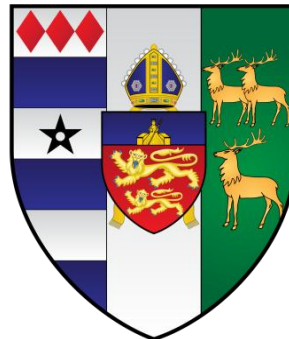
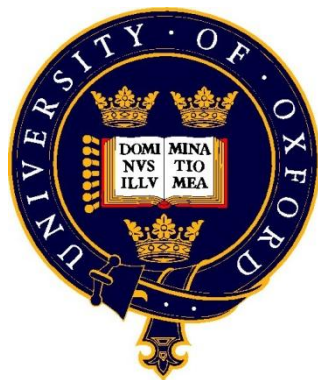


A REASSESSMENT OF THE INTERACTION BETWEEN COMPLEMENT C3d AND COMPLEMENT RECEPTOR CD21 SCR1-2

Thesis submitted to the Board of the Medical Sciences Division in partial
fulfilment of the requirements for the Degree of Doctor of Philosophy at the
University of Oxford



CYNTHIA K. W. TSO

LINCOLN COLLEGE

TRINITY TERM 2012

ABSTRACT

A reassessment of the interaction between complement C3d and complement receptor CD21 SCR1-2

Cynthia K. W. Tso

Department of Biochemistry, Lincoln College

Submitted for the Degree of Doctor of Philosophy, Trinity Term 2012

Biophysical characterisation of protein – ligand interactions can provide vital information to dissect complex biochemical binding mechanisms. C3d has been shown to interact with a number of different protein ligands, namely CD21 SCR1-2, *S. aureus* Efb-C, *S. aureus* Ehp, *S. aureus* Sbi and complement regulatory protein factor H. Although much is known about the relationship of C3d and CD21 in the induction of humoral immunity, the structural aspects of this interaction remained controversial until very recently. The aim of this thesis was to gain a detailed understanding of the C3d/CD21 SCR1-2 interaction using different biophysical methods and to identify potential low molecular weight inhibitors of the interaction.

A crystal structure of the C3d/CD21 complex solved by Szakonyi *et al.* in 2001 indicated the C3d binding site on CD21 was in the SCR2 domain. It did not agree with mutagenesis studies and recent NMR titration experiments show that only residues from the SCR1 domain of CD21 appear to mediate binding under physiologically relevant ionic strength. In the current work, NMR was employed to monitor ligand binding to C3d and to provide residue specific information that reflects a physiologically relevant binding mode to complement the crystallographic model solved by van den Elsen and Isenman in 2011. Microcalorimetric analysis on the site-directed mutagenesis studies also supported a model of hydrophobically- and electrostatically-driven protein-protein interaction for C3d and CD21 SCR1-2.

Complement C3d forms a non-specific thioester linkage with antigen, which then binds to CD21 SCR1-2 and coligates with membrane immunoglobulin of the B cell receptor. While the interactions triggers B cell activation and hence the production of antibody under normal circumstances, it has been demonstrated that the interactions also lead to undue B cell activation and auto-antibody production. There is a well established collagen-induced arthritis (CIA) mouse model to support the significance of C3d and CD21 in disease susceptibility. To this end, a high-throughput SPR-based screening platform was set up to screen a fragment library against C3d, so as to identify low molecular weight compounds that could serve as a starting point for drug development programme. Unfortunately, the work did not yield C3d-binding inhibitors and future work could include screening large fragment libraries that are designed to target protein-protein interfaces.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors, Prof. Jim McDonnell and Prof. Martin Noble, for giving me the opportunity to work on a fascinating structural biology project. The past four years of my D.Phil. have been a wonderful experience and I would like to give special thanks to past and present members of the McDonnell group: Jim (the big boss with an avid interest in collecting sea turtle figures), Norhakim (my absolute favourite lab buddy of all time and I love your eccentricity), Sushmita (our senior that was molten for four years by a molten globule), Johanna (time for another Monmouth coffee to continue our bohemian conversations), Seema (sorry that you have been unwell for a long time and please get well soon), Marie (you're super sweet) and Daniel (emerging star student of the McD lab). The support from you guys is invaluable and I am very grateful to have completed a doctoral degree in such a friendly, fun, motivating and pleasant workplace. Jim: Many many thanks for your guidance over the years and proof reading the thesis.

Oxford is such a marvellous place and I have met several lovely friends in this historic town. Robert: a massive thank you for all your care, encouragement and support throughout. I immensely enjoy living in our comfortable home on Iffley Turn with Blondie (my pet rabbit), exploring exquisite restaurants together in London and going on holidays in Europe. Rebecca, Sook Teng, Cecilia and Jasmine: I am so glad to have a group of amazing friends like you ladies. I cherish every moments we spent together, especially on our incredible skiing trip in Bulgaria and our epic holiday in Scandinavia. Lily: I thought of you every time I walked pass Trinity College. Hope the sky is bright and life is treating you well across the pond. Joseph: fellow biochem lincolnite! Thank you for your amazing hospitality in San Francisco. I for sure will come to California again and look forward to seeing you and Tara under glorious sunshine. James and Winnie: do you realize it has been eight years since we first met at Imperial? I am glad we have evolved from being undergrad classmates to enduring friends. Dawn: thank you for your encouragement and warm-hearted conversations from time to time. I really wish we could have seen each other more often! It has been fourteen years since I left home as a naive thirteen-year-old. Mum and Dad: thank you so much for your unconditional love and giving me the freedom to pursue whatever I have chosen to do. Thank you for everything and hope I have not let you down. This summer is a very special one too – Happy 90th Birthday, my dearest Grandma!

Lastly, I would like to thank Arthritis Research UK for providing a generous studentship to support this project and my participation at a scientific conference in Boston, USA. Also many thanks to the Department of Biochemistry, Oxford University and our collaborators at Randall Division of Cell & Molecular Biophysics, King's College London. This thesis documented my time at Oxford and placed a final mark to the end of my youthful student journey. I look forward to an engaging career in the pharmaceutical industry.



ABBREVIATIONS

ADME	Absorption, distribution, metabolism and excretion
Ag	antigen
BCR	B cell receptor
CMFF	Carboxymethyl fast flow
DMARDs	Disease-modifying anti-rheumatic drug
DSF	Differential scanning fluorimetry
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	Ethylenediamine tetraacetic acid
Efb	Extracellular fibrinogen-binding protein
Ehp	Efb-homologus protein
ESI-MS	Electrospray ionization mass spectroscopy
FDCs	Follicular dendritic cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear single quantum correlation
HTS	High-throughput screening
Ig	Immunoglobulin
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KA	Equilibrium association constant (M^{-1})
KD	Equilibrium dissociation constant (M)
kDa	kilo Dalton
mAb	monoclonal antibodies
MBL	Mannose binding lectin
mM	milli Molar
MS	Mass spectroscopy

MTX	Methotrexate
NSAIDs	Non-steroidal anti-inflammatory drugs
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	NOE spectroscopy
PCR	Polymerase chain reaction
PDB	Protein data bank
PDEA	2-(2-pyridinyldithio)ethaneamine
PPI	Protein-protein interface
ppm	Parts per million
PBS	Phosphate buffer saline
RA	Rheumatoid arthritis
RPM	Rotation per minute
Rmax	Maximum expected response
Sbi	Immunoglobulin binding protein
SCR	Short consensus repeat
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
TROSY	Transverse relaxation optimized spectroscopy
μM	micro Molar
1D	One dimensional
3D	three dimensional

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	ii
ABBREVIATIONS	iv
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xii
CHAPTER I INTRODUCTION	
1.1 The immune system	1
1.1.1 Brief overview of the immune system	1
1.1.2 Overview of the complement system	3
1.1.3 Complement C3d and its link to humoral immunity and autoimmunity	7
1.2 Structural basis of C3d-protein ligand interactions	9
1.2.1 Controversy in the crystal structure of C3d/CD21 SCR1-2 protein complex	9
1.2.2 Crystal structures of C3d in complex with other protein ligands	13
1.3 Therapeutics – The complement system & Arthritis	19
1.3.1 Therapeutic inhibition of the complement system	19
1.3.2 Therapeutic management of rheumatoid arthritis	22
1.4 Aim of thesis	25
CHAPTER II MATERIALS AND METHODS	
2.1 Buffers, solutions and media	27
2.2 Molecular biology	29
2.2.1 C3d DNA and vectors	29
2.2.2 CD21 SCR1-2 DNA and vectors	29
2.2.3 Site-directed mutagenesis	30
2.3 <i>E. coli</i> protein expression and purification	31
2.3.1 Expression and purification of recombinant C3d	31
2.3.2 Expression of ¹⁵ N- and ² H, ¹⁵ N, ¹³ C-labelled recombinant C3d	31

2.3.3 Expression of residue specific ¹⁵ N-leucine labelled recombinant C3d	32
2.3.4 Expression and purification of recombinant CD21 SCR1-2	32
2.4 Protein characterization techniques	33
2.4.1 Protein concentration	33
2.4.2 Mass spectrometry	33
2.4.3 SDS-PAGE	34
2.4.4 One-dimensional ¹ H NMR spectroscopy	34
2.5 Biophysical interaction analysis	34
2.5.1 Differential scanning fluorimetry	34
2.5.2 Surface Plasmon resonance	35
2.5.3 Isothermal titration calorimetry	36
2.6 Nuclear magnetic resonance	36
2.6.1 Sample preparation and data processing	36
2.6.2 NMR experiments for sequential backbone assignment	37
2.6.3 NMR titrations	42
2.7 Crystallization trials of recombinant C3d and the C3d/CD21 SCR1-2 protein complex	42
2.7.1 Crystallization and structure determination of recombinant C3d	42
2.7.2 Crystallization trials of the recombinant C3d/CD21 SCR1-2 protein complex	43
2.7.3 Post crystallization treatment and co-crystallization trials of C3d	43

CHAPTER III BIOPHYSICAL CHARACTERIZATION OF THE C3d/CD21 SCR1-2 INTERACTION AND STRUCTURE-DRIVEN SITE-DIRECTED MUTAGENESIS

3.1 Protein production of recombinant C3d and CD21 SCR1-2	44
3.1.1 Characterization of wild-type unlabelled C3d	44
3.1.2 Characterization of mutated and isotope-labelled C3d	46
3.1.3 Characterization of wild-type and mutated CD21 SCR1-2	47
3.2 SPR analysis on the wild-type C3d and CD21 SCR1-2 interaction	49
3.2.1 Brief introduction to surface plasmon resonance	49
3.2.2 Binding affinity and kinetics of the C3d/CD21 SCR1-2 interaction	50
3.3 ITC analysis on the C3d and CD21 SCR1-2 interaction	52
3.3.1 Brief introduction to isothermal titration calorimetry	52
3.3.2 Thermodynamics of the wild-type C3d/CD21 SCR1-2 interaction	55
3.3.3 Mutation analysis on the C3d/CD21 SCR1-2 interaction	57

3.3.4 Identification of salt bridge interactions by double mutant cycles	62
3.4 Discussion	67
 CHAPTER IV BACKBONE ASSIGNMENT OF C3d BY NMR	
4.1 Introduction to NMR spectroscopy of large proteins	69
4.2 NMR backbone resonance assignment of C3d	72
4.2.1 Sample optimization	72
4.2.2 Triple-resonance NMR experiments for sequential assignment	79
4.2.3 Computer-assisted backbone resonance assignment	82
4.2.4 Specific amino acid labelling to confirm identity of spin systems	89
4.2.5 NOESY experiments in sequential backbone resonance assignment	92
4.3 Discussion	94
 CHAPTER V C3d INTERACTION ANALYSIS BY NMR SPECTROSCOPY	
5.1 Introduction to ligand interaction analysis by NMR	97
5.2 Chemical shift perturbation analysis of C3d	99
5.2.1 The protein-protein interaction between C3d and CD21 SCR1-2	99
5.2.2 The interaction between C3d and factor H-derived peptide	109
5.3 Discussion	119
 CHAPTER VI CRYSTALLIZATION TRIALS OF THE C3d/CD21 SCR1-2 PROTEIN COMPLEX	
6.1 Structure determination of recombinant C3d	122
6.2 Crystallization trials of the protein complex	127
6.2.1 Analytical size exclusion chromatography	127
6.2.2 Crystallization trials of the C3d/CD21 SCR1-2 protein complex	132
6.3 Discussion	133
 CHAPTER VII FRAGMENT-BASED DRUG SCREENING	
7.1 Introduction to fragment-based drug discovery	138
7.1.1 Rational behind fragment-based drug design	138
7.1.2 Prospective small molecules binding groove on the surface of C3d	140
7.2 High-throughput fragment screening techniques	142

7.2.1 Differential scanning fluorimetry	142
7.2.2 Pilot assay with C3d and CD21 SCR1-2	143
7.2.3 DSF-based fragment library screening against C3d	145
7.2.4 Application of surface plasmon resonance in drug discovery	148
7.2.5 SPR-based fragment library screening against C3d	149
7.2.6 SPR-based dose-response experiments	151
7.2.7 DSF vs. SPR as primary fragment library screening methods	153
7.3 Fragment hit validation	156
7.3.1 Protein-based NMR spectroscopy	156
7.3.2 X-ray crystallography	158
7.4 Discussion	159
 CHAPTER VIII CONCLUDING REMARKS AND FUTURE DIRECTION	 162
 BIBLIOGRAPHY	 166
 APPENDIX	 190

LIST OF FIGURES

1.1 The complement activation pathway	6
1.2 Complement fragment C3d acts as a bridge between the innate and acquired immunity	7
1.3 Crystal structure of the C3d/CD21 SCR1-2 protein complex solved by Szakonyi <i>et al.</i> (PDB:1GHQ)	10
1.4 Overlay of the CD21 SCR1-2 ¹ H, ¹⁵ N-HSQC before and after addition of C3d	11
1.5 Crystal structure of C3d in complex with three different complement inhibitory protein from <i>S.aureus</i> ..	15
1.6 Crystal structure of the C3d/factor H SCR19-20 protein complex solved by Morgan <i>et al.</i> (PDB:3OXU)	18
2.1 Site-directed mutagenesis	30
2.2 HNCA	38
2.3 HN(CO)CA	39
2.4 HNCACB	39
2.5 HN(CO)CACB	40

2.6 HNCO	40
2.7 HN(CA)CO	41
3.1 Purification and characterization of wild type unlabelled C3d	45
3.2 1D proton NMR spectra of C3d mutant	46
3.3 Molecular mass profile of ¹⁵ N-labelled C3d obtained by maximum entropy processing of the ESI-MS spectrum	47
3.4 Purification and characterization of unlabelled CD21 SCR1-2	48
3.5 Schematic diagram of an SPR system	49
3.6 Representative SPR sensograms of the wild type C3d/CD21 SCR1-2 interaction	50
3.7 Binding constant determination of wild type C3d/CD21 SCR1-2 interaction	51
3.8 Simulated calorimetric binding curves adapted from Thermodynamics of Protein-Ligand Interactions: History, Presence and Future Aspects (Perozzo R <i>et al.</i> , 2004)	52
3.9 Schematic diagram of an ITC instrument	53
3.10 ITC analysis of the C3d/CD21 SCR1-2 interaction at physiological and high ionic strength	56
3.11 The effect of ionic strength on the C3d/CD21 SCR1-2 interaction	57
3.12 A summary of microcalorimetric analysis on the binding energetic from site-directed mutagenesis studies on C3d and CD21 SCR1-2	61
3.13 Diagram illustrating the double mutant cycle	62
3.14 A docking model of the C3d/CD21 SCR1-2 protein complex generated by ClusPro	66
3.15 A summary of the thermodynamics on enthalpy and entropy data from forty three protein-protein interactions gathered by Stites WE, 1997	68
4.1 ¹ H, ¹⁵ N-HSQC spectra of C3d collected on a Bruker Avance III 500 MHz spectrometer at 25°C	74
4.2 ¹ H, ¹⁵ N-TROSY-HSQC spectra of C3d collected on a Bruker Avance III 700 MHz spectrometer at 25°C	78
4.3 Specimen strip plot of selected spin system from TROSY-based triple resonance experiments	81
4.4 A summary of residues that were unambiguously assigned in the C3d protein sequence	84
4.5 Specimen strip plot to illustrate sequential connectivity of spin systems	85
4.6 Partially assigned ¹ H, ¹⁵ N-TROSY-HSQC spectrum of C3d	86
4.7 Mapping of NMR assigned residues to crystal structure of C3d	88
4.8 Molecular mass profile of ¹⁵ N-leucine labelled C3d obtained by maximum entropy processing of the ESI-MS spectrum	90

