangburn. 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 U S A* 87:3982-3986.
38. Pangburn, M.K., R.D. Schreiber, and H.J. Muller-Eberhard. 1977. Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein beta1H for cleavage of C3b and C4b in solution. *J Exp Med* 146:257-270.
39. Ripoche, J., A.J. Day, T.J. Harris, and R.B. Sim. 1988. The complete amino acid sequence of human complement factor H. *Biochem J* 249:593-602.
40. Ripoche, J., A.J. Day, A.C. Willis, K.T. Belt, R.D. Campbell, and R.B. Sim. 1986. Partial characterization of human complement factor H by protein and cDNA

sequencing: homology with other complement and non-complement proteins. *Biosci Rep* 6:65-72.

41. Kristensen, T., and B.F. Tack. 1986. Murine protein H is comprised of 20 repeating units, 61 amino acids in length. *Proc Natl Acad Sci U S A* 83:3963-3967.
42. Discipio, R.G., and T.E. Hugli. 1982. Circular dichroism studies of human factor H. A regulatory component of the complement system. *Biochim Biophys Acta* 709:58-64.
43. Perkins, S.J., P.I. Haris, R.B. Sim, and D. Chapman. 1988. A study of the structure of human complement component factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods. *Biochemistry* 27:4004-4012.
44. Dahlback, B., C.A. Smith, and H.J. Muller-Eberhard. 1983. Visualization of human C4b-binding protein and its complexes with vitamin K-dependent protein S and complement protein C4b. *Proc Natl Acad Sci U S A* 80:3461-3465.
45. Perkins, S.J., L.P. Chung, and K.B. Reid. 1986. Unusual ultrastructure of complement-component-C4b-binding protein of human complement by synchrotron X-ray scattering and hydrodynamic analysis. *Biochem J* 233:799-807.
46. Perkins, S.J., A.S. Nealis, and R.B. Sim. 1991. Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry* 30:2847-2857.
47. DiScipio, R.G. 1992. Ultrastructures and interactions of complement factors H and I. *J Immunol* 149:2592-2599.
48. Aslam, M., and S.J. Perkins. 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.
49. Barlow, P.N., M. Baron, D.G. Norman, A.J. Day, A.C. Willis, R.B. Sim, and I.D. Campbell. 1991. Secondary structure of a complement control protein module by two-dimensional 1H NMR. *Biochemistry* 30:997-1004.
50. Barlow, P.N., A. Steinkasserer, D.G. Norman, B. Kieffer, A.P. Wiles, R.B. Sim, and I.D. Campbell. 1993. Solution structure of a pair of complement modules by nuclear magnetic resonance. *J Mol Biol* 232:268-284.
51. Herbert, A.P., D. Uhrin, M. Lyon, M.K. Pangburn, and P.N. Barlow. 2006. Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. *J Biol Chem* 281:16512-16520.
52. Jokiranta, T.S., V.P. Jaakola, M.J. Lehtinen, M. Parepalo, S. Meri, and A. Goldman. 2006. Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *Embo J* 25:1784-1794.
53. Prosser, B.E., S. Johnson, P. Roversi, A.P. Herbert, B.S. Blaum, J. Tyrrell, T.A. Jowitt, S.J. Clark, E. Tarelli, D. Uhrin, P.N. Barlow, R.B. Sim, A.J. Day, and S.M. Lea. 2007. Structural basis for complement factor H linked age-related macular degeneration. *J Exp Med* 204:2277-2283.
54. Jouvin, M.H., M.D. Kazatchkine, A. Cahour, and N. Bernard. 1984. Lysine residues, but not carbohydrates, are required for the regulatory function of H on the amplification C3 convertase of complement. *J Immunol* 133:3250-3254.
55. Alsenz, J., J.D. Lambris, T.F. Schulz, and M.P. Dierich. 1984. Localization of the complement-component-C3b-binding site and the cofactor activity for factor I in the 38kDa tryptic fragment of factor H. *Biochem J* 224:389-398.
56. Alsenz, J., T.F. Schulz, J.D. Lambris, R.B. Sim, and M.P. Dierich. 1985. Structural and functional analysis of the complement component factor H with the use of different enzymes and monoclonal antibodies to factor H. *Biochem J* 232:841-850.

57. Misasi, R., H.P. Huemer, W. Schwaeble, E. Solder, C. Larcher, and M.P. Dierich. 1989. Human complement factor H: an additional gene product of 43 kDa isolated from human plasma shows cofactor activity for the cleavage of the third component of complement. *Eur J Immunol* 19:1765-1768.
58. Gordon, D.L., R.M. Kaufman, T.K. Blackmore, J. Kwong, and D.M. Lublin. 1995. Identification of complement regulatory domains in human factor H. *J Immunol* 155:348-356.
59. Kuhn, S., C. Skerka, and P.F. Zipfel. 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. Jokiranta, T.S., J. Hellwage, V. Koistinen, P.F. Zipfel, and S. Meri. 2000. Each of the three binding sites on complement factor H interacts with a distinct site on C3b. *J Biol Chem* 275:27657-27662.
61. Ormsby, R.J., T.S. Jokiranta, T.G. Duthy, K.M. Griggs, T.A. Sadlon, E. Giannakis, and D.L. Gordon. 2006. Localization of the third heparin-binding site in the human complement regulator factor H1. *Mol Immunol* 43:1624-1632.
62. Blackmore, T.K., J. Hellwage, T.A. Sadlon, N. Higgs, P.F. Zipfel, H.M. Ward, and D.L. Gordon. 1998. Identification of the second heparin-binding domain in human complement factor H. *J Immunol* 160:3342-3348.
63. Blackmore, T.K., T.A. Sadlon, H.M. Ward, D.M. Lublin, and D.L. Gordon. 1996. Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J Immunol* 157:5422-5427.
64. Hellwage, J., T.S. Jokiranta, M.A. Friese, T.U. Wolk, E. Kampen, P.F. Zipfel, and S. Meri. 2002. Complement C3b/C3d 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.
65. Sharma, A.K., and M.K. Pangburn. 1996. Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc Natl Acad Sci U S A* 93:10996-11001.
66. Pangburn, M.K. 2002. Cutting edge: localization of the host recognition functions of complement factor H at the carboxyl-terminal: implications for hemolytic uremic syndrome. *J Immunol* 169:4702-4706.
67. Pangburn, M.K., M.A. Atkinson, and S. Meri. 1991. Localization of the heparin-binding site on complement factor H. *J Biol Chem* 266:16847-16853.
68. Clark, S.J., V.A. Higman, B. Mulloy, S.J. Perkins, S.M. Lea, R.B. Sim, and A.J. Day. 2006. His-384 allotypic variant of factor H associated with age-related macular degeneration has different heparin binding properties from the non-disease-associated form. *J Biol Chem* 281:24713-24720.
69. Ferreira, V.P., A.P. Herbert, H.G. Hocking, P.N. Barlow, and M.K. Pangburn. 2006. Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J Immunol* 177:6308-6316.
70. Oppermann, M., T. Manuelian, M. Jozsi, E. Brandt, T.S. Jokiranta, S. Heinen, S. Meri, C. Skerka, O. Gotze, and P.F. Zipfel. 2006. The C-terminus of complement regulator Factor H mediates target recognition: evidence for a compact conformation of the native protein. *Clin Exp Immunol* 144:342-352.
71. Mold, C., H. Gewurz, and T.W. Du Clos. 1999. Regulation of complement activation by C-reactive protein. *Immunopharmacology* 42:23-30.
72. Sjoberg, A., P. Onnerfjord, M. Morgelin, D. Heinegard, and A.M. Blom. 2005. The extracellular matrix and inflammation: fibromodulin activates the classical pathway of complement by directly binding C1q. *J Biol Chem* 280:32301-32308.
73. Pio, R., A. Martinez, E.J. Unsworth, J.A. Kowalak, J.A. Bengoechea, P.F. Zipfel, T.H. Elsasser, and F. Cuttitta. 2001. Complement factor H is a serum-binding

protein for adrenomedullin, and the resulting complex modulates the bioactivities of both partners. *J Biol Chem* 276:12292-12300.