4.9 ^1H , ^{15}N -HSQC spectrum of ^{15}N -leucine C3d collected on a Bruker Avance III 700 MHz spectrometer at 25°C	91
4.10 Specimen strip plot of the ^1H , ^{15}N -NOESY spectrum of C3d collected on a Bruker Avance III 700 MHz spectrometer at 25°C	93
5.1 ^1H , ^{15}N -chemical shift perturbation experiments to map the interaction surfaces on C3d for CD21 SCR1-2	102
5.2 Insert showing expanded views on the indicated areas in the ^1H , ^{15}N -chemical shift perturbation experiments	103
5.3 The binding surfaces for CD21 SCR1-2 on C3d defined by ^1H , ^{15}N -chemical shift perturbation analysis	108
5.4 Rationale behind the design of a potential C3d-binding factor H-derived peptide	110
5.5 SPR analysis of the interaction between C3d and factor H-derived peptide	111
5.6 ITC analysis of the interaction between C3d and factor H-derived peptide	113
5.7 ^1H , ^{15}N -chemical shift perturbation experiments to map the interaction surfaces on C3d for factor H-derived peptide	115
5.8 Insert showing expanded views on the indicated areas in the ^1H , ^{15}N -chemical shift perturbation experiments	116
5.9 Saturation binding curves based on change in proton chemical shift position	117
5.10 The binding surfaces for factor H-derived peptide on C3d defined by ^1H , ^{15}N -chemical shift perturbation analysis	119
6.1 Intramolecular thioester bond in complement C3d	124
6.2 Crystal structure of C3d in P6 ₁ and P2 ₁ 2 ₁ 2 ₁ spacegroups	126
6.3 Analytical size exclusion chromatography profile of the molecular weight standards, C3d and the C3d/CD21 SCR1-2 complex	128
6.4 Size exclusion chromatography calibration curve for proteins with molecular weight between 30 kDa to 66 kDa	129
6.5 SDS-polyacrylamide gel showing fractions of the single peak eluted from a C3d/CD21 SCR1-2 mixture	129
6.6 The Influence of ionic strength on the formation of the C3d/CD21 SCR1-2 complex	131
6.7 Crystal structure of the C3d/CD21 SCR1-2 protein complex solved by Isenmen DE <i>et al.</i> (PDB:3OED)	135
6.8 Comparison of the C3d/CD21 SCR1-2 complex predicted by ClusPro and solved by X-ray crystallography	137
7.1 Prospective small molecule binding groove on C3d	141
7.2 Pilot DSF assay with C3d and CD21 SCR1-2	144

7.3 Representative denaturing curves showing the influence of fragments on the thermal stability of C3d	146
7.4 Representative SPR sensorgrams showing fragments that appeared to bind promiscuously to C3d	150
7.5 Representative SPR sensorgrams showing fragments with decent binding characteristics to C3d	151
7.6 Representative SPR sensorgrams showing fragment binding promiscuously to C3d	153
7.7 DSF and SPR as primary fragment library screening methods for hit detection	154
7.8 Results of DSF-based and SPR-based fragment library screening were sometimes inconsistent	155
7.9 Representative HSQC spectra of C3d collected on a Bruker 700 MHz spectrometer before and after the addition of potential fragment hits	157
7.10 Current hit detection methods and their sensitivity limit for detecting small molecule binding proteins	160

LIST OF TABLES

1.1 Summary of distinctive features of innate and acquired immunity	2
1.2 Comparison of C3d binding data from Efb-C, Ehp and Sbi-IV using ITC	14
2.1 Summary of samples used in NMR experiments	37
2.2 NMR titration parameters on ¹⁵ N-labelled C3d samples	42
2.3 Summary of experimental parameters used in analytical size exclusion chromatography	43
3.1 The effect of ionic strength on the C3d/CD21 SCR1-2 interaction	57
3.2 Affinities and thermodynamic parameters of C3d and CD21 SCR1-2 mutants	59
3.3 A summary of double mutant cycle analyses	65
4.1 Molecular weight distribution of solution NMR structures deposited in the RCSB Protein Data Bank as of January 2012	70
4.2 Completeness of the NMR spectra	80
5.1 Summary of NMR kinetics	98
5.2 NMR chemical shift perturbation analysis of the C3d/CD21 SCR1-2 interaction	101
5.3 Thermodynamic parameters of the interaction between C3d/CD21 SCR1-2 and C3d/factor H-derived peptide in 20 mM NaH ₂ PO ₄ , 150 mM NaCl, pH 6.8	113
6.1 Summary of data collection and refinement statistic	125
6.2 Crystallization trial of the C3d/CD21 SCR1-2 protein complex	132

CHAPTER I

INTRODUCTION

1.1 THE IMMUNE SYSTEM

1.1.1 BRIEF OVERVIEW OF THE IMMUNE SYSTEM

Immunity protects host organisms against pathogens and can be broadly categorized into innate immunity and acquired immunity (Table 1.1). Innate immunity relies on germline-encoded pattern-recognition receptors (PRRs) that recognize evolutionary conserved pathogen-associated molecular patterns (PAMPS) on the surface of microorganisms, such as lipopolysaccharides, peptidoglycans, non-methylated CpG and double-stranded RNA (Janeway CA *et al.*, 2002; Beutler B, 2003; Beutler B, 2004; Raulet DH, 2004). These patterns are absent from eukaryotic cells and allow discrimination between pathogens and self (Medzhitov R *et al.*, 1997; Hoffmann JA *et al.*, 1999). The Toll-like receptor (TLR) family is a major class of mammalian PRR for detecting

pathogenic non-self molecules (Medzhitov R *et al.*, 1997). Akin to the TLR family, the complement system is considered a central component of innate immunity (Hosea SW *et al.*, 1980) and can be activated by PRRs. The recognition of pathogenic components causes induction of genes encoding cationic antimicrobial peptides which trigger host defense by damaging the cell membranes of pathogens (Shai Y, 1999; Lehrer RI and Ganz T, 1999). Furthermore, the activation of the innate immune response is typically a prerequisite for the initiation of acquired immunity (Fearon DT and Locksley RM, 1996; Schnare M *et al.*, 2001).

In contrast to the innate immune system, acquired immunity is mediated by clonally distributed B and T lymphocytes that differentiate from stem-cell precursors within primary haematopoietic tissues and the thymus, respectively (Pancer Z *et al.*, 2004). The acquired immune system utilizes a vast repertoire of receptors, such as immunoglobulins (Ig) and T-cell receptors (TCR), which are generated through recombination of variable gene segments in the Ig and TCR gene loci. Ig-bearing B lymphocytes recognize unprocessed antigens and TCR-bearing T lymphocytes recognize antigen fragments presented by major histocompatibility complex (MHC) class I and II molecules (Rudolph MG and Wilson IA, 2002). Together these receptors presented a huge diversity in the recognition of any possible antigen and each receptor of an unique specificity is expressed in a clone of lymphocytes. Some B and T lymphocytes will become memory cells after activation and allow the acquired immune system to mount a quicker and more vigorous response against subsequent exposures to previously encountered antigens.

Table 1.1 Summary of distinctive features of innate and acquired immunity		
	Innate immunity	Acquired immunity
Receptors	Germ line-encoded	Products of site-specific somatic recombination
Receptor expression	Restricted to certain cell-types but non-clonal	Clonally-distributed via B and T lymphocytes
Repertoire	Limited	Vast
Memory	No	Yes

1.1.2 OVERVIEW OF THE COMPLEMENT SYSTEM

The complement system provides an important link between the innate and acquired immunity, most notably through the activated complement C3 components (Fearon DT and Carter RH, 1995; Carroll MC, 2004). These C3 components have been shown to modulate humoral immune response in acquired immunity (Carroll MC, 1998; Dempsey *et al.*, 1996), as well as innate immune response through interactions with complement regulatory proteins (Longhi MP *et al.*, 2006). The complement system is comprised of more than thirty soluble serum proteins and activation of the complement system leads to rapid generation of various complement cleavage products, which recruit and activate effector immune cells to promote the lysis or phagocytosis of invading pathogens (Müller-Eberhard HJ, 1988). The complement system can be activated by three distinct pathways, namely the classical pathway (Reid KBM and Porter RR, 1981), lectin pathway (Reid K *et al.*, 1994; Fujita T *et al.*, 2004) and alternative pathway (Fearon DT and Austen KF, 1975). The pathways differ in the recognition mechanism of antigens but their activation steps converge in the cleavage of central complement component C3 protein (Figure 1.1). The activation of the complement system comprises a cascade of initiation and amplification steps which are regulated at multiple levels (Parker CJ, 1992; Liszewski MK and Atkinson JP, 1996). The regulation facilitates recognition of self from non-self antigens and modulates potent cell-destruction capability of the complement system to cellular targets.

(i) Classical pathway

The classical pathway is initiated when complement component C1 binds to antigen-bound immunoglobulin or C-reactive protein (Szalai AJ *et al.*, 1997) on the surface of foreign or apoptotic cells. Antigens expressed on cell surfaces are tagged by circulating natural antibody (*e.g.* IgM) or antibody released after a humoral response (*e.g.* IgG). C-reactive protein is a pattern-recognition receptor that binds to phosphocholine expressed on the surface of apoptotic cells and certain types of bacteria. The complement component C1 constitutes a receptor C1q (Korb LC and Ahearn LM, 1997; Taylor PR *et al.*, 2000) and proteases C1r and C1s, which are consecutively activated upon cell surface recognition of C1q (Liszewski MK *et al.*, 1996). C1q recognizes antigen-bound IgG/IgM and C-reactive protein on cellular surface and induces the formation of C3

convertase through cleavage of component C2 and C4 by protease C1s. Cleavage of C4 and C2 leads to covalent deposition of C4b on cell surfaces in the immediate vicinity of activated antigen-bound C1q clusters (opsonization), as well as generation of C3 convertase that cleaves C3 to initiate downstream immune responses. C1 esterase inhibitor (C1-inh) is a complement regulatory protein that controls this activation step. It irreversibly binds to C1r and C1s after their dissociation from C1q, thereby preventing spontaneous cleavage of subsequent complement components.

(ii) Lectin pathway

The lectin pathway utilizes pattern-recognition receptors that predominantly recognize carbohydrate patterns expressed on the surface of pathogens, such as mannose-binding lectin (Epstein J *et al.*, 1996) and ficolins (Matsushita M and Fujita T, 2002). Mannose-binding lectin (MBL) is a C-type lectin which is structurally similar to complement component C1q. It has six globular carbohydrate-recognition domains that bind to N-acetyl glucosamine- and mannose-bearing glycoconjugates on the surface of pathogens. Ficolins have similar functional properties to lectins and bind to acetyl groups in both carbohydrates and non-carbohydrate cell surface antigens. Both MBL and ficolins assemble with MBL-associated serine proteases (MASPs) that are structurally and functionally similar to complement C1r and C1s in the classical pathway, leading to the generation of C3 convertase (Fujita T *et al.*, 2004; Stahl PD and Ezekowitz RA, 1998). Small MBL-associated protein (sMAP) and MAP-1 are the complement regulatory proteins that control this activation step by binding to MBL and ficolins to compete with MASPs and to inhibit C4 deposition.

(iii) Alternative pathway

The alternative pathway is continuously switched-on to provide a low background level of complement protection and allows constant probing of cells (Fearon DT and Austen KF, 1975; Daha MR *et al.*, 1976). Minimal spontaneous hydrolysis of complement component C3 and allows recruitment of factor B proteases to C3. The transient complex is cleaved by factor D to generate C3 convertase that subsequently cleaves complement C3 into its active fragment C3a and C3b. The complement component C3 contains a thioester

bond formed between side chains of a specific glutamine and cysteine residues that becomes highly reactive upon cleavage into C3b (Bhagavan NV, 2002). This reactive transient thioester group on C3b covalently attaches to nucleophilic antigens presented on cell surfaces, such as amine or hydroxyl groups on glycoconjugates.

Regardless of the complement pathways, all surface-bound C3 convertases can induce the amplification branch of the alternative pathway by cleaving complement C3. The amplification process creates an efficient cycle of C3 activation and convertase assembly in the presence of factor B and factor D, thereby increasing the responses from crucial effector branches of the complement system. Innate defense mechanisms triggered by the complement system include recruitment of inflammatory cells by anaphylatoxins (e.g. complement C3a and C5a), opsonization of pathogens by C3b and C3d, as well as lytic attack of pathogens by the formation of membrane attack complex. Following the cleavage of C3 into C3a and C3b, the C3 convertase recruits one C3b fragment to form C5 convertase. Reactions downstream of C5 convertase generate precursors of the membrane attack complex (Müller-Eberhard HJ, 1985) to form lytic pore on surfaces of target cells. The C3 and C5 convertases are regulated by complement receptor 1 (CD35), decay-accelerating factor (DAF), membrane co-factor protein (MCP) and factor H. Factor H is the most abundant regulator and acts by accelerating decay of C3 convertases or by promoting the degradation of C3b (Whaley K and Ruddy S, 1976; Weiler JM *et al.*, 1976). Complement fragment C3d is the final degradation product of the antigen-bound C3b (Seya T and Nagasawa S, 1985). The co-recognition of antigen-bound C3d by complement receptor CD21 and B-cell receptor (BCR) on surfaces of B lymphocytes and follicular dendritic cells (FDCs) has been shown to form an important link between the innate and acquired immunity (Dempsey PW *et al.*, 1996; Toapanta FR and Ross TM, 2006).

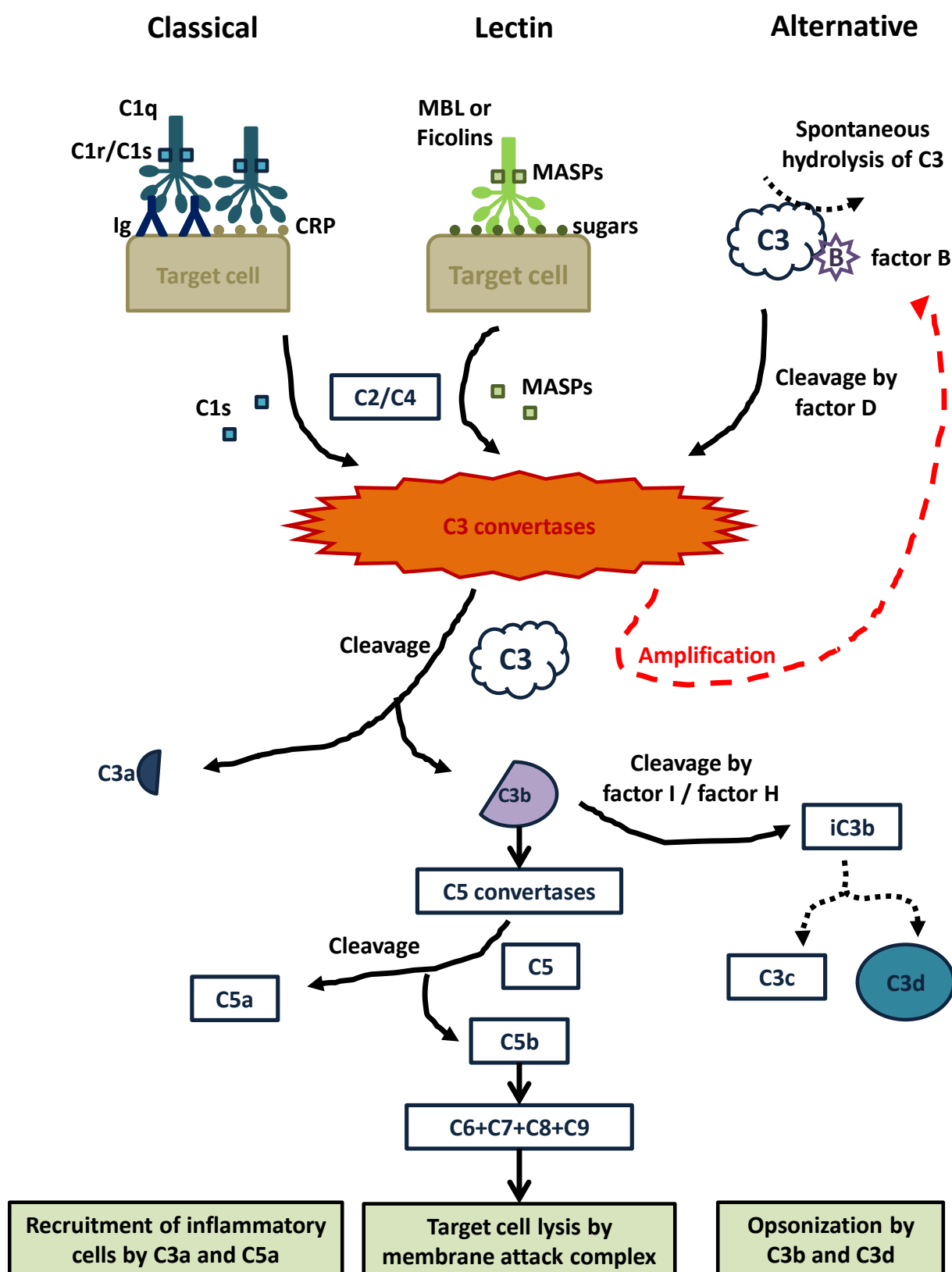


Figure 1.1] The complement activation pathways

The complement system can be activated by the classical, lectin and alternative pathways and they all converge to a central step when complement component C3 is cleaved by C3 convertase.

1.1.3 COMPLEMENT C3d AND ITS LINK TO HUMORAL IMMUNITY AND AUTOIMMUNITY

Although the complement system is considered a chief innate host defense system, its involvement in acquired immunity was demonstrated by early experimental results showing that a transient reduction in circulating complement component C3 led to an impairment in antibody response (Pepys MB, 1976). Complement enhances humoral immunity predominantly through complement receptor CD21 and CD35 expressed on B lymphocytes and follicular dendritic cells (FDCs) (Ahearn JM *et al.*, 1996; Molina H *et al.*, 1996). CD21 forms a receptor complex with a signaling molecule CD19 and binds to C3d (Matsumoto AK *et al.*, 1991; Barrington RA *et al.*, 2009). When circulating naïve B lymphocytes encounter antigen and receive additional signal from T-helper lymphocytes, they become activated and differentiate into antibody-secreting plasma B lymphocytes or memory B lymphocytes. The co-engagement of CD21 and the B cell antigen receptor (BCR) mediated by antigen-bound C3d enhances B lymphocyte activation and provides signal for cell survival (Carter RH and Fearon DT, 1992). It has been demonstrated that antigen-bound C3d lowered the threshold for B lymphocyte activation 10,000-fold more efficiently than antigen alone (Dempsey PW *et al.*, 1996). Hence, C3d acts during the inductive phase of the antibody response through its covalent attachment to antigen and co-engagement of CD21 and the BCR on the surface of B lymphocytes (Figure 1.2). Subsequent B lymphocyte differentiation leads to antibody production that confers long-lasting acquired immunity against specific antigens recognized by the BCR (Carroll MC, 1998).

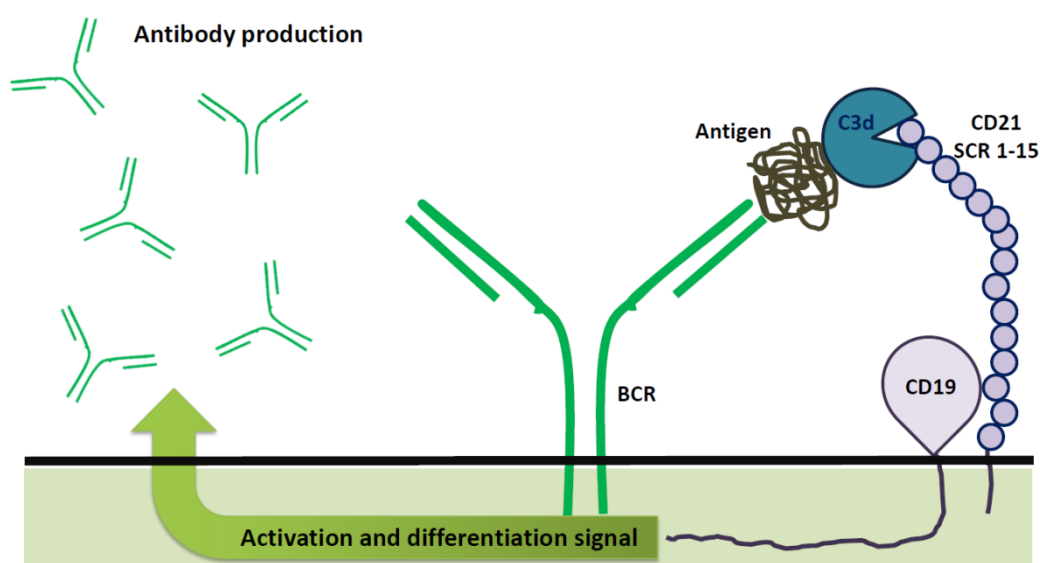


Figure 1.2 | Complement fragment C3d acts as a bridge between the innate and acquired immunity

However, given the non-specific nature of the thioester covalent attachment of antigen to C3d, it is important to note that both foreign and self antigens are equally capable of being opsonized by C3d. Studies in mice with targeted deficiencies of complement C3 (Fischer MB *et al.*, 1996), CD21 and CD35 receptors (Ahearn J *et al.*, 1996; Molina H *et al.*, 1996) have shown the involvement of complement in multiple stages of B lymphocyte differentiation. Failed clearance of immune complexes and apoptotic cells are thought to increase the likelihood of exposure and reactivity to self-antigens. Consequently, undue activation of the complement system is implicated in numerous pathological conditions, in particular autoimmune disorders such as rheumatoid arthritis (Gulati P *et al.*, 1994; Morgan BP *et al.*, 1988). Using a classical murine model of self-antigen collagen type II-induced inflammatory arthritis (CIA), it has been reported that attachment of self-antigen to C3d can promote joint inflammation and auto-destruction in the presence of the CD21-CD19 co-receptor on B lymphocytes or FDCs (Hietala MA *et al.*, 2002; Del Nagro CJ *et al.*, 2005). In a more recent study, it has been demonstrated that self-antigen-bound C3d can induce autoantibody production in a murine model of B lymphocyte anergy (Lyubchenko T *et al.*, 2007). B lymphocyte anergy is an inhibitory mechanism in which defective B lymphocytes displaying autoreactive BCR were deactivated and undergo clonal deletion upon self-antigen stimulation. Autoreactive B lymphocytes account for about 20% of the mature B lymphocyte population in healthy individuals and up to 50% in rheumatoid arthritis patients (Samuels J *et al.*, 2005). Co-engagement of CD21 and the BCR on autoreactive B lymphocytes mediated by self-antigen-bound C3d can breach the inhibitory mechanism and leads to auto-antibody production. By blocking the association of C3d and CD21, C3d-based inhibitors may dampen the complement-induced humoral response associated with autoimmune diseases such as rheumatoid arthritis.

1.2 STRUCTURAL BASIS OF C3d–PROTEIN LIGAND INTERACTIONS

1.2.1 CONTROVERSY IN THE CRYSTAL STRUCTURE OF C3d-CD21 SCR1-2 PROTEIN COMPLEX

Although much is known about the relationship of C3d and CD21 in the induction of humoral immunity, the structural aspects of this interaction remain controversial. CD21 is a type I transmembrane protein with a molecular weight of approximately 145 kDa. There are two polymorphic forms of CD21 in which the extracellular domain comprises fifteen or sixteen short consensus repeats (SCRs) (Weis JJ *et al.*, 1988). Each SCR domain comprises 60 - 70 amino acid residues, which is folded in a compact β -barrel configuration containing four conserved cysteine residues and an invariant tryptophan. The conserved cysteine residues form a pattern of disulfide bridges connecting Cys1-Cys3 and Cys2-Cys4. It has been demonstrated that the two N-terminal domains of CD21, *i.e.* SCR1 - SCR2, were sufficient to mediate high affinity interaction with C3d (Molina H *et al.*, 1995).

A crystal structure of C3d-CD21 SCR1-2 protein complex determined in 2001 (PDB:1GHQ) indicated that the C3d binding site on CD21 was in the SCR2 domain (Szakonyi G *et al.*, 2001). In the crystal structure, CD21 SCR1-2 formed a non-physiological homodimer with only one SCR2 domain in direct contact with C3d (Figure 1.3). Previous X-ray scattering and analytical ultracentrifugation studies have unambiguously shown that CD21 SCR1-2 existed as monomer whether bound to C3d or free in solution (Guthridge JM *et al.*, 2001; Gilbert HE *et al.*, 2005). The interactions between C3d and CD21 SCR2 in the crystal structure involved an extensive use of backbone atoms on both proteins in forming hydrogen bonds and van der Waals interactions mediated by water molecules at the binding interface. The backbone interactions observed in the crystal structure do not agree with the experimental and computational analysis supporting electrostatically driven and ionic strength dependent characteristics of the C3d-CD21 SCR1-2 interaction (Clemenza L *et al.*, 2000; Zhang L *et al.*, 2007; Bomb M, 2008; Guthridge JM *et al.*, 2001).

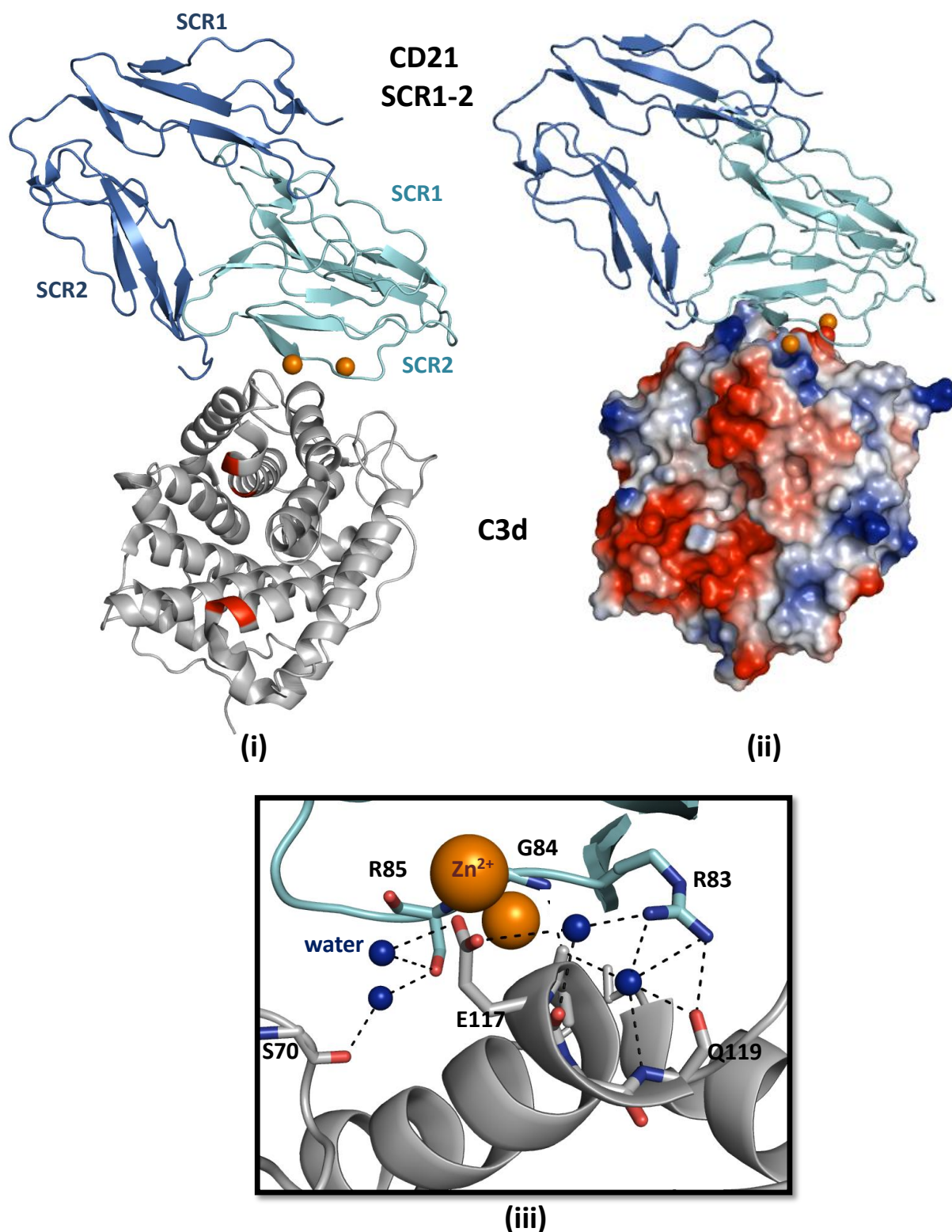


Figure 1.3 | Crystal structure of the C3d-CD21 SCR1-2 protein complex solved by Szakonyi G *et al.* (PDB: 1GHQ)

- (i) Structure of the C3d-CD21 SCR1-2 protein complex showing overall molecular architecture as cartoon diagram. CD21 SCR1-2 forms a non-physiological homodimer (light blue and dark blue) and two zinc ions are found at the C3d-CD21 SCR1-2 binding interface (orange spheres). Mutagenesis studies indicated residues colored in red in C3d were shown to be key residues responsible for binding to CD21 SCR1-2.
- (ii) Structure of the protein complex with C3d shown as electrostatic surface representation.
- (iii) Close-up diagram of the crystallographic C3d-CD21 SCR2 binding interface.

Furthermore, residues that were shown to mediate protein complex formation in the crystal structure were inconsistent with site-directed mutagenesis studies (Clemenza L and Isenman DE, 2000; Hannan JP *et al.*, 2005; Bomb M, 2008). Mutational studies indicated residues that mediated high-affinity binding to C3d were located on CD21 SCR1. Also, mutating residues at the crystallographic interface on both proteins had no major effect towards binding interaction. C3d possesses a negatively charged concave surface, and likewise there is a positively charged convex surface on CD21 SCR1 (Figure 1.4). Neutralizing one or more of charged residues in the respective region on C3d or CD21 has been shown to diminish their association significantly. Previous CD21 SCR1-2 NMR backbone assignment and subsequent titration experiments with C3d from our group supported the findings from mutagenesis studies (Hannan JP *et al.*, 2005; Bomb M, 2008). Employing NMR chemical shift perturbation experiments, unlabeled C3d was titrated against ^{15}N -labelled CD21 SCR1-2 (Bomb M, 2008). Only a small number of residues on CD21 SCR1-2 were affected by addition of C3d (Figure 1.4). At physiologically relevant ionic strength (150 mM NaCl), only residues from CD21 SCR1 were perturbed and these residues formed a single contiguous surface when mapped onto the structure of CD21 SCR1-2. This binding region consists of positively charged residues and a central hydrophobic patch lying along the convex surface of the SCR1 domain. The result was consistent with early mutagenesis studies (Hannan JP *et al.*, 2005; Bomb M, 2008), but not was not in full agreement with a recent NMR analysis published elsewhere (Kovacs JM *et al.*, 2009).