74. Zipfel, P.F., C. Skerka, J. Hellwage, S.T. Jokiranta, S. Meri, V. Brade, P. Kraiczy, M. Noris, and G. Remuzzi. 2002. Factor H family proteins: on complement, microbes and human diseases. *Biochem Soc Trans* 30:971-978.
75. Zipfel, P.F., R. Wurzner, and C. Skerka. 2007. Complement evasion of pathogens: common strategies are shared by diverse organisms. *Mol Immunol* 44:3850-3857.
76. Rooijakkers, S.H., and J.A. van Strijp. 2007. Bacterial complement evasion. *Mol Immunol* 44:23-32.
77. McRae, J.L., T.G. Duthy, K.M. Griggs, R.J. Ormsby, P.J. Cowan, B.A. Cromer, W.J. McKinstry, M.W. Parker, B.F. Murphy, and D.L. Gordon. 2005. Human factor H-related protein 5 has cofactor activity, inhibits C3 convertase activity, binds heparin and C-reactive protein, and associates with lipoprotein. *J Immunol* 174:6250-6256.
78. Jozsi, M., and P.F. Zipfel. 2008. Factor H family proteins and human diseases. *Trends Immunol* 29:380-387.
79. Kavanagh, D., T.H. Goodship, and A. Richards. 2006. Atypical haemolytic uraemic syndrome. *Br Med Bull* 77-78:5-22.
80. Perez-Caballero, D., C. Gonzalez-Rubio, M.E. Gallardo, M. Vera, M. Lopez-Trascasa, S. Rodriguez de Cordoba, and P. Sanchez-Corral. 2001. Clustering of missense mutations in the C-terminal region of factor H in atypical hemolytic uremic syndrome. *Am J Hum Genet* 68:478-484.
81. Pickering, M.C., H.T. Cook, J. Warren, A.E. Bygrave, J. Moss, M.J. Walport, and M. Botto. 2002. Uncontrolled C3 activation causes membranoproliferative glomerulonephritis in mice deficient in complement factor H. *Nat Genet* 31:424-428.
82. Pickering, M.C., E.G. de Jorge, R. Martinez-Barricarte, S. Recalde, A. Garcia-Layana, K.L. Rose, J. Moss, M.J. Walport, H.T. Cook, S.R. de Cordoba, and M. Botto. 2007. Spontaneous hemolytic uremic syndrome triggered by complement factor H lacking surface recognition domains. *J Exp Med* 204:1249-1256.
83. Verdugo, M.E., and J. Ray. 1997. Age-related increase in activity of specific lysosomal enzymes in the human retinal pigment epithelium. *Exp Eye Res* 65:231-240.
84. Meri, S., and M.K. Pangburn. 1994. Regulation of alternative pathway complement activation by glycosaminoglycans: specificity of the polyanion binding site on factor H. *Biochem Biophys Res Commun* 198:52-59.
85. Jozsi, M., S. Heinen, A. Hartmann, C.W. Ostrowicz, S. Halbich, H. Richter, A. Kunert, C. Licht, R.E. Saunders, S.J. Perkins, P.F. Zipfel, and C. Skerka. 2006. Factor H and atypical hemolytic uremic syndrome: mutations in the C-terminus cause structural changes and defective recognition functions. *J Am Soc Nephrol* 17:170-177.
86. Manuelian, T., J. Hellwage, S. Meri, J. Caprioli, M. Noris, S. Heinen, M. Jozsi, H.P. Neumann, G. Remuzzi, and P.F. Zipfel. 2003. Mutations in factor H reduce binding affinity to C3b and heparin and surface attachment to endothelial cells in hemolytic uremic syndrome. *J Clin Invest* 111:1181-1190.
87. Jokiranta, T.S., Z.Z. Cheng, H. Seeberger, M. Jozsi, S. Heinen, M. Noris, G. Remuzzi, R. Ormsby, D.L. Gordon, S. Meri, J. Hellwage, and P.F. Zipfel. 2005. Binding of complement factor H to endothelial cells is mediated by the carboxy-terminal glycosaminoglycan binding site. *Am J Pathol* 167:1173-1181.
88. Sanchez-Corral, P., C. Gonzalez-Rubio, S. Rodriguez de Cordoba, and M. Lopez-Trascasa. 2004. Functional analysis in serum from atypical Hemolytic Uremic Syndrome patients reveals impaired protection of host cells associated with mutations in factor H. *Mol Immunol* 41:81-84.

89. Sanchez-Corral, P., D. Perez-Caballero, O. Huarte, A.M. Simckes, E. Goicoechea, M. Lopez-Trascasa, and S.R. de Cordoba. 2002. Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome. *Am J Hum Genet* 71:1285-1295.

90. Heinen, S., P. Sanchez-Corral, M.S. Jackson, L. Strain, J.A. Goodship, E.J. Kemp, C. Skerka, T.S. Jokiranta, K. Meyers, E. Wagner, P. Robitaille, J. Esparza-Gordillo, S. Rodriguez de Cordoba, P.F. Zipfel, and T.H. Goodship. 2006. De novo gene conversion in the RCA gene cluster (1q32) causes mutations in complement factor H associated with atypical hemolytic uremic syndrome. *Hum Mutat* 27:292-293.

91. Vaziri-Sani, F., L. Holmberg, A.G. Sjoholm, A.C. Kristoffersson, M. Manea, V. Fremeaux-Bacchi, I. Fehrman-Ekholm, R. Raafat, and D. Karpman. 2006. Phenotypic expression of factor H mutations in patients with atypical hemolytic uremic syndrome. *Kidney Int* 69:981-988.

92. Shaul, Y., and G. Schreiber. 2005. Exploring the charge space of protein-protein association: a proteomic study. *Proteins* 60:341-352.

93. Smith, R.J., J. Alexander, P.N. Barlow, M. Botto, T.L. Cassavant, H.T. Cook, S.R. de Cordoba, G.S. Hageman, T.S. Jokiranta, W.J. Kimberling, J.D. Lambris, L.D. Lanning, V. Levidiotis, C. Licht, H.U. Lutz, S. Meri, M.C. Pickering, R.J. Quigg, A.L. Rops, D.J. Salant, S. Sethi, J.M. Thurman, H.F. Tully, S.P. Tully, J. van der Vlag, P.D. Walker, R. Wurzner, and P.F. Zipfel. 2007. New approaches to the treatment of dense deposit disease. *J Am Soc Nephrol* 18:2447-2456.

94. Appel, G.B., H.T. Cook, G. Hageman, J.C. Jennette, M. Kashgarian, M. Kirschfink, J.D. Lambris, L. Lanning, H.U. Lutz, S. Meri, N.R. Rose, D.J. Salant, S. Sethi, R.J. Smith, W. Smoyer, H.F. Tully, S.P. Tully, P. Walker, M. Welsh, R. Wurzner, and P.F. Zipfel. 2005. Membranoproliferative glomerulonephritis type II (dense deposit disease): an update. *J Am Soc Nephrol* 16:1392-1403.

95. Pickering, M.C., and H.T. Cook. 2008. Translational mini-review series on complement factor H: renal diseases associated with complement factor H: novel insights from humans and animals. *Clin Exp Immunol* 151:210-230.

96. Mullins, R.F., N. Aptsiauri, and G.S. Hageman. 2001. Structure and composition of drusen associated with glomerulonephritis: implications for the role of complement activation in drusen biogenesis. *Eye* 15:390-395.

97. Abrera-Abeleda, M.A., C. Nishimura, J.L. Smith, S. Sethi, J.L. McRae, B.F. Murphy, G. Silvestri, C. Skerka, M. Jozsi, P.F. Zipfel, G.S. Hageman, and R.J. Smith. 2006. Variations in the complement regulatory genes factor H (CFH) and factor H related 5 (CFHR5) are associated with membranoproliferative glomerulonephritis type II (dense deposit disease). *J Med Genet* 43:582-589.

98. Hageman, G.S., D.H. Anderson, L.V. Johnson, L.S. Hancox, A.J. Taiber, L.I. Hardisty, J.L. Hageman, H.A. Stockman, J.D. Borchardt, K.M. Gehrs, R.J. Smith, G. Silvestri, S.R. Russell, C.C. Klaver, I. Barbazetto, S. Chang, L.A. Yannuzzi, G.R. Barile, J.C. Merriam, R.T. Smith, A.K. Olsh, J. Bergeron, J. Zernant, J.E. Merriam, B. Gold, M. Dean, and R. Allikmets. 2005. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci U S A* 102:7227-7232.

99. Thakkinstian, A., P. Han, M. McEvoy, W. Smith, J. Hoh, K. Magnusson, K. Zhang, and J. Attia. 2006. Systematic review and meta-analysis of the association between complement factor H Y402H polymorphisms and age-related macular degeneration. *Hum Mol Genet* 15:2784-2790.

100. Giannakis, E., T.S. Jokiranta, D.A. Male, S. Ranganathan, R.J. Ormsby, V.A. Fischetti, C. Mold, and D.L. Gordon. 2003. A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein. *Eur J Immunol* 33:962-969.

101. Sjoberg, A.P., L.A. Trouw, S.J. Clark, J. Sjolander, D. Heinegard, R.B. Sim, A.J. Day, and A.M. Blom. 2007. The factor H variant associated with age-related macular degeneration (His-384) and the non-disease-associated form bind differentially to C-reactive protein, fibromodulin, DNA, and necrotic cells. *J Biol Chem* 282:10894-10900.
102. Blackmore, T.K., V.A. Fischetti, T.A. Sadlon, H.M. Ward, and D.L. Gordon. 1998. M protein of the group A Streptococcus binds to the seventh short consensus repeat of human complement factor H. *Infect Immun* 66:1427-1431.
103. Kraiczy, P., J. Hellwage, C. Skerka, H. Becker, M. Kirschfink, M.M. Simon, V. Brade, P.F. Zipfel, and R. Wallich. 2004. Complement resistance of Borrelia burgdorferi correlates with the expression of BbCRASP-1, a novel linear plasmid-encoded surface protein that interacts with human factor H and FHL-1 and is unrelated to Erp proteins. *J Biol Chem* 279:2421-2429.
104. Laine, M., H. Jarva, S. Seitsonen, K. Haapasalo, M.J. Lehtinen, N. Lindeman, D.H. Anderson, P.T. Johnson, I. Jarvela, T.S. Jokiranta, G.S. Hageman, I. Immonen, and S. Meri. 2007. Y402H polymorphism of complement factor H affects binding affinity to C-reactive protein. *J Immunol* 178:3831-3836.
105. Yu, J., P. Wiita, R. Kawaguchi, J. Honda, A. Jorgensen, K. Zhang, V.A. Fischetti, and H. Sun. 2007. Biochemical analysis of a common human polymorphism associated with age-related macular degeneration. *Biochemistry* 46:8451-8461.
106. Skerka, C., N. Lauer, A.A. Weinberger, C.N. Keilhauer, J. Suhnel, R. Smith, U. Schlotzer-Schrehardt, L. Fritsche, S. Heinen, A. Hartmann, B.H. Weber, and P.F. Zipfel. 2007. Defective complement control of factor H (Y402H) and FHL-1 in age-related macular degeneration. *Mol Immunol* 44:3398-3406.
107. Johnson, P.T., K.E. Betts, M.J. Radeke, G.S. Hageman, D.H. Anderson, and L.V. Johnson. 2006. Individuals homozygous for the age-related macular degeneration risk-conferring variant of complement factor H have elevated levels of CRP in the choroid. *Proc Natl Acad Sci U S A* 103:17456-17461.
108. Hakobyan, S., C.L. Harris, C. van den Berg, M.B. Pepys, and M. P.B. Binding of factor H to C-reactive protein occurs only when the latter has undergone non-physiologic denaturation. *Mol Immunol* 44:3983-3984.
109. Cereghino, J.L., and J.M. Cregg. 2000. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev* 24:45-66.
110. Cregg, J.M., J.L. Cereghino, J. Shi, and D.R. Higgins. 2000. Recombinant protein expression in Pichia pastoris. *Mol Biotechnol* 16:23-52.
111. Cregg, J.M., T.S. Vedvick, and W.C. Raschke. 1993. Recent advances in the expression of foreign genes in Pichia pastoris. *Biotechnology (N Y)* 11:905-910.
112. Gellissen, G. 2000. Heterologous protein production in methylotrophic yeasts. *Appl Microbiol Biotechnol* 54:741-750.
113. Macauley-Patrick, S., M.L. Fazenda, B. McNeil, and L.M. Harvey. 2005. Heterologous protein production using the Pichia pastoris expression system. *Yeast* 22:249-270.
114. Clare, J.J., F.B. Rayment, S.P. Ballantine, K. Sreekrishna, and M.A. Romanos. 1991. High-level expression of tetanus toxin fragment C in Pichia pastoris strains containing multiple tandem integrations of the gene. *Biotechnology (N Y)* 9:455-460.
115. URL. 26.08.2008. http://www.expasy.ch/tools/protparam.html. In.
116. Lyon, M., J.A. Deakin, D. Lietha, E. Gherardi, and J.T. Gallagher. 2004. The interactions of hepatocyte growth factor/scatter factor and its NK1 and NK2 variants with glycosaminoglycans using a modified gel mobility shift assay. Elucidation of the minimal size of binding and activatory oligosaccharides. *J Biol Chem* 279:43560-43567.

117. Catlow, K.R., J.A. Deakin, Z. Wei, M. Delehedde, D.G. Fernig, E. Gherardi, J.T. Gallagher, M.S. Pavao, and M. Lyon. 2008. Interactions of hepatocyte growth factor/scatter factor with various glycosaminoglycans reveal an important interplay between the presence of iduronate and sulfate density. *J Biol Chem* 283:5235-5248.
118. Bodenhausen, G., and D.J. Ruben. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* 69:185.
119. Vuister, G.W., and A. Bax. 1992. Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling. *J. Magn. Reson.* 98:428-435.
120. Grzesiek, S., and A. Bax. 1992. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* 114:6291.
121. Grzesiek, S., and A. Bax. 1992. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* 99:201.
122. Grzesiek, S., and A. Bax. 1993. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J. Biomol. NMR* 3:185-204.
123. Wang, A.C., P.J. Lodi, J. Qin, G.W. Vuister, A.M. Gronenborn, and G.M. Clore. 1994. An Efficient Triple-Resonance Experiment for Proton-Directed Sequential Backbone Assignment of Medium-Sized Proteins. *J. Magn. Reson. Ser. B* 105:196.
124. Grzesiek, S., and A. Bax. 1992. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* 96:432.
125. Clubb, R.T., V. Thanabal, and G. Wagner. 1992. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C' chemical shifts in 15N---13C-labelled proteins. *J. Magn. Reson.* 97:213.
126. Montelione, G.T., B.A. Lyons, S.D. Emerson, and M. Tashiro. 1992. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.* 114:10974-10975.
127. Sklenar, V., M. Piotto, R. Leppik, and V. Saudek. 1993. Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson. Ser. A* 102:241.
128. Kay, L.E., G.Y. Xu, A.U. Singer, D.R. Muhandiram, and J.D. Formankay. 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.
129. Yamazaki, T., J.D. Forman-Kay, and L.E. Kay. 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.
130. Pascal, S.M., D.R. Muhandiram, T. Yamazaki, J.D. Formankay, and L.E. Kay. 1994. Simultaneous Acquisition of 15N- and 13C-Edited NOE Spectra of Proteins Dissolved in H2O. *J. Magn. Reson. Ser. B* 103:197.
131. Vranken, W.F., W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, and E.D. Laue. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687-696.
132. Herrmann, T., P. Guntert, and K. Wuthrich. 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.
133. Schubert, M., D. Labudde, H. Oschkinat, and P. Schmieder. 2002. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *J Biomol NMR* 24:149-154.