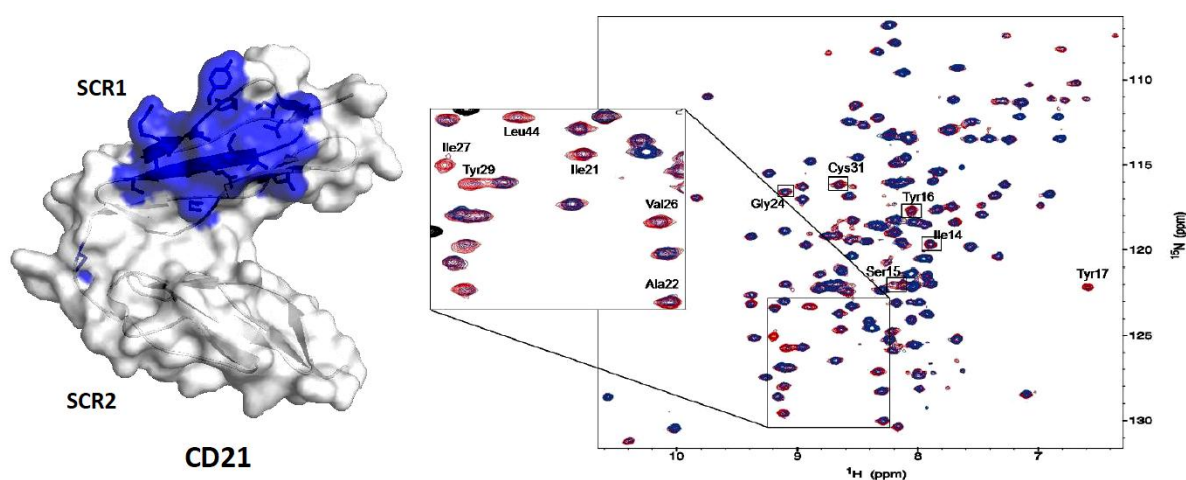


Figure 1.4 | Overlay of the CD21 SCR1-2 ^1H , ^{15}N HSQC before and after the addition of C3d

The spectra were collected on a Bruker 500 MHz spectrometer at 25 °C with 100 μM CD21 SCR1-2 in 20 mM NaHPO_4 , 150 mM NaCl, pH6.8 (Bomb M, 2008). Red = starting spectra; Blue = at 1:1 stoichiometry of CD21 SCR1-2:C3d. Residues with the most marked changes were indicated. When these residues were mapped on the CD21 SCR1-2 structure, they formed a large contiguous binding surface for C3d in the SCR1 domain.

NMR chemical shift perturbation analysis published by Kovacs *et al.* suggested that residues from both CD21 SCR1 and SCR2 were involved in the binding to C3d in 1/3 PBS solution (*i.e.* ~50 mM NaCl). The authors proposed a high-affinity and a low-affinity C3d binding site on CD21 SCR1 and SCR2, respectively. Our NMR data at low ionic strength also showed residues perturbed outside of CD21 SCR1 domain; these residues did not form a contiguous binding site and were scattered over both domains of CD21. The scattered additional sites most likely represent non-specific binding between two oppositely charged molecules at low ionic strength. Analytical ultracentrifugation studies have shown that C3d existed in monomer-dimer and monomer-trimer equilibria in 50 mM NaCl, but as a monomer only in 137 mM NaCl (Li K *et al.*, 2008). It is possible the low-affinity C3d binding site on CD21 SCR2 proposed by Kovacs *et al.* is an artefact of non-specific association of multimeric C3d to CD21 SCR1-2 at low ionic strength. In the NMR studies carried out by Bomb M, the non-specific binding sites disappeared when the ionic strength was raised to 150 mM NaCl and only residues in CD21 SCR1 appeared to mediate binding to C3d under physiological relevant ionic strength. With respect to the crystal structure, Bomb's data also indicated that residues observed at the crystallographic interface were unperturbed in the NMR chemical shift perturbation analysis.

1.2.2 CRYSTAL STRUCTURES OF C3d IN COMPLEX WITH OTHER PROTEIN LIGANDS

In addition to CD21 SCR1-2, C3d also binds to several other proteins for which crystal structures of the complex were available. Native protein ligands of C3d include *Staphylococcus aureus* extracellular fibrinogen-binding protein Efb-C (Hammel M *et al.*, 2007), Efb-homologous protein Ehp (Hammel M *et al.*, 2007) and immunoglobulin-binding protein Sbi (Clark EA *et al.*, 2011), as well as host complement regulatory protein factor H (Morgan HP *et al.*, 2011). Interestingly, the CD21 SCR1-2 binding site on C3d confirmed by mutational studies shows a high degree of adaptability and is also utilized by these native C3d ligands. An understanding of their interaction sites could provide insights into the development of inhibitors targeting complement fragment C3d.

(i) Complement inhibitory proteins from *S. aureus*

S. aureus triggers all three activation pathways of the complement system (Neth O *et al.*, 2002; Kawasaki A *et al.*, 1897; Bredius RG *et al.*, 1992; Verbrugh HA *et al.*, 1979; Wilkinson BJ *et al.*, 1978) and the bacterium secretes complement inhibitory proteins that mimic the function of host complement regulatory proteins to attenuate complement response (Lee LY *et al.*, 2004). Currently six intrinsic complement inhibitory proteins secreted by *S. aureus* have been identified and three of which are ligands of complement C3 and its fragment C3d. The C-terminal domain of extracellular fibrinogen-binding protein (Efb-C) is a 16 kDa protein secreted by *S. aureus* to inhibit the deposition of C3b/C3d onto sensitized surfaces (Lee LY *et al.*, 2004). It has been shown that *S. aureus* strains incapable of expressing Efb are less virulent than a wild-type strain in a mouse model of wound healing (Palma M *et al.*, 1996). The bacterium also secretes Ehp, a homologue of Efb-C, which consists of eighty amino acid residues and shares 44% sequence identity with Efb-C. Findings in literature suggest the binding of either protein induces a conformational change in complement component C3 so that C3 is incapable of participating in the generation of C3b and C3 convertase (Hammel M *et al.*, 2007). Immunoglobulin-binding protein (Sbi) is the most recent characterized complement inhibitory protein and affects the acquired immune system by sequestering host IgG through the formation of insoluble complexes (Atkins KL *et al.*, 2008). In addition to Ig-binding domains I and II, Sbi contains two further domains (Sbi-III and

IV) that can bind C3d and scavenge the molecule from circulation (Burman JD *et al.*, 2008). Sbi-IV is a 11 kDa domain and only shares 19% sequence identity with Efb-C.

Crystal structures of C3d in complex with Efb-C, Ehp and Sbi-IV (PDB:2GOX, PDB:2NOJ and PDB:2WY8, respectively) provide insights into complement evasion and immunosuppression mechanism utilized by *S. aureus*. The proteins display very similar binding modes to C3d (Figure 1.5). For instance, the C3d-binding domain of Sbi (Sbi-IV) shares significant structural and functional similarities with Efb-C and Ehp. Despite minimal overall sequence identity, the three proteins display a mutual three-helix bundle fold and share two conserved C3d anchoring residues (R131 and N138 in Efb-C; R75 and N82 in Ehp; R45 and N52 in Sbi-IV). The conserved asparagine in Efb-C, Ehp and Sbi-IV forms a network of intermolecular hydrogen bonds with the main-chain atoms of residues V97, I100, A101 and I102 of C3d. Mutating the conserved arginine or asparagine on Efb-C, Ehp and Sbi to alanine resulted in a total loss of C3d-binding activity with regards to Efb-C and Sbi-IV (Hammel M *et al.*, 2007; Upadhyay A *et al.*, 2008) and a 400-fold decrease in the binding affinity to C3d with regards to Ehp (Hammel M *et al.*, 2007). Moreover, their interaction with C3d shares a high level of redundancy and utilizes the same residues on the negatively-charged concave surface of C3d (*i.e.* the binding surface for CD21 SCR1-2). It has been demonstrated that these proteins are capable of blocking the association of C3d and CD21 SCR1-2 (Ricklin D *et al.*, 2008; Isenman DE *et al.*, 2010), thereby interfering with the crucial link between the innate and acquired immunity. Table 1.2 summarizes the binding interactions of Efb-C, Ehp and Sbi-IV with C3d characterized by ITC.

Table 1.2 | Comparison of C3d binding data from Efb-C, Ehp and Sbi-IV using ITC

Ligand	n	K _D (nM)	ΔH (kJ/mol)	ΔS (J/mol)	IC ₅₀ (μM)*	Reference
Efb-C	1.0	2.0	-35.6	48.1	0.41	Hammel M <i>et al.</i> , 2007
Ehp	0.9	0.18	-36.6	62.8	0.12	Hammel M <i>et al.</i> , 2007
Sbi-IV	1.2	360	-15.9	69.9	10	Upadhyay A <i>et al.</i> , 2008

* Inhibition of the complement alternative pathway

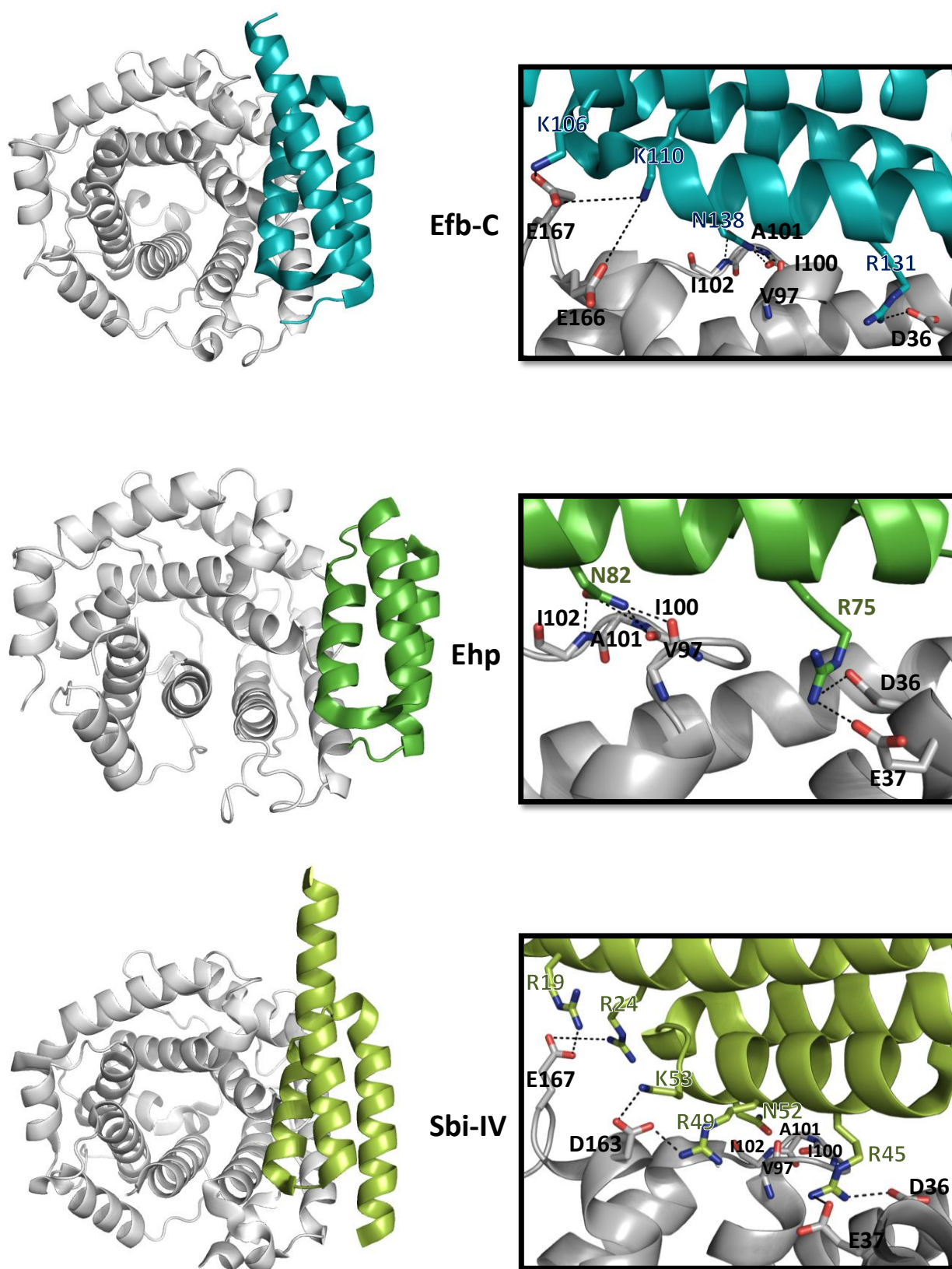


Figure 1.5 | Crystal structure of the C3d in complex with three different complement inhibitory proteins from *S. aureus*

Crystal structures of the protein complex C3d-Efb-C (top), C3d-Ehp (middle) and C3d-Sbi-IV (bottom). C3d and the *S. aureus* proteins are shown as cartoon ribbons. Detailed interactions between C3d and the *S. aureus* proteins are shown on the corresponding diagrams on the left.

(i) Complement regulatory protein factor H

Rapid responsiveness of the complement system is tightly controlled by complement regulatory proteins at multiple levels of the activation cascade to minimize damage to the host. Factor H acts on complement component C3b to ensure the alternative pathway of complement activation is maintained at a level sufficient to eliminate foreign antigen invasion without causing deteriorating effects to host cells (Panburn MK et al., 1977; Weiler JM et al., 1976; Whaley K and Ruddy S, 1976). It is a 155 kDa glycoprotein comprised of twenty SCR domains. Factor H is not localized on a cell membrane and utilizes its sixteen C-terminal SCR domains to recognize polyanionic self-surface makers to increase its retention on host cell surface. It associates most efficiently with C3b and C3 convertases that are attached to self-surfaces which bear unique polyanionic markers, such as glycosaminoglycans (GAGs) and sialic acid (Meri S and Pangburn MK, 1994; Pangburn MK, 2000). Factor H competes with factor B for binding with C3b so as to block the assembly of C3 convertase, as well as accelerates the decay of any C3 convertase that forms irreversibly. It is also a cofactor for factor I-mediated C3b degradation to yield iC3b that is incapable of binding factor B. Hence, the preferential enrichment of factor H on host cell surfaces over bacterial surfaces helps innate immunity to discriminate between self and pathogens. In the presence of factor H, host cells are protected from complement attack mediated by activation cascade downstream of C3 convertase. It has been shown that factor H SCR19-20 binds to both C3b and its final degradation fragment C3d (Lehtinen MJ *et al.*, 2009). A crystal structure of C3d in complex with factor H SCR19-20 (Morgan HP *et al.*, 2011; PDB:3OXU) is a surrogate for the experimentally yet unattainable C3b-factor H SCR19-20 structure and imitates the interaction between the C-terminus of factor H and C3b.