134. Guntert, P., C. Mumenthaler, and K. Wuthrich. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273:283-298.
135. Guntert, P. 2004. Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353-378.
136. Linge, J.P., M.A. Williams, C.A. Spronk, A.M. Bonvin, and M. Nilges. 2003. Refinement of protein structures in explicit solvent. *Proteins* 50:496-506.
137. Brunger, A.T., P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Panriu, R.J. Read, L.M. Rice, T. Simonson, and G.L. Warren. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905-921.
138. Nilges, M., G.M. Clore, and A.M. Gronenborn. 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.
139. Nilges, M. 1995. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J Mol Biol* 245:645-660.
140. Koradi, R., M. Billeter, and K. Wuthrich. 1996. *J Mol Graph* 14: 51-55, 29-32.
141. DeLano, W.L.T.P.M.G.S. 2002. The PyMOL Molecular Graphics System. In., Palo Alto, CA.
142. Nicholls, A., K.A. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11:281-296.
143. Dam, J., and P. Schuck. 2004. Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. *Methods Enzymol* 384:185-212.
144. Lebowitz, J., M.S. Lewis, and P. Schuck. 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci* 11:2067-2079.
145. Hocking, H. 2008. The structure of an active N-terminal fragment of human complement factor H. In University of Edinburgh, Edinburgh.
146. Higgins, D.R., and J.M. Cregg. 1998. Introduction to Pichia pastoris. *Methods Mol Biol* 103:1-15.
147. Grinna, L.S., and J.F. Tschopp. 1989. Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, Pichia pastoris. *Yeast* 5:107-115.
148. Hocking, H.G., A.P. Herbert, D. Kavanagh, D.C. Soares, V.P. Ferreira, M.K. Pangburn, D. Uhrin, and P.N. Barlow. 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.
149. Ahmed, A.E., and K. Whaley. 1988. Purification of a plasma protein that inhibits complement-mediated prevention of immune precipitation. *Immunology* 64:45-50.
150. Emberson, L.M., A.J. Trivett, P.J. Blower, and P.J. Nicholls. 2005. Expression of an anti-CD33 single-chain antibody by Pichia pastoris. *J Immunol Methods* 305:135-151.
151. Reverter, D., S. Ventura, V. Villegas, J. Vendrell, and F.X. Aviles. 1998. Overexpression of human procarboxypeptidase A2 in Pichia pastoris and detailed characterization of its activation pathway. *J Biol Chem* 273:3535-3541.
152. Ghosalkar, A., V. Sahai, and A. Srivastava. 2008. Secretory expression of interferon-alpha 2b in recombinant Pichia pastoris using three different secretion signals. *Protein Expr Purif* 60:103-109.

153. Werten, M.W., and F.A. de Wolf. 2005. Reduced proteolysis of secreted gelatin and Yps1-mediated alpha-factor leader processing in a Pichia pastoris kex2 disruptant. *Appl Environ Microbiol* 71:2310-2317.

154. Ledgerwood, E.C., P.M. George, R.J. Peach, and S.O. Brennan. 1995. Endoproteolytic processing of recombinant proalbumin variants by the yeast Kex2 protease. *Biochem J* 308 (Pt 1):321-325.

155. Bader, O., Y. Krauke, and B. Hube. 2008. Processing of predicted substrates of fungal Kex2 proteinases from Candida albicans, C. glabrata, Saccharomyces cerevisiae and Pichia pastoris. *BMC Microbiol* 8:116.

156. Rockwell, N.C., and J.W. Thorner. 2004. The kindest cuts of all: crystal structures of Kex2 and furin reveal secrets of precursor processing. *Trends Biochem Sci* 29:80-87.

157. Cabral, K.M., M.S. Almeida, A.P. Valente, F.C. Almeida, and E. Kurtenbach. 2003. Production of the active antifungal Pisum sativum defensin 1 (Psd1) in Pichia pastoris: overcoming the inefficiency of the STE13 protease. *Protein Expr Purif* 31:115-122.

158. Daly, R., and M.T. Hearn. 2005. Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. *J Mol Recognit* 18:119-138.

159. Raemaekers, R.J., L. de Muro, J.A. Gatehouse, and A.P. Fordham-Skelton. 1999. Functional phytohemagglutinin (PHA) and Galanthus nivalis agglutinin (GNA) expressed in Pichia pastoris correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. *Eur J Biochem* 265:394-403.

160. Kobayashi, K., S. Kuwae, T. Ohya, T. Ohda, M. Ohyama, H. Ohi, K. Tomomitsu, and T. Ohmura. 2000. High-level expression of recombinant human serum albumin from the methylotrophic yeast Pichia pastoris with minimal protease production and activation. *J Biosci Bioeng* 89:55-61.

161. Kennedy, J.F., and A. Rosevear. 1973. An assessment of the fractionation of carbohydrates on concanavalin A-sepharose 4B by affinity chromatography. *J Chem Soc [Perkin 1]* 19:2041-2046.

162. URL. 20.10.2008. http://www.neb.com/nebecomm/products/productP0703.asp. In.

163. Hamilton, S.R., and T.U. Gerngross. 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol* 18:387-392.

164. URL. 12.10.2008. http://swift.cmbi.ru.nl/servers/html/oldqua.html. In WHATIF.

165. Morris, A.L., M.W. MacArthur, E.G. Hutchinson, and J.M. Thornton. 1992. Stereochemical quality of protein structure coordinates. *Proteins* 12:345-364.

166. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan. 1963. Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95-99.

167. Jenkins, H.T., L. Mark, G. Ball, J. Persson, G. Lindahl, D. Uhrin, A.M. Blom, and P.N. Barlow. 2006. Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor. *J Biol Chem* 281:3690-3697.

168. Hutchinson, E.G., and J.M. Thornton. 1994. A revised set of potentials for beta-turn formation in proteins. *Protein Sci* 3:2207-2216.

169. Laue, T.M., and W.F. Stafford, 3rd. 1999. Modern applications of analytical ultracentrifugation. *Annu Rev Biophys Biomol Struct* 28:75-100.

170. Szakonyi, G., J.M. Guthridge, D. Li, K. Young, V.M. Holers, and X.S. Chen. 2001. Structure of complement receptor 2 in complex with its C3d ligand. *Science* 292:1725-1728.

171. Kuhn, S., and P.F. Zipfel. 1996. Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur J Immunol* 26:2383-2387.

172. Ogata, R.T., P. Mathias, B.M. Bradt, and N.R. Cooper. 1993. Murine C4b-binding protein. Mapping of the ligand binding site and the N-terminus of the pre-protein. *J Immunol* 150:2273-2280.

173. Krych-Goldberg, M., R.E. Hauhart, T. Porzukowiak, and J.P. Atkinson. 2005. Synergy between two active sites of human complement receptor type 1 (CD35) in complement regulation: implications for the structure of the classical pathway C3 convertase and generation of more potent inhibitors. *J Immunol* 175:4528-4535.

APPENDIX A

REAGENTS

Standard chemicals for preparation of buffers and media, as well as standard laboratory consumables were purchased either from Sigma-Aldrich Company Ltd. (Gillingham, UK) or (Fisher Scientific UK Ltd. (Loughborough, UK).

All media containing antibiotics were stored at 4°C.

Antibiotica

- Zeocin™ (ant-zn-1) was acquired from Autogen Bioclear UK Ltd. (Calne, UK).
- Ampicillin (A1001-5) was bought from Cambridge BioScience (Cambridge, UK).

Enzymes

All restriction enzymes (*Xba*I – R0145S; *Pst*I – R0140S; *Sac*I – R0156S; *Bts*I – R0614S; *Hph*I – R0158S), as well as EndoH$_f$® (P0703S) and the Quick Ligation® Kit (M2200S) were purchased from New England Biolabs Ltd. (Hitchin, UK).

SDS-PAGE

- Pre-stained Protein Marker, Broad Range (6-175 kDa) [P7708S] and Quick-Load® were obtained from New England Biolabs UK.
- Precision Plus Protein All Blue Standards (161-0373), Bio-Safe Coomassie stain (161-0786); Polyacrylamide Ready-Gel 4–20% (161-1105) were purchased from Bio-Rad Laboratories Ltd. (Hemel Hempstead, UK).
- NuPAGE® Novex 4-12% Bis-Tris Gel, NuPAGE® Sample Reducing Agent, NuPAGE® MOPS and MES buffers were obtained from Invitrogen (Paisley, UK).

DNA-Agarose-gel

- 100 base pair (N3231S) and 1 kb DNA Ladder (N0468S) were obtained from New England Biolabs UK.
- Sybrsafe® was obtained from Invitrogen (Paisley, UK).

PCR

- PCR Master mix (M7502) was bought from Promega Cooperations (Southampton, UK)
- PfuTurbo® Hotstart DNA Polymerase (#600320) and Herculase® Hotstart DNA Polymerase (#600310) were purchased from Stratagene (Stratagene European Headquarters, Amsterdam Zuidoost, The Netherlands)

Vectors

All vectors used were purchased from Invitrogen (Paisley, UK).

- *P. pastoris* expression vectors pPICZαA and pPICZαB.
- Blunt end TOPO®-cloning vectors pUB/Bsd-TOPO® and pCR®4Blunt-TOPO®

Cells

P. pastoris strain KM71H and chemical competent *E. coli* one shot TOP10® cells were bought from Invitrogen (Paisley, UK).