Interestingly, the proteins crystallized as a 2:1 complex with two copies of C3d and one factor H SCR19-20 molecules (Figure 1.6). Two distinct interaction sites were observed in the crystal structure – factor H SCR19 was bound to a surface comprised of α -helix 4 and α -helix 7 on C3d, whereas factor H SCR20 was bound to the negatively charged concave surface on C3d. Protein NMR spectroscopy and mutagenesis studies were used by the authors to investigate the C3d-factor H SCR19-20 interaction independently (Morgan HP *et al.*, 2011). Using chemical shift perturbation analysis, it was shown that eleven backbone resonances originally

presented in the HSQC spectrum of ^{15}N -labeled factor H SCR19-20 disappeared upon addition of C3d. Seven of these residues were strongly perturbed (>90% loss in signal intensities) and formed a contiguous binding surface on SCR19. They also corresponded to the residues found at the crystallographic C3d – factor H SCR 19 interface. Mutagenesis studies confirmed that D1119 and Q1139 in factor H SCR19, as well as E117 and D122 in C3d were key interacting residues in the protein complex. Double mutation of C3d E117A/D122A had the most deleterious effect on the binding to factor H SCR19-20 and decreased the binding affinity by 1600-fold. The result was consistent with the interface identified from NMR. All together the findings indicated factor H SCR19 was an imperative domain in mediating the interaction with C3d and the C3d-SCR19 crystallographic interface represented a physiologically relevant interaction. The authors suggested that negatively charged concave surface in C3d presumably participated in electrostatic steering effects upon binding to factor H SCR19-20. This phenomenon was also reported for the C3d-CD21 and C3d/Efb-C interactions (Morikis D and Lambris JD, 2004; Ricklin D *et al.*, 2008).

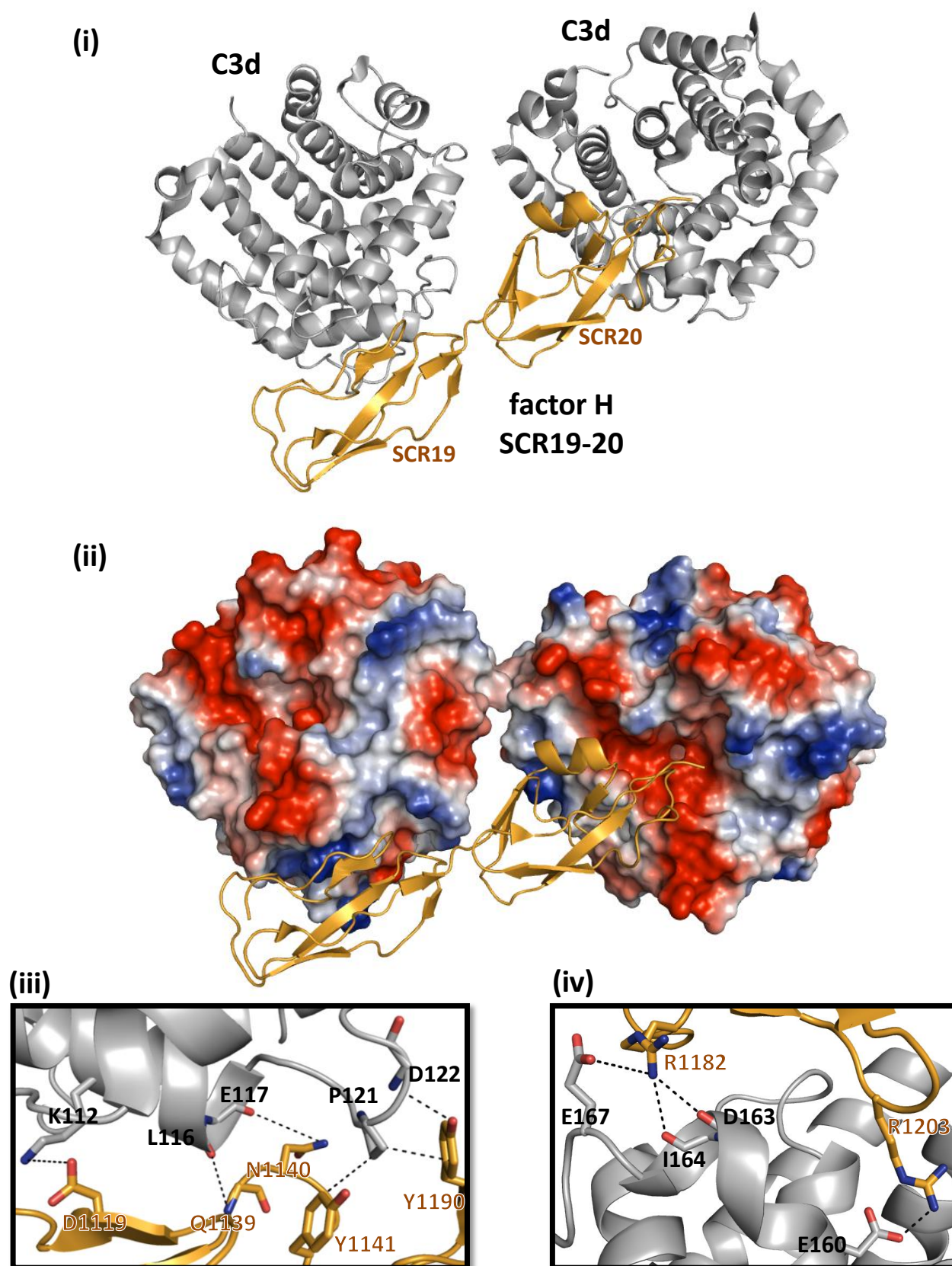


Figure 1.6 | Crystal structure of the C3d/factor H SCR19-20 protein complex solved by Morgan HP *et al.* (PDB: 3OXU)

- (i) Structure of the C3d/factor H SCR19-20 protein complex showing overall molecular architecture as cartoon diagram.
- (ii) Structure of the protein complex with C3d shown as electrostatic surface representation.
- (iii) Close-up diagram of the crystallographic C3d/factor H SCR19 binding interface.
- (iv) Close-up diagram of the crystallographic C3d/factor H SCR20 binding interface.

1.3 THERAPEUTICS – THE COMPLEMENT SYSTEM & ARTHRITIS

1.3.1 THERAPEUTIC INHIBITION OF THE COMPLEMENT SYSTEM

Upsetting the delicate balance of activation and regulation of the complement system can cause an overreaction that leads to destruction of the host's cells. Undue activation of the complement system triggers excessive inflammatory and cell lysis activities that are associated with a number of autoimmune, inflammatory and neurodegenerative diseases (Walport MJ, 2001; Volanakis JE and Frank M, 1998; Sahu A and Lambris JD, 2000). For instance, binding of complement component C1q to immune complexes deposited on host cell surfaces is thought to initiate the classical pathway and contribute to the pathogenesis of systemic lupus erythematosus and rheumatoid arthritis (Manderson AP *et al.*, 2004). In addition, C1q also serves as a PRR for amyloid plaques which is responsible for the pathogenesis of Alzheimer's disease (Bonifati DM and Kishore U, 2007). Complement fragment C5a was shown to have a role in complement-mediated inflammatory diseases such as asthma (Guo RF and Ward PA, 2005). Hence, although inhibition of complement activity has been recognized as a possible therapeutic strategy, numerous attempts in the development of complement-specific drugs have been largely disappointing (Sahu A and Lambris JD, 2000; Bureeva S *et al.*, 2005; Makrides SC, 1998; Lambris JD and Holers VM, 2000; Holland MC *et al.*, 2004). The complex nature and redundancy of the complement activation cascade and its diseases involvement is further complicated by the involvement of the complement system in acquired immunity. The lack of clarity in the exact disease-related mechanisms of complement at the molecular and clinical level often hinders specific targeting. Furthermore, many fundamental steps in the complement activation system utilize large protein-protein interactions that do not have a high druggability. Druggability refers to disease-related proteins that have a potential to be targeted by drug-like compounds (Keller TH *et al.*, 2006). A 'druggable genome' analysis was carried out to map the targets of currently marketed drugs to their protein families (Hopkins AL and Groom CR, 2002). Enzymes represent the dominant class of drug target (50%), closely followed by G-protein-coupled receptors (30%) and ion channels (7%). According to Hopkins and Groom, the only targets with promising druggability in the complement system were therefore serine proteases and anaphylatoxin

receptors. At present, an anti-C5 monoclonal antibody is the first and only complement-specific drug approved by the U.S. Food and Drug Administration (FDA) in 2007. Historic development of complement inhibitors of various classes was reviewed in the following section.

(i) Serine protease inhibitors

Early drug discovery efforts have focused on serine protease inhibitors as they appeared to be the most druggable targets in the complement system (Narayana SV *et al.*, 2000). There are eight integral serine proteases of the complement activation cascade, namely C1r, C1s, MASP-1, MASP-2, C3 convertases, factor D, factor B and factor I (Sim RB and Tsiftoglou SA, 2004). Development of small-molecule drug candidates targeting factor D (Szalai AJ *et al.*, 2000) and C1s (Muerke M *et al.*, 2001) has been largely unsuccessful due to their short half-lives and lack of specificity. C1-inhibitor (C1-inh) is a plasma-derived heavily glycosylated protein and is the only complement protease inhibitor currently marketed for the treatment of hereditary angioedema (Agostoni A *et al.*, 1980; De Serres J *et al.*, 2003; Wagenaar-Bos IG and Hack CE, 2006). The disease is characterized by a native C1-inh deficiency or mutation in the native C1-inh that leads to inactivity in the human body.

(ii) Anaphylatoxin receptor antagonists

Most of the pro-inflammatory activities appeared to signal through complement C5a receptor (C5aR), which is a G-protein-coupled receptor expressed on numerous immune cells. The most promising drug candidate targeting C5aR up to date is a cyclic peptidomimetic PMX-53, which was developed by Peptech (Kohl J, 2006). Although the drug showed positive results in a phase 1a/2b clinical trials with rheumatoid arthritis patients, its poor half-life (70 minutes) and rapid degradation in intestinal mucosa severely limits its use. The drug and its analogues were omitted from the development pipeline after subsequent company acquisition by Cephalon.

(iii) Soluble complement regulators

Complement regulatory proteins are native modifiers of the complement system to prevent host cell destruction and therefore have been considered for therapeutic use in the early stages of complement drug discovery (Brooks E *et al.*, 2005). Soluble forms of complement receptor 1 (sCR1), MCP, DAF, CD95 and their analogues have been developed but sCR1 was discontinued from clinical trials and others were no longer listed in companies' pipeline (Sahu A and Lambris JD, 2000; Li JS *et al.*, 2006; Avant Immunotherapeutics Inc, 2007).

(iv) Therapeutic antibodies

Monoclonal antibodies (mAb) against complement component C5 (eculizumab; Soliris™) was the first and only FDA-approved complement-specific drug in the market (US Food and Drug Administration, 2007; Alexion Pharmaceuticals Inc., 2007). The mAb was developed more than two decades ago (Frei Y *et al.*, 1987) and Alexion Pharmaceuticals demonstrated its effectiveness in a mouse model of rheumatoid arthritis (Wang Y *et al.*, 1995). The drug was approved for treatment of paroxysmal nocturnal hemoglobinuria (PNH) and tested in clinical studies for other pathological conditions in Nephrology, Hematology, Neurology and Ophthalmology (Alexion Pharmaceuticals Inc.).

(v) Low molecular weight complement component inhibitors

Compstatin is the only complement-based small-molecule inhibitor that has successfully entered clinical trials. It is a cyclic tridecapeptide that prevents the cleavage of complement component C3 to C3a and C3b (Sahu A *et al.*, 1996), thereby inhibiting the crucial central step in the complement activation cascade. A co-crystal structure of compstatin in complex with complement fragment C3c suggested that an interruption in protein-protein interactions during the formation of C3 convertase is responsible for its inhibitory activity (Janssen BJ *et al.*, 2007). The drug demonstrated effectiveness in a variety of experimental disease models (Holland MC, Morikis D and Lambris JD, 2004). A compstatin analog (POT-4) with improved inhibitory efficacy of 264-fold completed phase 1 clinical study for treatment of age-related macular degeneration in 2008

(Katragadda M *et al.*, 2006), but no further studies have been initiated recently.

1.3.2 THERAPEUTIC MANAGEMENT OF RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is a chronic inflammatory autoimmune disease characterized by destruction of joint tissues that often leads to disability. It affects 0.5 – 1% of the population in developed countries and 50% of patients are unable to sustain full-time occupation within ten years of disease onset (Gabriel SE, 2001; Wolfe FE and Hawley DJ, 1998; World Health Organization: <http://www.who.int/chp/topics/rheumatic/en/>). The disease has been estimated to create an economic burden of €30 billion per year in Europe alone (Lundkvist J *et al.*, 2008). The cause of RA remains elusive and many immunological pathways have been shown to be involved in the generation of the disease. Notably, there is a strong association between RA and several autoantibodies (Steiner G and Smolen JS, 2002), such as rheumatoid factor which targets the Fc fragment of IgG. A number of autoantigens are also commonly known to trigger B-lymphocyte and T-lymphocyte activation leading to inflammatory and autoimmune responses (Fritsch R *et al.*, 2002; Schellekens GA *et al.*, 1998). Although the disease can be relatively well-controlled with current therapeutic agents, the treatment for RA is associated with high costs and can cause a reduction in life expectancy if not handled appropriately (Pincus TE *et al.*, 1994; Yelin E and Wanke LA, 1999; Pugner KM *et al.*, 2000). There are two principle RA drug therapies for disease management – symptomatic treatment with non-steroidal anti-inflammatory drugs (NSAIDs) and disease-modifying anti-rheumatic drugs (DMARDs). NSAIDs provide symptomatic relief namely by hampering the generation of prostaglandin by cyclooxygenase during inflammatory responses, but do not interfere with underlying joint destruction mechanisms. In contrast, DMARDs modulate all aspects of the disease process and symptomatic therapies can be omitted once DMARDs become efficacious in disease management. However, remissions only occur in approximately a quarter of the patient population and discontinuing DMARD therapy is often associated with a significant risk of flares (ten Wolde S *et al.*, 1996). Current drug therapies do not entirely cure RA, and using a single drug alone acting on individual protein target has proven insufficient to retard disease progression owing to pathogenic complexity and redundancy

of the immune system. Therefore, it is necessary to inhibit several targets in alliance until the cause of RA is fully understood. Two classes of DMARDs are presently used in RA, namely small-molecule inhibitors and monoclonal antibodies (mAb).

(i) Small-molecule DMARDs

Methotrexate (MTX) forms a foundation in the treatment of RA and is currently the gold standard of DMARD therapy (Cronstein BN, 2005). It is usually given concomitantly with other DMARDs for disease management. Although initially introduced for chemotherapy in the treatment of cancer, MTX has been employed in the treatment of RA since the 1980's (Weinblatt ME *et al.*, 1985). It dampens the metabolism of folic acid by inhibiting dihydrofolate reductase and thymidylate synthase during DNA and RNA synthesis. Folic acid is crucial for the *de novo* synthesis of the nucleoside thymidine and purine synthesis. Its mechanisms of action in the treatment of RA remains unclear, but it has been shown that the anti-inflammatory action of MTX is mediated by the suppression of T-lymphocyte activation and cell adhesion molecule (*e.g.* ICAM-1) expression by T-lymphocytes (Johnston A *et al.*, 2005). Leflunomide is another small-molecule inhibitor for RA of note and acts on dihydroorotate dehydrogenase to inhibit the *de novo* synthesis of pyrimidine (Smolen JS *et al.*, 1999). The drug interferes with the generation of pro-inflammatory cytokines, immune cell proliferation and migration which are involved in the pathogenesis of RA (Manna SK and Aggarwal BB, 1999; Siemasko KF *et al.*, 1998). All small-molecule DMARDs currently used have limited efficacy and potential adverse side-effects in prolonged use.

(ii) Monoclonal antibody DMARDs

Monoclonal antibodies targeting pro-inflammatory cytokines (*e.g.* TNF and interleukins) have been developed to modulate the progression of RA. TNF- α can trigger systemic inflammatory response and is secreted by a variety of immune cells, such as monocytes, macrophages and other lymphoid cells. It has been demonstrated that an over-expression of TNF in transgenic mice leads to the development of severe arthritis (Keffer J *et al.*, 1991). Consequently, there are five anti-TNF mAb on the market presently used for RA, with

Golimumab (Mazumdar S and Greenwald D, 2009) and Certolizumab pegol (Goel N and Stephens S, 2010) being the latest addition and both received FDA approval in 2009. There are three further mAbs in the market targeting interleukin-6 receptor or other proteins expressed on the surface of immune cells, namely abatacept (Moreland L *et al.*, 2005), rituximab (Cohen SB *et al.*, 2006) and tocilizumab (Venkiteshwaran A, 2009), which received FDA approval in 2010. Monoclonal antibody DMARDs are considered a newer class of RA drugs and are only issued to patients who haven't responded or have adverse side-effects to conventional therapy. Although recent mAb therapy appeared to be more efficacious in RA patients who have responded in clinical practice, this class of drugs do have side effects (Bathon JM *et al.*, 2000; Elliott MJ *et al.*, 1994; Lipsky PE *et al.*, 2000; Weinblatt ME *et al.*, 1999, Moreland LW *et al.*, 1997; Weinblatt ME *et al.*, 2003; Kempeni J, 2000; Furst D *et al.*, 2003; Keane J *et al.*, 2001; Mohan N *et al.*, 2001). Furthermore, large scale production of therapeutic mAbs is expensive and the high financial burden associated with the new mAb therapy is likely to become a societal issue given the prevalence and chronic nature of RA. Consequently, there is a growing interest in finding novel small-molecule inhibitors for disease management that are relatively inexpensive to manufacture.

As discussed in section 1.1.3, the C3d/CD21 SCR1-2 interaction is implicated in the development of rheumatoid arthritis and other autoimmune disorders. There is a potential to therapeutically exploit the C3d/CD21 SCR1-2 interaction through the rational design of inhibitors that would disrupt the protein-protein interaction. C3d is the final degradation product of the complement activation cascade and thus inhibitors of the interaction would not interfere with upstream complement activation steps. As a result, developing small molecule inhibitors of the C3d/CD21 SCR1-2 interaction would allow alternative treatment to current mAb DMARDs. It is hoped that inhibitors arose from this project would provide a more cost effective therapeutic management for rheumatoid arthritis with minimum immunosuppression side effects. Although the controversy regarding atomic level details of the C3d/CD21 SCR1-2 binding interface is still unresolved, recent mutagenesis studies of the binding interface will provide the platform for the design of compounds that exploit this interface as a therapeutic target (Clemenza L and Isenman DE, 2000; Hannan JP *et al.*, 2005; Bomb

M, 2008). The surface identified on C3d is a particularly promising site for the development of small molecule inhibitors. The binding site presents a deep hydrophobic groove, suggesting that the hydrophobic residues on the CD21 SCR1-2 surface cannot insert deeply enough to make full use of the potential binding surface. Consequently, a well designed small molecule could make optimal use of the groove, achieving better affinity and specificity than the native protein ligand.

1.4 AIM OF THESIS

The crystal structure of the C3d-CD21 SCR1-2 protein complex (PDB: 1GHQ) and NMR chemical shift perturbation analysis presented in the literatures do not seem to describe a model that represents physiologically relevant interactions between the proteins. Further structural clarification of the C3d-CD21 SCR1-2 interaction is needed to address the precise binding model of the proteins and to define physiological relevant binding sites for structure-based drug-design campaign. Complement C3d is the final degradation product of component C3 and apparently presents an optimal drug target within the complement system. Prevention of terminal complement activities, *e.g.* inhibition of the C3d-CD21 SCR1-2 interaction, is likely to suppress pathogenic complement-associated activities without compromising its native immune defense functions when treating chronic pathological conditions like arthritis. Successful small-molecule inhibitors could serve as a potential therapeutic agent to complement traditional treatment for rheumatoid arthritis or *S. aureus*-induced pathological conditions. Consequently, the following studies have been proposed in this thesis:

- Further mutational studies and microcalorimetric analysis to delineate the respective binding site on C3d and CD21 SCR1-2 and their binding model in solution.
- NMR backbone assignment of C3d to enable further assessment of the CD21 SCR1-2 binding site on C3d to complement results in the literature. It would also enable NMR-based screening and characterization of small-molecule inhibitors targeting C3d.
- Fragment-based drug screening against C3d to identify potential small-molecule inhibitors.

- Production of diffraction quality C3d crystals to provide a platform for determination of co-crystal structures of C3d with small ligands.
- Crystallization of the C3d-CD21 SCR1-2 protein complex to obtain a new crystal structure that represents physiologically relevant interactions.

CHAPTER II

MATERIALS AND METHODS

2.1 BUFFERS, SOLUTIONS AND MEDIA

CMFF loading buffer: 20 mM sodium acetate, 20 mM NaCl, 1 mM EDTA, pH 5.5

CMFF elution buffer: 20 mM sodium acetate, 300 mM NaCl, 1 mM EDTA, pH 5.5

Coomassie de-stain solution: 10 % acetic acid, 20 % methanol, 70 % H₂O

Coomassie stain solution: 0.1 % coomassie dye, 10 % acetic acid, 40 % methanol, 50 % H₂O

EDC: 400 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride

HEPES buffer: 10 mM HEPES, 150 mM NaCl, pH 6.8

LB media: 1 % peptone, 0.5 % yeast extract, 1.0 % NaCl, pH 7.0

LB agar: LB medium plus 1.5 % agar

MonoQ loading buffer: 10 mM NaH₂PO₄, 40 mM NaCl, 2 mM EDTA, pH 7.1

MonoQ elution buffer: 10 mM NaH_2PO_4 , 1M NaCl, 2 mM EDTA, pH 7.1

M9 media: 42 mM Na_2HPO_4 , 22 mM KH_2PO_4 , 9 mM NaCl, 19 mM NH_4Cl , 0.2 % (v/v) glucose, 1 mM MgSO_4 and 100 μM CaCl_2 , pH 7.4

NHS: 100 mM N-hydroxysuccinimide

Ni loading buffer: 20 mM NaH_2PO_4 , 500 mM NaCl, pH 7.5

Ni elution buffer: 10 mM NaH_2PO_4 , 500 mM NaCl, 500 mM imidazole, pH 7.5

Non-labeled amino acid mix: nineteen out of twenty proteogenic amino acids (excluding isotope-labeled amino acid) at concentrations of 50 mg / L each.

PBS: 10 mM NaH_2PO_4 , 137 mM NaCl, 2.7 mM KCl, pH 7.4

PDEA: 80 mM 2-(2-pyridinyldithio)ethaneamine in 100 mM sodium borate, pH 8.5

Phosphate buffer: 20mM NaH_2PO_4 , 150 mM NaCl, 0.01% NaN_3 , pH 6.8

Refolding buffer: 0.02 M ethanolamine, 1 mM EDTA, 0.5 M arginine, 1 mM cysteine, 2 mM cystine, pH 9.5

SDS-PAGE loading buffer: 100 mM Tris-HCl, 200 mM DTT, 4 % SDS, 20 % (v/v) glycerol, 0.2 % bromophenol blue

SDS-PAGE resolving gel: 15 %: 0.4 mM Tris, 15 % acrylamide, 0.01 % SDS, pH 8.8. Polymerised with 0.01 % ammonium persulphate and 0.5 % TEMED

SDS-PAGE running buffer: 25 mM Tris, 250 mM glycine, 3.5 mM SDS, pH 8.3

SDS-PAGE stacking gel: 0.43 mM Tris, 4 % acrylamide, 0.01 % SDS, pH 6.8. Polymerised with 0.025 % ammonium persulphate and 1.25 % TEMED

Solubilisation buffer: 8M urea, 10 mM Tris-HCl, 1 mM EDTA, 25 mM DTT, pH 8.0

Solubilisation dialysis buffer: 6M urea, 10 mM HCl

SPR running buffer: 20 mM sodium phosphate, 150 mM NaCl, 0.005 % tween-20, pH 6.8.

Thiol blocking solution: 50 mM cysteine in 100 mM NaCl pH 4.0

2.2 MOLECULAR BIOLOGY

2.2.1 C3d DNA VECTORS

C3d comprises the amino acids Asp 996 to Arg 1303 of mature C3 using the numbering from the Swiss-Prot accession code P01024. C3d contains three cysteine residues at positions 1010, 1101 and 1158. Cys 1010 forms part of the reactive thioester moiety in C3. This was mutated to an alanine residue to ensure that only one disulphide linkage could form between Cys 1101 and Cys 1158 in recombinant C3d. A *pET 15b* expression construct of C3d (Cys 1010 Ala) was kindly donated by Prof. David Isenman, University of Toronto. The expressed protein contains an additional eight amino acids (MLDAERLK) as compared to enzymatically produced C3d. Whereas the first two amino acids (ML) are vector derived, the other six (D996 – K1001) correspond to the C-terminal fragment of C3dg. The amino acid numbering used forthwith in this thesis corresponds to the C3d residue numbers (Met1 – Arg310) to keep in line with the published literature. The vector was selected by resistance to ampicillin (Melford) and propagated in the *E. coli* DH5 α strain (Invitrogen) for subcloning or BL21 (DE3)* strain (Invitrogen) for protein expression. The C3d protein construct is documented in Appendix II.

2.2.2 CD21 SCR1-2 DNA VECTORS

CD21 SCR1-2 comprises the amino acids Ser 2 to Ser 128 using the numbering from the experimentally determined N-terminal sequence of the mature protein (Fingerroth JD, Clabby ML and Strominger JD, 1988). This corresponds to residues Ser 22 – Ser 148 of the unprocessed precursor, containing the signal peptide, using the numbering of the Swiss-Prot accession code P20023. The construct used differs by one residue from that used in the crystal structure of C3d/CD21 SCR1-2 complex (Szakonyi *et al.*, 2001). A pIVEX 2.3 expression construct of CD21 SCR1-2 fused with a C-terminal hexa-His tag was kindly donated by Professor Joyce Fingerroth, Harvard Medical School. The vector was selected by resistance to kanamycin (Invitrogen) and propagated in the *E. coli* DH5 α strain (Invitrogen) for subcloning or BL21 (DE3)* strain (Invitrogen) for protein expression. The CD21 SCR1-2 protein construct is documented in Appendix II.

2.2.3 SITE-DIRECTED MUTAGENESIS

Site-directed mutagenesis was carried out using a QuickChange II site directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. A number of point mutations were introduced in both the C3d and CD21 SCR1-2 constructs. The primers that were used are listed in Appendix 1.

Template specific forward and reverse primers were designed using Stratagene's web-based software programme <http://labtools.stratagene.com/QC>. The mutagenic oligonucleotide primers contained desired mutation in the middle of the primer with 10-15 bases of correct sequence on both sides. Primers were between 25 -45 bases in length, with a melting temperature (T_m) of $\geq 78^\circ\text{C}$. The site-directed mutagenesis was performed using *PfuUltra™* high-fidelity DNA polymerase for mutagenic primer-directed replication of both plasmid strands with the highest fidelity. The procedure utilized a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest and two synthetic oligonucleotide primers, both containing the desired mutation (Figure 2.1). The mutagenic oligonucleotide primers, each complementary to opposite strands of the vector, annealed to the plasmid and were extended during temperature cycling by *PfuUltra* HF DNA polymerase, without primer displacement. Following temperature cycling, the product was treated with *Dpn* I endonuclease. The endonuclease is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue supercompetent cells.

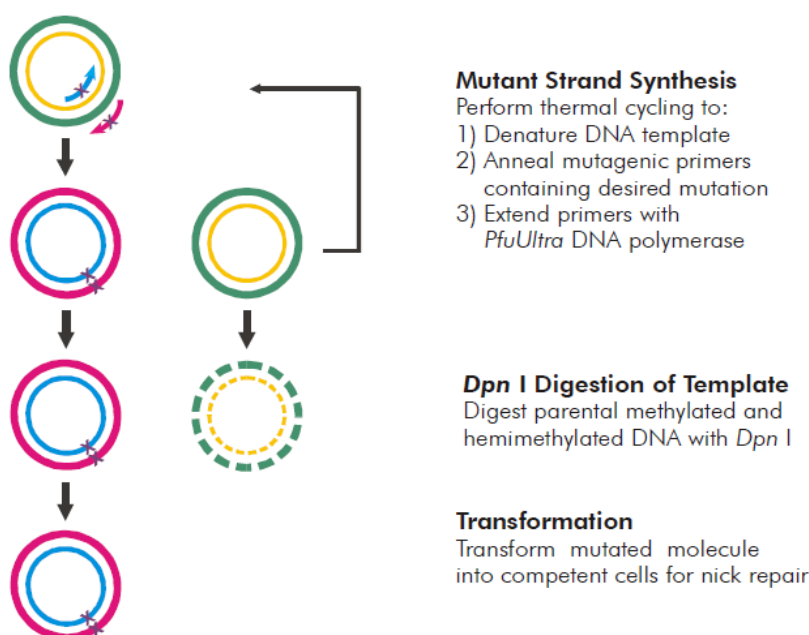


Figure 2.1 | Site-directed mutagenesis

Figure adapted from QuickChange II site directed mutagenesis kit's manual. Parental plasmid is in green/yellow and mutagenic primers are in blue and pink. The parental plasmid is digested with *Dpn* I after PCR reaction.

2.3 E. coli PROTEIN EXPRESSION AND PURIFICATION

2.3.1 EXPRESSION AND PURIFICATION OF RECOMBINANT C3d

C3d was expressed in *E. coli* BL21 (DE3) grown in LB medium containing 100 µg/ml ampicillin, from frozen glycerol stock. For protein production, a 5 ml starter culture was diluted to 1 litre LB and grown at 37°C until an A600 of 0.8, at which point addition was made to achieve final concentration of 0.25 mM IPTG. After further growth for 20 hours at 16°C, bacteria were harvested by centrifugation for 20 min at 5000 x g.

The pelleted cells were suspended in 20 ml of CMFF loading buffer containing 10X protease inhibitors (Sigma) and subjected to five cycles of sonication (Misonix XL-2000, 40 s each) with 15 min periods of cooling on ice between sonication steps. After centrifugation for 30 min at 20,000 x g, the supernatant was filtered through a Millipore Millex-GP 0.22 µm filter and loaded onto a HiPrep 16/10 carboxymethyl fast flow column (CMFF column, GE Healthcare) previously equilibrated in CMFF loading buffer. The column was washed with five column volumes of CMFF loading buffer and C3d was eluted in 8 ml fractions with CMFF elution buffer. Protein fractions were pooled and dialyzed three times against 500 ml of MonoQ loading buffer and loaded onto a HiPrep 16/10 Mono Q column (GE Healthcare) previously equilibrated in the MonoQ loading buffer. C3d was eluted in 8 ml fractions with a gradient to 35% MonoQ elution buffer (10 mM NaH₂PO₄, 1M NaCl, 2 mM EDTA, pH 7.1) over 6 column volumes. Protein fractions were pooled and dialyzed three times against 500 ml of PIPES buffer (10 mM PIPES, 150 mM NaCl, pH6.0). The dialyzed sample was concentrated with a Vivaspin 20 centrifugal concentrator MWCO 10,00 Da (Sigma-Aldrich) and applied to a HiLoad 26/60 Superdex 75 gel filtration column (GE Healthcare) previously equilibrated in the phosphate buffer. Fractions containing monomeric C3d were pooled and concentrated.