APPENDIX B

BUFFER and MEDIA

Media	**Composition** (w/v unless otherwise stated)
Basal Salts for unlabelled fermentor growth	4% (v/v) glycerol 0.413% potassium hydroxide 1.5% magnesium sulphate heptahydrate 1.82% potassium sulphate 0.093% calcium sulphate 2.67% (v/v) ortho-phosphoric acid
Basal Salts for ^{15}N-labelled fermentor growth	3.125% (v/v) glycerol 1.5% magnesium sulphate heptahydrate 1% potassium sulphate 0.093% calcium sulphate 10% (v/v) 200 mM potassium phosphate pH 5
Basal Salts for ^{15}N,^{13}C-labelled fermentor growth	1.5% magnesium sulphate heptahydrate 1% potassium sulphate 0.093% calcium sulphate 10% (v/v) 200 mM potassium phospate pH 5
BMG (Buffered minimal glycerol)	100 mM Potassium phosphate pH 6 1.34% YNB 4 x 10-5 % biotin 1% glycerol
BMM (Buffered minimal methanol)	100 mM potassium phosphate pH 6 1.34% YNB 4 x 10-5 % biotin 0.5% methanol
LB (Lysogeny Broth) Lennox (Low salt (less than 5 g/L) is required for efficient selection with Zeocin™)	0.5% yeast extract 1% tryptone 0.5% sodium chloride +/- 1.5% agar
LB (Broth) Miller	0.5% yeast extract

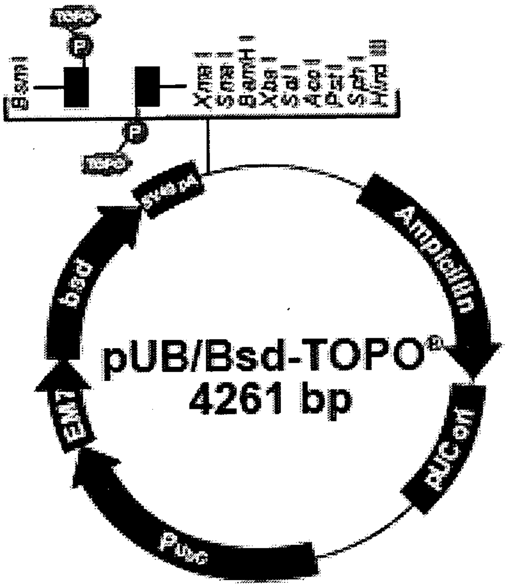
	1% tryptone 1% sodium chloride +/- 1.5% agar
SOC (Super Optimal broth with Catabolite repression)	0.5% yeast extract 2% tryptone 10 mM sodium chloride 2.5 mM potassium chloride 10 mM magnesium chloride 10 mM magnesium sulphate 20 mM D-glucose
YNB (Yeast Nitrogen Base)	10x YNB stock (Sigma) with ammonium sulphate without amino acids
YPD (Yeast Extract, Peptone, Dextrose)	1% yeast extract 2% peptone 2% dextrose (D-glucose) +/- 1.5% agar
YPDS (Yeast Extract, Peptone, Dextrose Sorbitol)	1% yeast extract 2% peptone 2% dextrose (D-glucose) 1 M sorbitol +/- 1.5% agar +/- 100-300 μg/ml Zeocin
EDTA stock solution (ethylenediaminetetraacetic acid)	0.5 M stock adjusted to pH 8 with sodium hydroxide pellets
SDS-PAGE sample loading buffer	50 mM Tri-HCl 100 mM %-mercaptoethanol 2% Sodium-dodecyl-sulfate 0.1% bromophenol blue 10% glycerol
Towbin buffer	10% Methanol 25 mM Tris Base 192 mM Glycine
TCA (trichloroacetic acid)	30% stock (Sigma) made up in H2O
Tris-Glycine-SDS Buffer (TGS)	0.025 M Tris-Base 0.192 M Glycine 0.1% SDS (w/v) pH 8.3

TAE buffer 50x in distilled H2O	Tris-acetate (2 M) EDTA (100 mM)
TAE Agarose	agarose 1 % (w/v) in TAE buffer

APPENDIX C

VECTOR MAPS

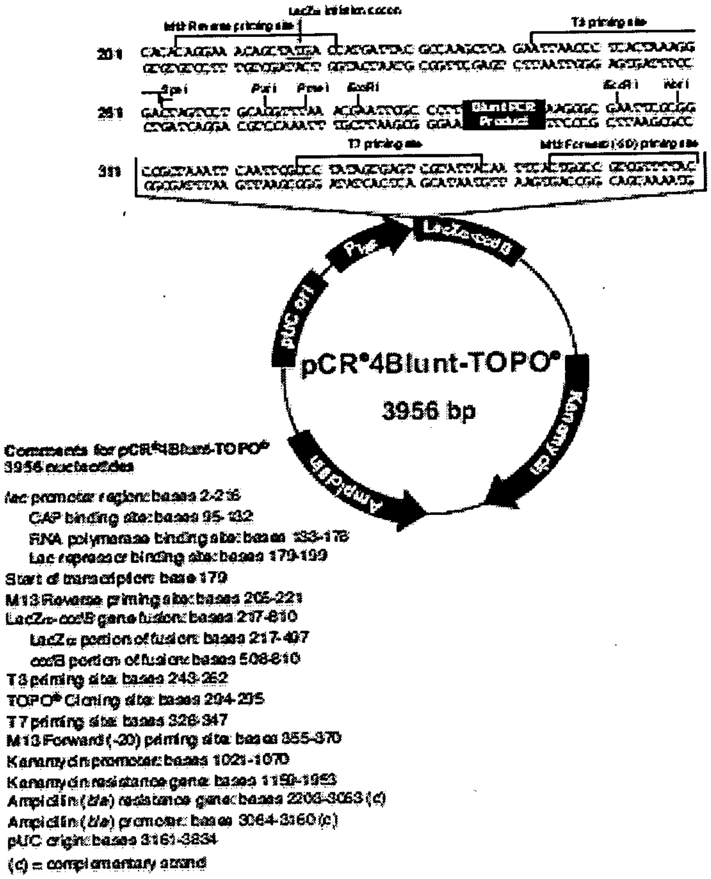
A) pUB/Bsd-TOPO®:


TOPO
P
Bsm I
Xma I
Sma I
BamH I
Xba I
Sal I
Acc I
Pst I
Sph I
Hind III
P
TOPO
SV40 pA
Ampicillin
pUC ori
P_{UbC}
EM7
bsd
pUB/Bsd-TOPO®
4261 bp

Comments for pUB/Bsd-TOPO®
4261 nucleotides

SV40 early polyadenylation sequence: bases 1-131
TOPO® cloning site: bases 141-142
bla promoter: bases 694-793
Ampicillin (*bla*) resistance gene: bases 793-1653
pUC origin: bases 1798-2471
Human ubiquitin C (*UbC*) promoter: bases 2533-3747
EM7 promoter: bases 3753-3803
Blasticidin (*bsd*) resistance gene: bases 3822-4220

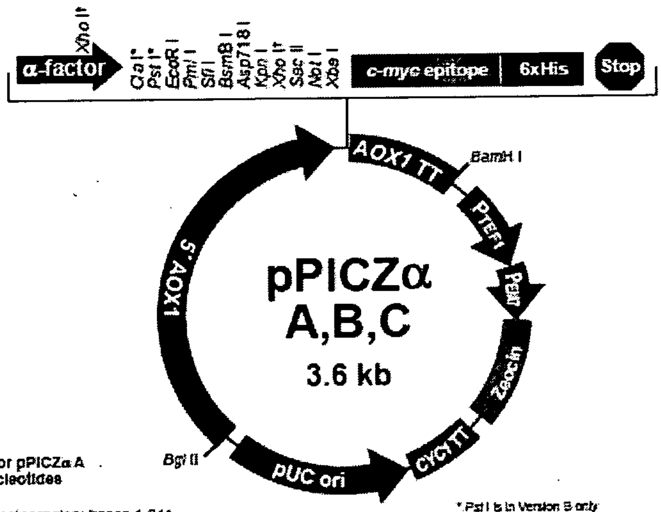
B) pCR®4Blunt-TOPO®:


LacZα initiation codon
M13 Reverse priming site
T3 priming site
201
251
311
Spe I
Pst I
Pme I
EcoR I
EcoR I
Not I
Blunt PCR Product
T7 priming site
M13 Forward (-20) priming site
Plac
LacZα-ccdB
pUC ori
Kanamycin
Ampicillin
pCR®4Blunt-TOPO®
3956 bp

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
    ccdB 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

C) pPICZαA:


α-factor
Xho I†
Cla I*
Pst I*
EcoR I
Pml I
Sfi I
BsmB I
Asp718 I
Kpn I
Xho I†
Sac II
Not I
Xba I
c-myc epitope
6xHis
Stop
AOX1 TT
BamH I
P TEF1
P EM7
Zeocin
CYC1 TT
pUC ori
Bgl II
5' AOX1
pPICZα
A,B,C
3.6 kb

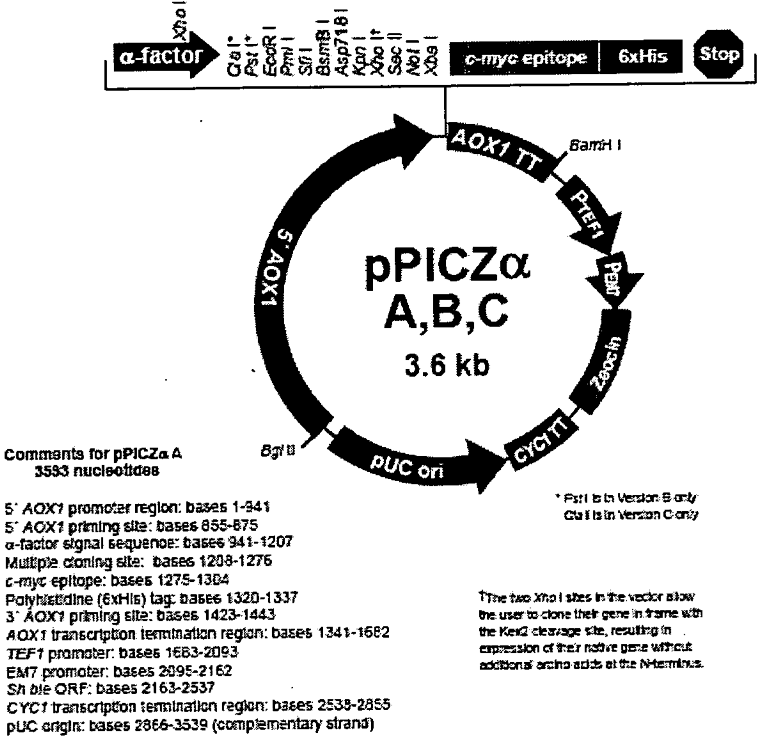
Comments for pPICZα A
3593 nucleotides

5' *AOX1* promoter region: bases 1-941
5' *AOX1* priming site: bases 855-875
α-factor signal sequence: bases 941-1207
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.

D) pPICZαB:


Xho I†
α-factor
Cla I*
Pst I*
EcoR I
Pml I
Sfi I
BsmB I
Asp718 I
Kpn I
Xho I†
Sac II
Not I
Xba I
c-myc epitope
6xHis
Stop
AOX1 TT
BamH I
PTEF1
PEM7
Zeocin
CYC1 TT
pUC ori
Bgl II
5' AOX1
pPICZα
A,B,C
3.6 kb

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
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)

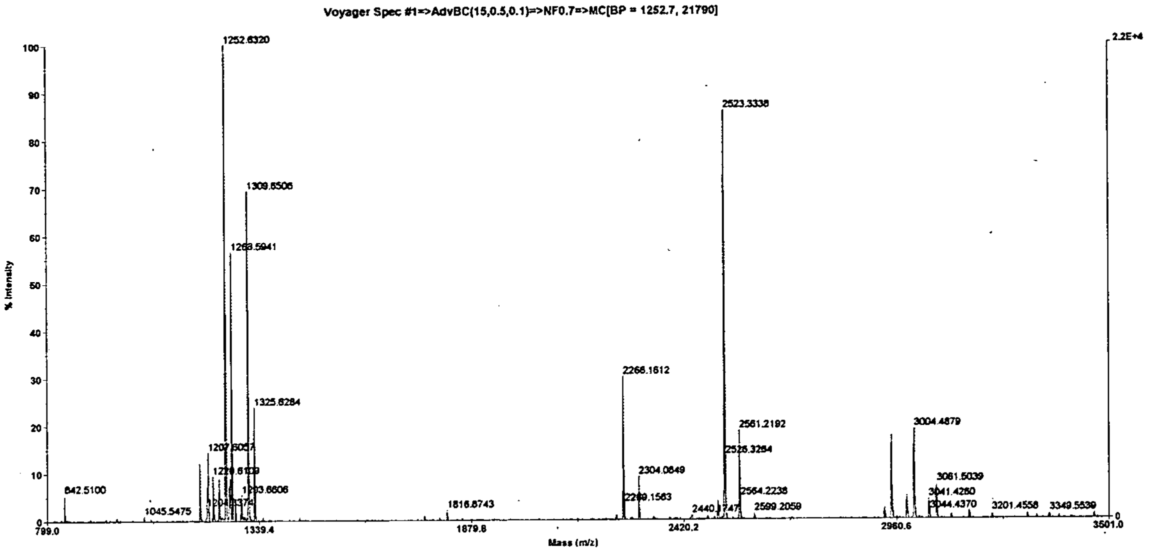
* *Pst* I is in Version B only
Cla I is in Version C only

†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.

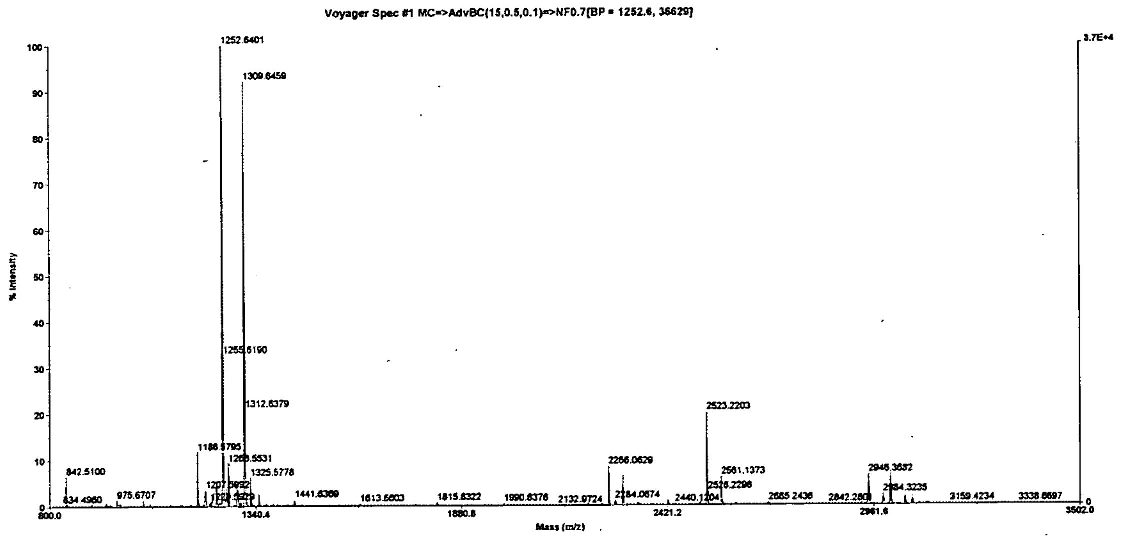
APPENDIX D

MALDI MASS SPECTRA OF TRYPTIC-DIGEST PEPTIDE FINGERPRINTING

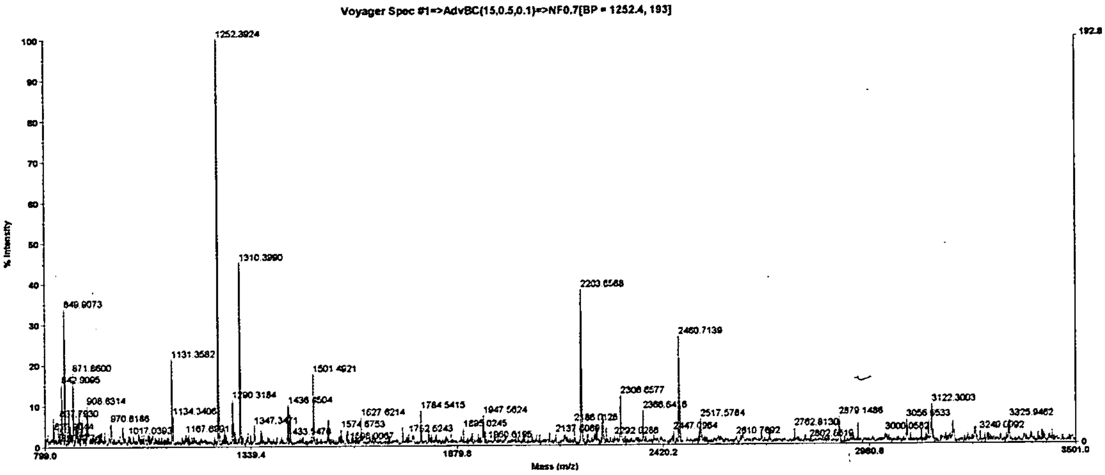
FH-8-9:


Voyager Spec #1=>AdvBC(15,0.5,0.1)=>NF0.7=>MC[BP = 1252.7, 21790]

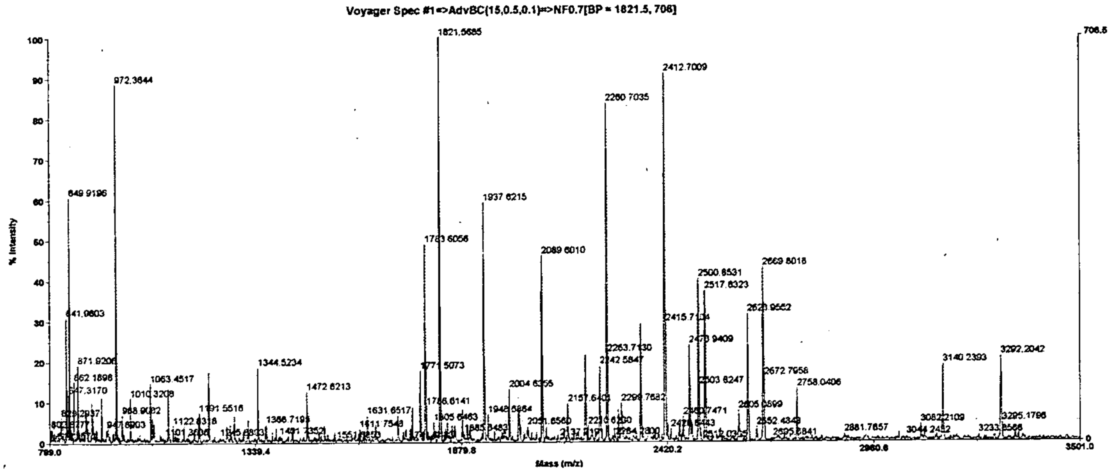
FH-$_{\text{RR}}$8-9:


Voyager Spec #1 MC=>AdvBC(15,0.5,0.1)=>NF0.7[BP = 1252.6, 36629]

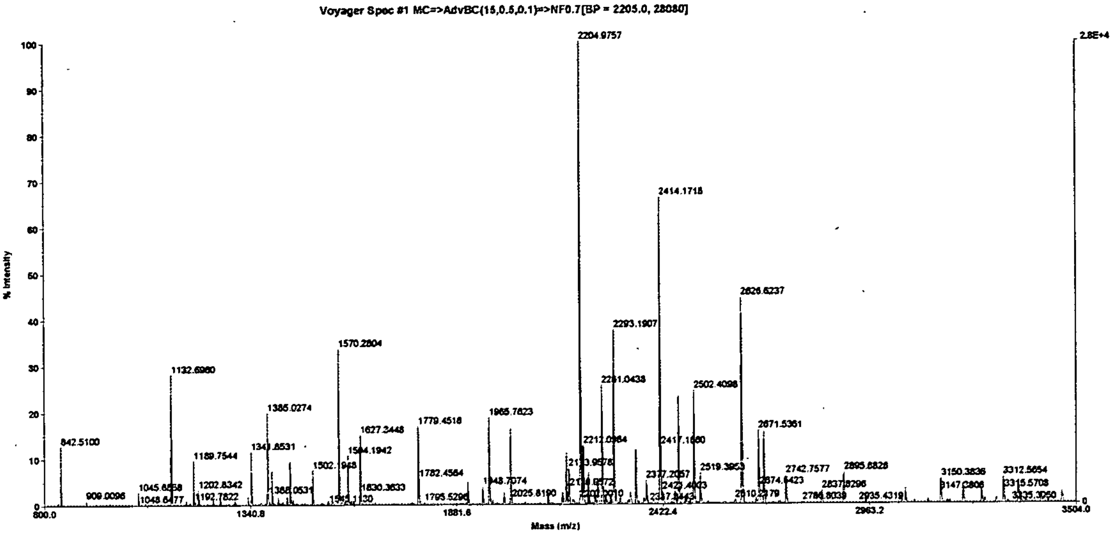
FH-8-15:



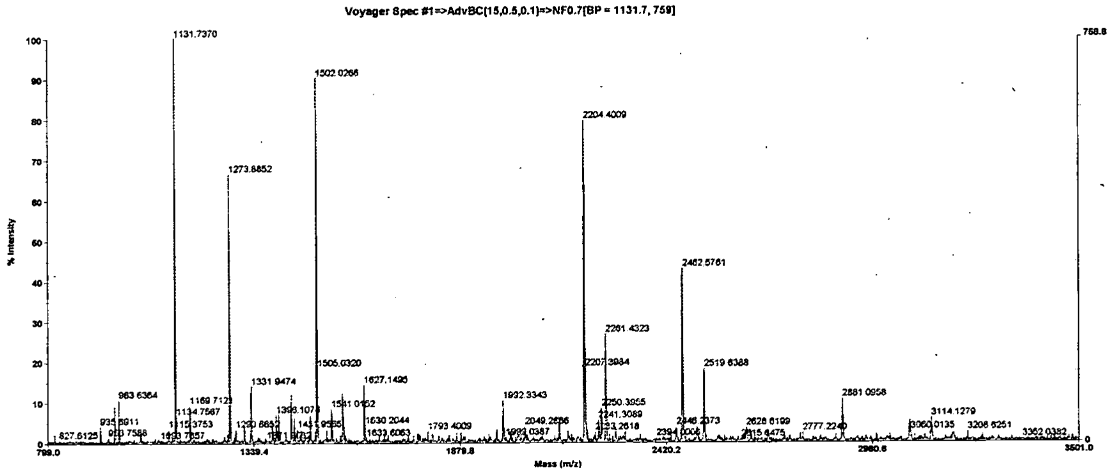
FH-10-12:



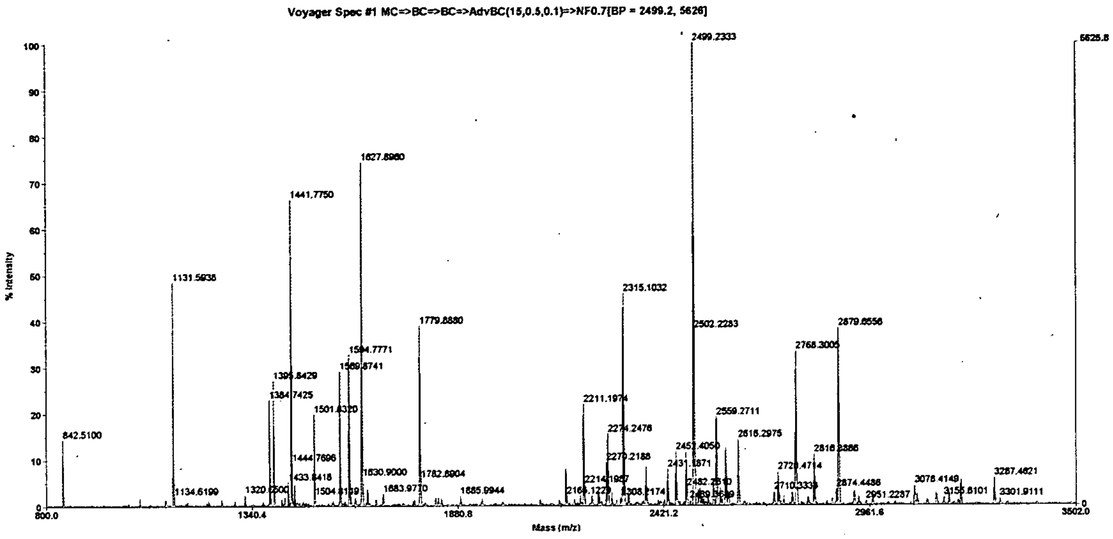
FH-10-15:



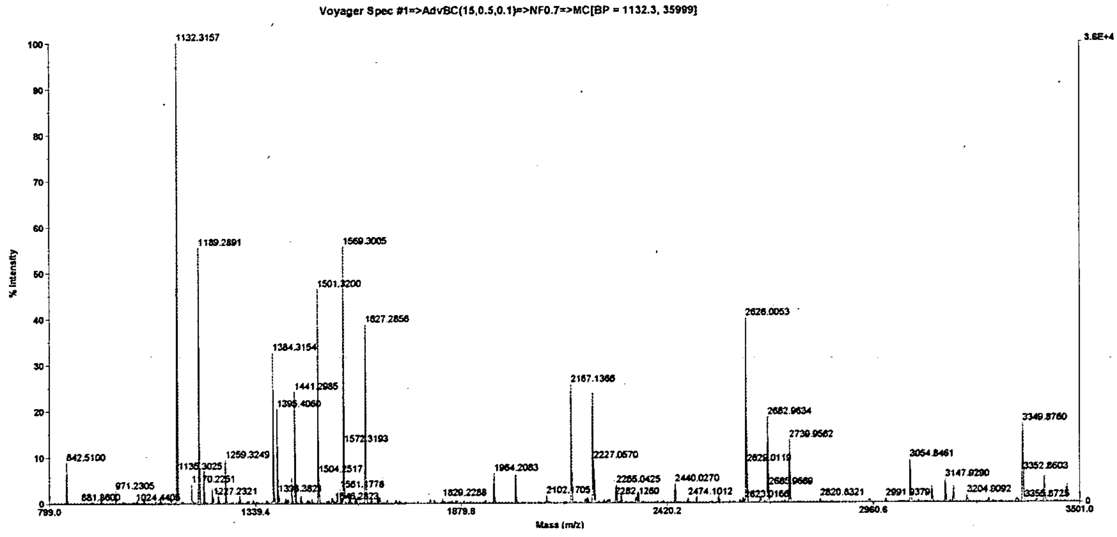
FH11-14:



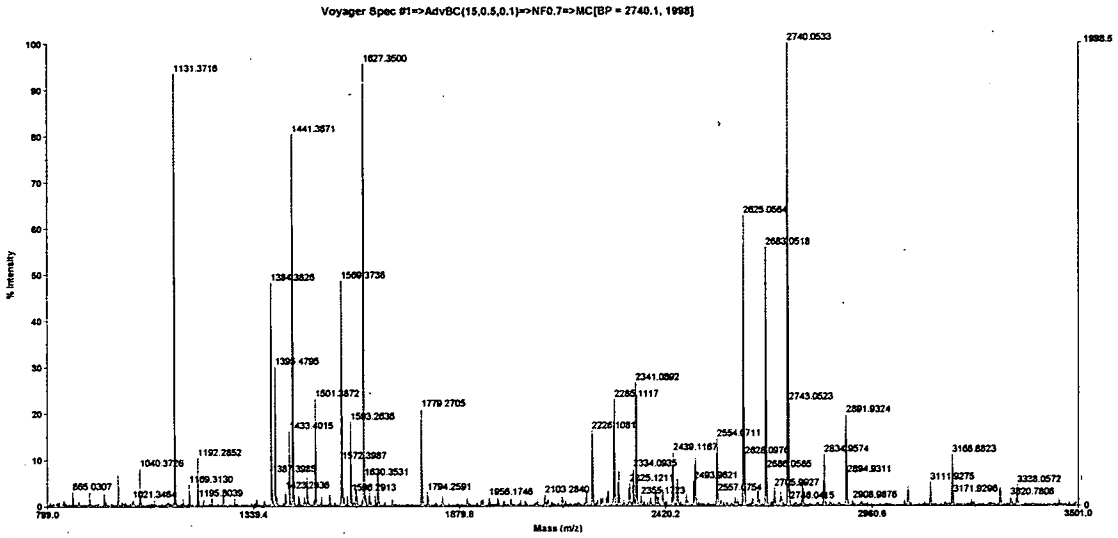
FH12-13:



FH-13-15:

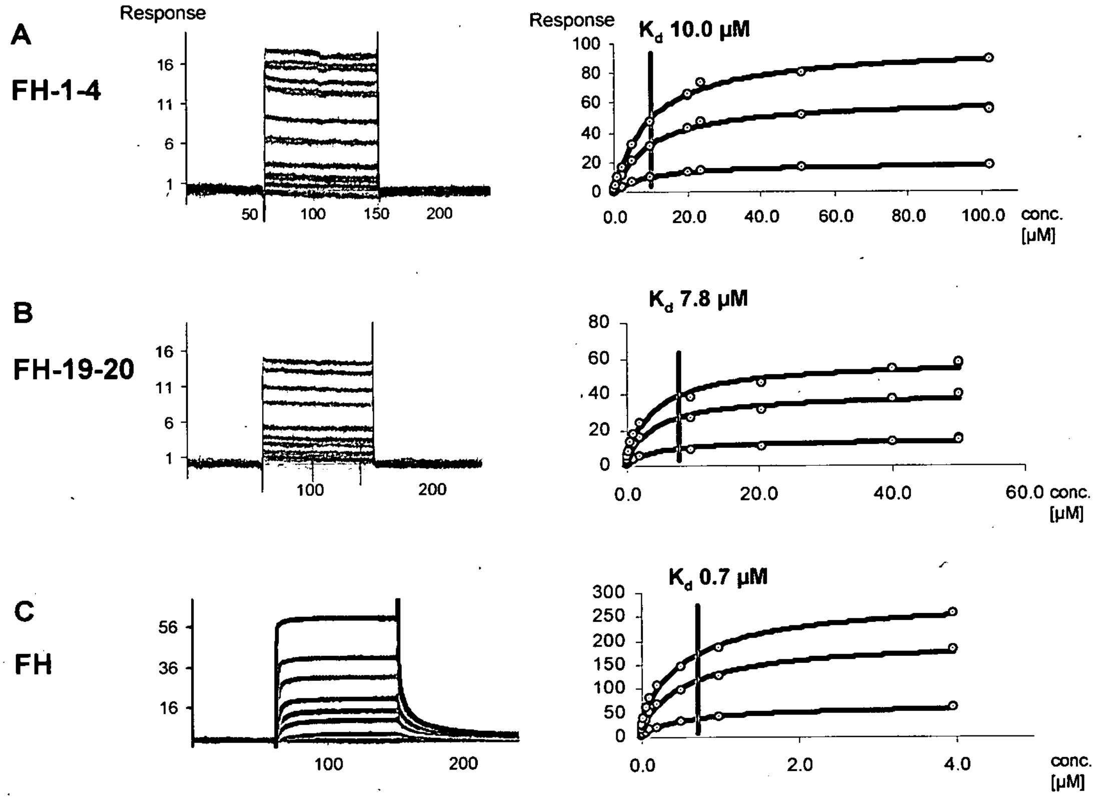


FH-13-14:


Voyager Spec #1=>AdvBC(15,0.5,0.1)=>NF0.7=>MC[BP = 2740.1, 1998]
% Intensity
100
90
80
70
60
50
40
30
20
10
0
1131.3716
1627.3500
1441.3671
1384.3826
1569.3738
1396.4796
1501.3872
1779.2705
1583.2636
1433.4015
1192.2852
1040.3726
1572.3987
1387.3925
1630.3531
1169.3130
865.0307
1423.2936
1586.2913
1794.2591
1021.3464
1195.3039
1956.1746
2103.2840
2740.0533
1998.5
2625.0564
2683.0518
2341.0892
2285.1117
2743.0523
2891.9324
2226.1081
2554.0711
2439.1167
2628.0976
2834.9574
3168.8823
2334.0935
2686.0565
2894.9311
2325.1211
2493.9621
2705.9927
3111.9275
3328.0572
2355.1723
2557.0754
2746.0485
2908.9876
3171.9296
3320.7806
0
799.0
1339.4
1879.8
2420.2
2960.6
3501.0
Mass (m/z)

APPENDIX D

USE OF SPR TO MEASURE DISSOCIATION CONSTANTS FOR MAJOR C3B-BINDING SITES IN FH.



Duplicate sensorgrams are shown for (A) FH-1-4, (B) FH-19-20, and (C) full-length FH at a range of analyte concentrations (FH-1-4, 0.05-102 μM; FH-19-20, 0.05-50 μM; FH, 0.01-3.93 μM). These data are illustrative and show results obtained using flow cell 2 of C1-chip. 140, 500 and 750 RU of C3b were amine coupled onto flow-cells 2,3 and 4, respectively. Flow-cell 1 served as a reference surface.