2.3.2 EXPRESSION OF ¹⁵N AND ²H, ¹⁵N, ¹³C-LABELLED RECOMBINANT C3d

Isotope labelled C3d was expressed in the *E. coli* host strain BL21 (DE3)*. ¹⁵N- and ²H, ¹⁵N, ¹³C-labeled C3d was prepared in M9 media with the replacement of 1 g/L ¹⁵NH₄Cl or 1 g/L ¹⁵NH₄Cl and 3 g/L ¹³C-glucose in 1L D₂O (Spectra Stable Isotopes), respectively. Overnight cultures were diluted 50 fold into fresh media and grown in an orbital shaker at 37°C until OD600 = 0.8 at 250 rpm. IPTG was added to 1 mM to induce protein

expression. After further growth for 20 hours at 16°C, bacteria were harvested by centrifugation for 20 min at 5000 x g.

2.3.3 EXPRESSION OF RESIDUE SPECIFIC ¹⁵N-LEUCINE LABELED RECOMBINANT C3d

Isotope labelled C3d was expressed in the *E. coli* host strain BL21 (DE3)*. ¹⁵N-Leucine labeled C3d was prepared in M9 media. Overnight cultures were diluted 50 fold into fresh media and grown in an orbital shaker at 37°C until OD600 = 0.8 at 250 rpm. IPTG was added to 1 mM to induce protein expression. The following components were added forty minutes after the induction: 50 mg / L of rifampicin, 50 mg / L of non-labeled amino acid mix, 70 mg / L of ¹⁵N-labeled amino acid and 110 mg / L of aminooxyacetic acid. After further growth for three hours at 37°C, bacteria were harvested by centrifugation for 20 min at 5000 x g.

2.3.4 EXPRESSION AND PURIFICATION OF RECOMBINANT CD21 SCR1-2

CD21 SCR1-2 was expressed in *E. coli* BL21 (DE3) grown in LB medium containing 30 µg/ml kanamycin from frozen glycerol stock. For protein production, a 5 ml starter culture was diluted to 1 litre LB and grown at 37°C until an A600 of 0.8, at which point addition was made to achieve final concentration of 1mM IPTG. After further growth for 4 hours, bacteria were harvested by centrifugation for 20 min at 5000 x g.

The pelleted cells were suspended in 20 ml of 1x PBS and subjected to five cycles of sonication (Misonix XL-2000, 40 s each), with 5 min periods of cooling on ice between sonication steps. Inclusion bodies were solubilised in 20 ml of solubilisation buffer for 2 hours at 4 °C and acidified to pH 4.0 prior to centrifugation at 25,000 x g for 30 min. The 8M urea and 25 mM DTT in solubilisation buffer were used to denature expressed proteins and to prevent disulphide bonds formation prior to refolding, respectively. The supernatant was dialyzed against three changes of 25 volumes of solubilisation dialysis buffer and diluted slowly into fifty volumes of degassed refolding buffer. The refolding buffer contained 0.02 M ethanolamine and 0.5 M arginine to help stabilize the refolding process by preventing protein aggregation. It also contained 1 mM cysteine and 2 mM cystine that served as thiol/disulfide redox reagents to facilitate renaturation of the refolding process (Raman B *et al.*, 1996). The protein was left to refold overnight at 4°C and concentrated to 50 ml using a

tandem Sartorius Vivaflow 200 modules 5000 MWCO PES. The sample was dialyzed four times against ten volumes of Ni loading buffer and loaded onto a 10 ml freshly generated Ni-NTA agarose column (Qiagen) previously equilibrated in Ni loading buffer. CD21 was eluted in 4 ml fractions with a gradient to 100% Ni elution buffer at 1 ml/min over 50 min. Protein fractions were pooled and dialyzed three times against 500 ml of phosphate buffer. The dialyzed sample was concentrated with a Vivaspin 20 centrifugal concentrator MWCO 5,000 Da (Sigma-Aldrich) and applied to a HiLoad 26/60 Superdex 75 gel filtration column (GE Healthcare) previously equilibrated in phosphate buffer. Fractions containing monomeric CD21 SCR1-2 were pooled and concentrated.

2.4 PROTEIN CHARACTERIZATION TECHNIQUES

2.4.1 PROTEIN CONCENTRATION

Concentration of protein samples were determined by absorbance at 280 nm (A₂₈₀) using a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The absorption at 280 nm gives a measure of protein concentration, based on a molar extinction coefficient from the amino acid composition of a protein (Gill S.C. and von Hippel P.H., 1989). The exact molar extinction coefficients of wild type C3d and CD21 SCR1-2 were determined by amino acid analysis (Sittampalam GS, Ellis RM, Miner DJ, Rickard EC, Clodfelter DK, 1988) and the experiments were performed at the protein characterization facility, Department of Biochemistry, University of Oxford. The molar extinction coefficient values determined for C3d and CD21 SCR1-2 are 67130 M⁻¹ cm⁻¹ and 26199 M⁻¹ cm⁻¹, respectively.

2.4.2 MASS SPECTROMETRY

The molecular masses of C3d and CD21 SCR1-2 protein samples were determined by mass spectrometry (MS). Purified protein samples were de-salted using reversed phase HPLC on a C4 column (Agilent Technologies) prior to MS analysis. All MS spectra were recorded on a Bruker MaXis ESI-Q TOF mass spectrometer at Randall Divison of Cell & Molecular Biophysics, King's College London.

2.4.3 SDS-PAGE

Protein purity was assessed using standard SDS-PAGE techniques (Sambrook J and Russell D, 2001). Protein samples were boiled for 5 minutes at 100 °C in 1 x loading dye, containing 200 mM DTT for reducing samples. They were loaded onto 12.5 % polyacrylamide gels (BioRad Mini Gel system). Low-molecular weight markers (GE Healthcare) were run alongside protein samples. Typically 5 – 10 µg of protein samples were loaded in a total volume of 20 µl. The gels were run in 1 x SDS-PAGE running buffer at 150 V until the dye front reached the bottom of the plate. The gels were removed from the tank and placed into coomassie stain solution for 3 hours. Gels were destained in coomassie de-stain solution and dried using the DryEase Mini-Gel Drying System (Invitrogen).

2.4.4 ONE DIMENSIONAL ¹H-NMR SPECTROSCOPY

¹H-NMR spectra of unlabelled protein samples were recorded on a Bruker Avance III 700 MHz spectrometer equipped with a cryoprobe at 25 °C to check for correct protein folding. Characteristic upfield shifted methyl peaks beyond 0 ppm were used as indicators of correct folding.

2.5 BIOPHYSICAL INTERACTION ANALYSIS

2.5.1 DIFFERENTIAL SCANNING FLUORIMETRY

All experiments were carried out using a Stratagene Mx3005P QPCR system (Agilent Technologies) in 96-well plate format according to a well-established differential scanning fluorimetry protocol (Niesen FH, Berglund H and Vedadi M, 2007). For a full-plate of fragment screen, a 3 ml protein solution consisting of 5 µM wild type C3d in phosphate buffer supplemented with 4% DMSO and 5x SYPRO orange (Invitrogen) was prepared. A fixed 19.6 µl aliquot of the protein solution was dispensed into each of the wells of the PCR plate using a 50 µl multichannel pipette. A fragment library comprised one thousand diverse drug-like low molecular weight fragment (< 300 MW) was kindly provided by Dr. David Mountford (Institute of Pharmaceutical Science, King's College London) at concentrations of 50 mM in 100% DMSO. Fragment library was dispensed into each of the wells of the PCR plate with a Mosquito Nanodrop Crystallization robot (Molecular Dimension) typically at 400

nl to yield a final fragment concentration of 1 mM per well. The PCR plate was sealed with an optical transparent seal and placed into the PCR instrument. The temperature scan was run from 25 to 95 °C at 1 °C min⁻¹. Data analysis was carried out with the 'DSF analysis' Excel file that is available for download from <ftp://ftp.sgc.ox.ac.uk/pub/biophysics>.

2.5.2 SURFACE PLASMON RESONANCE

All experiments were carried out using a Biacore™ T100 system at 25 °C according to the manufacturer's guidelines (Biacore AB). A C3d construct, engineered to have a free exposed thiol group spatially on the opposite surface of the CD21 binding site, was generated by reversing the initial Cys1010Ala mutation (corresponding to Cys 17 of the expressed protein). The resulting mutant protein, A17C, was immobilized onto a CM4 sensor chip by thiol coupling of the free cysteine using a ligand thiol coupling kit (BIAcore AB). The sensor surface was activated with a 50 µl pulse of EDC/NHS and reactive disulfide groups were introduced with a 50 µl pulse of PDEA. C3d A17C in 20 mM phosphate, 150 mM NaCl, pH 6.8 was immobilized to densities of 100 – 7000 RU according to experiment type. Excess reactive groups were deactivated with a 40 µl pulse of thiol blocking solution. All steps carried out at a flow rate of 10 µl/min. All SPR experiments were carried out in SPR running buffer at a flow rate of 30 µl/min to minimize mass transport effects. The ligand contact time was 60 seconds with a dissociation time of 1200 seconds. A blank flow cell was used as a control surface.

For fragment-based screening experiments, C3d A17C was immobilized to densities of at least 5000 RU. Fragments were screened in a 96-well plate format and analyte concentrations were at 500 µM in SPR running buffer supplemented with 4% DMSO. Throughout the experiments, the SPR running buffer was supplemented with 4% DMSO to ensure ligand solubility. Buffer blanks and CD21 SCR1-2 were injected at eighteen different time points during the experiment to monitor reliability of the sensor surface. Fragments that showed reasonable binding profiles in the preliminary screen were subjected to secondary screening. In the secondary screen, analyte concentrations were injected in a titration series of "low to high" analyte concentration from 63 µM to 2 mM. All other experimental parameters were kept the same as the

preliminary screen. For affinity/kinetic analysis, C3d A17C was immobilized to densities of 100 RU. Analyte concentrations were injected in a titration series of “low to high” analyte concentration. Buffer blanks were also included in the titration series. SPR data was processed with the Biaevaluation software 3.1 (Biacore AB). Curve fitting and other analysis was performed using MicroCal Origin v7 (OriginLab).

2.5.3 ISOTHERMAL TITRATION CALORIMETRY

Experiments were performed using a iTC₂₀₀ microcalorimeter (MicroCal) at 25 °C. For most experiments, ITC buffer consisted of 20 mM sodium phosphate, 50 mM NaCl, pH 6.8. The wildtype interaction between C3d and CD21 SCR1-2 was also repeated at physiological relevant ionic strengths (20 mM sodium phosphate, 150 mM NaCl, pH 6.8). The protein to be titrated, wildtype or mutated CD21 SCR1-2 or wildtype or mutated C3d, was placed in the calorimetric cell at a concentration of 30-50 µM. This was titrated by a series of small injections of the appropriate ligand, present at a concentration of 300-600 µM in the ITC syringe, until a molar ratio of 3:1 to 4:1 was reached. Since the heat evolved and the binding affinity of the interaction is directly linked to the concentrations of both interaction partners used, a control heat of dilution experiment was performed in each case and subtracted from the actual protein-ligand interaction experiment. The control experiment consisted of adding protein into the calorimetric cell containing only buffer. Results were analyzed using the Origin Microcal v7 LLC ITC software (MicroCal).

2.6 NUCLEAR MAGNETIC RESONANCE

2.6.1 SAMPLE PREPARATION AND DATA PROCESSING

Protein samples were concentrated to volumes of 300 µl, supplemented with 5 % v/v D₂O and transferred into Shigemi tubes (Shigemi Inc.). Table 2.1 shows a summary of sample conditions and experimental application. All NMR experiments were performed on a Bruker Avance III 500 MHz spectrometer equipped with a TCI ¹H/¹³C/¹⁵N CryoProbe or a Bruker Avance III 700 MHz spectrometer equipped with a QCI ¹H/¹³C/¹⁵N/³¹P CryoProbe (Bruker BioSpin). All NMR experiments were recorded at 25°C. The proton carrier frequency was set on water. Spectra were processed with TopSpinTM software (Bruker BioSpin) and analyzed

using CcpNmr Analysis (CCPN, University of Cambridge). Sequence specific backbone assignment was assisted by the automated assignment program PACES (Coggins BE and Zhou P, 2003) and UNIO-MATCH (Volk J, Herrmann T and Wüthrich K, 2008).

Table 2.1 | Summary of samples used in NMR experiments

Sample	Concentration (μM)	Buffer conditions	Experimental Application
^{15}N-labeled C3d	200	20 mM NaHPO_4 , 50 mM NaCl, pH 6.8	Salt titration and optimizing experimental conditions
	400	20 mM NaHPO_4 , 50 mM NaCl, pH 6.8	Salt titration and optimizing experimental conditions
	180	20 mM NaHPO_4 , 150 mM NaCl, pH 6.8	Chemical shift perturbation analysis. Ligand titrated: <i>CD21 SCR1-2</i>
	180	20 mM NaHPO_4 , 400 mM NaCl, pH 6.8	Chemical shift perturbation analysis. Ligand titrated: <i>CD21 SCR1-2</i>
	180	20 mM NaHPO_4 , 150 mM NaCl, pH 6.8	Chemical shift perturbation analysis. Ligand titrated: <i>factor H peptide</i>
	150	20 mM NaHPO_4 , 150 mM NaCl, pH 6.8 in 4% DMSO	Fragment hits validation. Ligand titrated: <i>chemical fragments</i>
^2H, ^{13}C, ^{15}N-labeled C3d	350	20 mM NaHPO_4 , 150 mM NaCl, pH 6.8	Triple-resonance NMR experiments ^{15}N -NOESY-HSQC experiment
^{15}N-leucine labeled C3d	100	20 mM NaHPO_4 , 150 mM NaCl, pH 6.8	Selective ^{15}N -Leucine labeling to confirm partial backbone assignment

2.6.2 NMR EXPERIMENTS FOR SEQUENTIAL BACKBONE ASSIGNMENT

Sequence-specific assignments of backbone resonances were obtained by using standard NMR experiment as described below, namely 2D [^1H , ^{15}N] TROSY-HSQC and TROSY-based 3D [^1H , ^{15}N , ^{13}C]-HNCA, -HN(CO)CA, -HNCACB, -HN(CO)CACB, -HNCO, -HN(CA)CO, as well as [^1H , ^1H , ^{15}N]-HSQC NOESY experiments.

2.6.2 (i) [^1H , ^{15}N] TROSY-HSQC

Standard ^1H , ^{15}N -heteronuclear single quantum coherence (HSQC) experiments (Cavanagh J, Fairbrother WJ, Palmer AG, Rance M and Skelton NJ, 2007) were acquired. Transverse relaxation optimized spectroscopy

(TROSY)-HSQC experiments (Pervushin K, Riek R, Wider G and Wüthrich K, 1997) were also acquired to aid studies of large protein molecules. During the experiment, magnetization is transferred from hydrogen to attached ^{15}N nuclei through J-coupling. The chemical shift is evolved on the nitrogen and the magnetization is transferred back to the hydrogen for detection. A typical HSQC spectra show all H-N correlations which predominately are backbone amide groups, but Trp, Asn and Gln side-chain are also visible. TROSY-HSQC improves spectral resolution by reducing linewidths and eliminates Asn and Gln side-chain resonance to reduce spectral overlap issue.

2.6.2 (ii) HNCA

Magnetization is passed from ^1H to ^{15}N and then via N- $\text{C}\alpha$ J-coupling to the $^{13}\text{C}\alpha$ and then back again to ^{15}N and ^1H hydrogen for detection (Figure 2.2). The chemical shift is evolved for $^1\text{H}^{\text{N}}$ as well as the $^{15}\text{N}^{\text{H}}$ and $^{13}\text{C}\alpha$, resulting in a three-dimensional spectrum (Kay LE *et al.*, 1990; Farmer BT *et al.*, 1992). Since the amide nitrogen is coupled both to

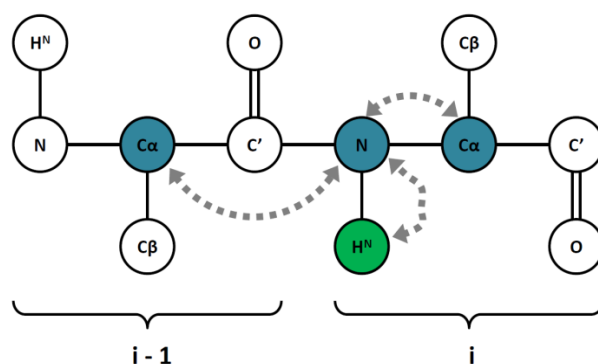


Figure 2.2 | HNCA

the $\text{C}\alpha$ of its own residue (i) and that of the preceding residue (i-1), both these transfers occur and peaks for both $\text{C}\alpha$ are visible in the spectrum. However, the coupling to the directly bonded $\text{C}\alpha$ is generally stronger and therefore these peaks will appear with greater intensity in the spectra. This experiment is useful for backbone assignment when used in conjunction with the HN(CO)CA spectrum.

2.6.2 (iii) *HN(CO)CA*

Magnetization is passed from ^1H to ^{15}N and then to ^{13}CO . From here it is transferred to $^{13}\text{C}\alpha$ and then back via ^{13}CO to ^{15}N and ^1H for detection (Figure 2.3). The chemical shift is only evolved for the $^1\text{H}^{\text{N}}$, the $^{15}\text{N}^{\text{H}}$ and the $^{13}\text{C}\alpha$, but not for the ^{13}CO (Bax A and Ikura I, 1991). This results in a spectrum which is similar to the HNCA, except it is selective for the $\text{C}\alpha$ of the preceding residue (i-1).

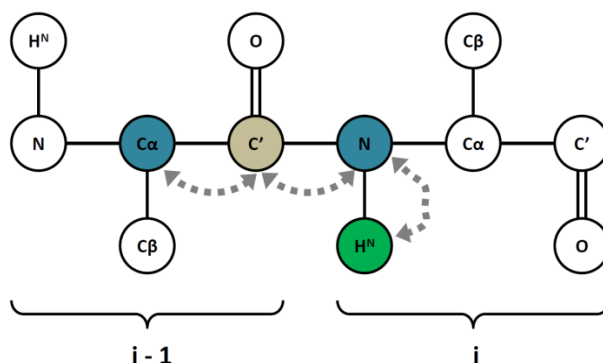


Figure 2.3 | HN(CO)CA

2.6.2 (iv) *HNCACB*

Magnetization is transferred ^1H to ^{15}N and then via N- $\text{C}\alpha$ J-coupling to the $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$. From here it is transferred back first to $^{13}\text{C}\alpha$ and then to $^{15}\text{N}^{\text{H}}$ and to $^1\text{H}^{\text{N}}$ for detection (Figure 2.4; Grzesiek S and Bax A, 1992). Because magnetization is transferred to $^{15}\text{N}_i$ from both $^{13}\text{C}\alpha_i$ and $^{13}\text{C}\alpha_{i-1}$, as a result for each NH group

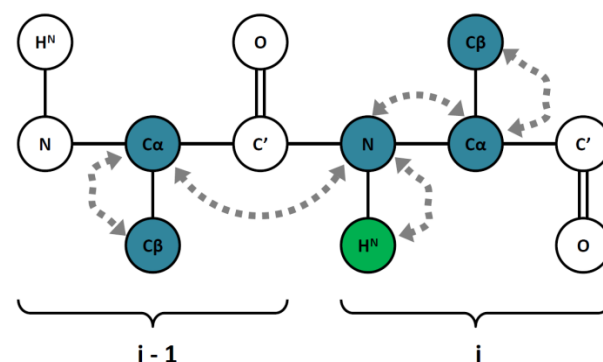


Figure 2.4 | HNCACB

there are two $\text{C}\alpha$ and $\text{C}\beta$ peaks visible, *i.e.* $\text{C}\alpha$ and $\text{C}\beta$ from of its own residue (i) and that of the preceding residue (i-1). The chemical shift is evolved simultaneously on $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$, so these appear in one dimension. The chemical shifts evolved in the other two dimensions are $^{15}\text{N}^{\text{H}}$ and $^1\text{H}^{\text{N}}$. This experiment is particularly useful for backbone assignment when used in conjunction with the HN(CO)CACB spectrum.

2.6.2 (v) *HN(CO)CACB*

Magnetization is passed from ^1H to ^{15}N and then to ^{13}CO . From here it is transferred to $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$, then back via ^{13}CO to ^{15}N and ^1H for detection. (Figure 2.5; Grzesiek S and Bax A, 1992). The chemical shift is evolved simultaneously on $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$, so these appear in one dimension. This results in a spectrum which is similar to the HNCACB, except it is selective for the $\text{C}\alpha$ and $\text{C}\beta$ of the preceding residue ($i-1$).

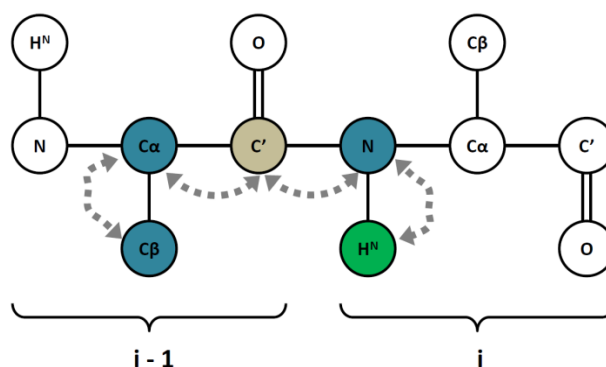


Figure 2.5 | HN(CO)CACB

2.6.2 (vi) *HNCO*

Magnetization is passed from hydrogen to ^{15}N and then selectively to the carbonyl ^{13}C via $^{15}\text{N}^{\text{H}} - ^{13}\text{CO}$ J-coupling (Figure 2.6). Magnetization is then passed back via ^{15}N to hydrogen for detection (Kay LE *et al.*, 1990; Ikura M *et al.*, 1990). The chemical shift is evolved on all three nuclei. This is the most sensitive triple-resonance experiment and is used to obtain CO_{i-1} chemical shifts. This experiment can be useful for backbone assignment when used in conjunction with the HN(CA)CO spectrum.

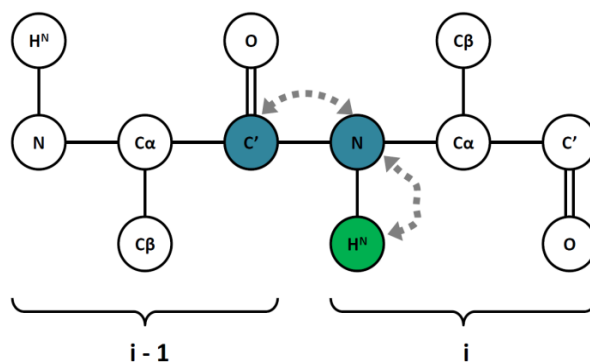


Figure 2.6 | HNCO

2.6.2 (vii) *HN(CA)CO*

Magnetization is transferred from ^1H to ^{15}N and then via N-C α J-coupling to the $^{13}\text{C}\alpha$. From there it is transferred to ^{13}CO via the $^{13}\text{C}\alpha$ - ^{13}CO J-coupling (Figure 2.7). For detection the magnetization is transferred back the same way: from ^{13}CO to $^{13}\text{C}\alpha$, ^{15}N and finally ^1H . The chemical shift is only evolved on ^1H , ^{15}N and ^{13}CO and not on the $^{13}\text{C}\alpha$

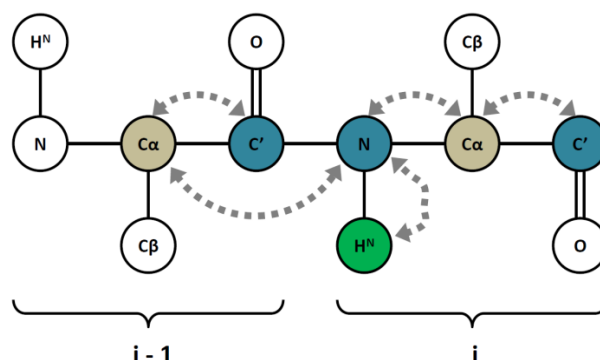


Figure 2.7 | HN(CA)CO

(Clubb RT, Thanabal VT and Wagner G, 1992). Because the amide nitrogen is coupled both to the $\text{C}\alpha$ of its own residue (i) and that of the preceding residue (i-1), transfer to both ^{13}CO nuclei occurs. Thus for each NH group, two carbonyl groups are observed in the spectrum. But because the coupling between N_i and $\text{C}\alpha_i$ is stronger than that between N_i and $\text{C}\alpha_{i-1}$, the CO_i peak generally ends up being more intense than the CO_{i-1} peak.

2.6.2 (viii) $[^1\text{H}, ^1\text{H}, ^{15}\text{N}]$ -NOESY-HSQC

Magnetization is exchanged between all hydrogens using the nuclear overhauser effect (NOE). NOE is a through-space phenomenon used in the study of three-dimensional structure and conformation (Cavanagh J, Fairbrother WJ, Palmer AG, Rance M and Skelton NJ, 2007). This experiment establishes correlations between protons which are physically close to each other (within $5\text{\AA}$) regardless of covalent bond structure. Then the magnetization is transferred to neighboring ^{15}N nuclei and back to ^1H for detection (Marion D, Kay LE, Sparks SW, Torchia DA and Bax A, 1989). This spectrum can be used to identify sequential through-space connectivities by defining protons that are close in space (Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn G and Clore M, 1989).

2.6.3 NMR TITRATIONS

Two salt titrations using standard HSQC experiments were carried out on a Bruker Avance III 500 MHz spectrometer at protein concentrations of 200 and 400 μ M, respectively. Three different titrations with protein/peptide ligands as well as nineteen fragment hit validations using standard HSQC experiments were carried out on a Bruker Avance III 700 MHz spectrometer. The parameters used in each of the titrations were shown in Table 2.2.

Table 2.2 NMR titration parameters on ^{15} N-labeled C3d samples			
Ligand	Titration points	No. of titration set	Remarks
NaCl	50 mM and 100 mM	2	Tested high and low C3d concentrations
CD21 SCR 1-2	0, 25, 50, 75, 100, 200 %*	2	Tested high and moderate ionic strengths [†]
factor H peptide	0, 0.5, 0.75, 1.5, 2.25, 3.0, 4.0 mM	1	-
chemical fragment	0 and 1 mM	19	Screened nineteen potential fragment hits

* Titration points for CD21 SCR1-2 corresponded to amount of ligand added in molar ratio with starting point.

[†] Protein samples were in 20 mM sodium phosphate, pH6.8 at 150 mM NaCl and 400 mM NaCl, respectively.

2.7 CRYSTALLIZATION TRIALS OF RECOMBINANT C3d AND THE C3d/CD21 SCR1-2 PROTEIN COMPLEX

2.7.1 CRYSTALLIZATION AND STRUCTURE DETERMINATION OF RECOMBINANT C3d

Crystals of C3d were obtained following screening of commercial sparse-matrix of crystallization conditions and grown at 4°C, using sitting drop method (0.1 μ l of protein and 0.1 μ l of reservoir buffer in a drop). The protein solution contained 18 mg/ml C3d in 20 mM sodium phosphate, 150 mM NaCl, pH 6.8. The reservoir solution contained either 0.2 M potassium nitrate, 20 w/v polyethylene glycol (PEG) 3350 or 10%w/v PEG 1000, 10% PEG 8000. Native crystals were cryo-protected and frozen in liquid nitrogen for data collection. Diffraction data were collected at either ESRF or Diamond synchrotron and processed with MOSFLM (Leslie AGW and Powell HR, 2007) and SCALA (Evans PR, 2005). Data collection statistics are presented in Chapter VI, Table 6.1. The structure was determined by molecular replacement method using previously determined C3d (PDB: 1C3D) as a search model. Refinement statistics are presented in Chapter VII, Table 6.1.

2.7.2 CRYSTALLIZATION TRIALS OF THE RECOMBINANT C3d/CD21 SCR1-2 PROTEIN COMPLEX

Initial trays were set up with six commercially available 96 well crystallization screens using sitting drop method. Protein samples were concentrated to 450 μ M. The protein solution comprised a mixture of C3d and CD21 SCR1-2 at a molar ratio of 1:1, 1:1.25 and 1:2 in 10 mM HEPES, 150 mM NaCl, pH 6.8. Only a few salt crystals were obtained. The effect of ionic strength and temperature in protein complex formation was further investigated by analytical gel filtration chromatography. All experiments were carried out using a Superdex™ 75 HR 10/30 column (GE Healthcare) previously equilibrated in HEPES buffer. Molecular standards and monomeric C3d sample were included in some runs for calibration. Table 2.3 showed a summary of protein samples used and experimental parameters for the size exclusion chromatography experiments. Subsequently, sixty-four trays were setup with protein solution in 10 mM HEPES, 100 mM NaCl, pH 6.8 as detailed in Chapter VI, Table 6.2. Crystallization trial variables included molar ratio of protein mixture, drop ratio of protein sample and mother liquor and temperature during tray set up.

Table 2.3 | Summary of experimental parameters used in analytical size exclusion chromatography

Molar ratio of C3d:CD21	Concentration of NaCl (mM)	Temperature (°C)
1:1	150	4
1:1.3	50, 100, 150, 250	4
1:1.3	150	25
1:2	150	4

2.7.3 POST CRYSTALLIZATION TREATMENT AND CO-CRYSTALLIZATION TRIALS OF C3d

C3d crystals were subjected to soaking experiments in 15 mM factor H peptide or 20 mM fragment hits in a buffer comprised mother liquor supplemented with 4% DMSO and cryo-protectant, which is typically 20% ethylene glycol. Soaking time varied from 30 mins to 15 hours depending on crystal tolerability. For co-crystallization trials of C3d with factor H peptide or fragment hits, C3d was concentrated to 18 mg/ml in HEPES buffer supplemented with 4% DMSO. Factor H peptide and fragment hits were then added to the protein sample at 15 mM and 20 mM, respectively. The mixture solution was subjected to screening of commercial sparse-matrix of crystallization conditions at 4 and 25 °C, using the sitting drop method.

CHAPTER III

BIOPHYSICAL CHARACTERIZATION OF THE C3d/CD21 SCR1-2 INTERACTION AND STRUCTURE-DRIVEN SITE-DIRECTED MUTAGENESIS

3.1 PROTEIN PRODUCTION OF RECOMBINANT C3d AND CD21 SCR1-2

3.1.1 CHARACTERIZATION OF WILD-TYPE UNLABELED C3d

C3d was expressed in soluble form and purified by two separate ion exchange chromatography steps to isolate the protein from crude cell lysate. C3d, being a negatively charged protein with a theoretical pI of 5.9, allowed efficient purification using a weak cation exchanger carboxymethylase (CM) followed by a strong anion exchanger Mono Q. The sample was subjected to size exclusion chromatography on S75 column (Figure 3.1a) and yielded a single band on the Coomassie stained gel (Figure 3.1b). Mass spectrometry of C3d confirmed the protein exists as a single homogenous species with an observed mass of 34,710 Da (Figure 3.1c). A typical 1D proton NMR spectrum of C3d showed good signal dispersion upfield of 0 ppm and downfield of 8.5 ppm (Figure 1d), which are characteristic of well structured proteins (Rehm T *et al.*, 2002).

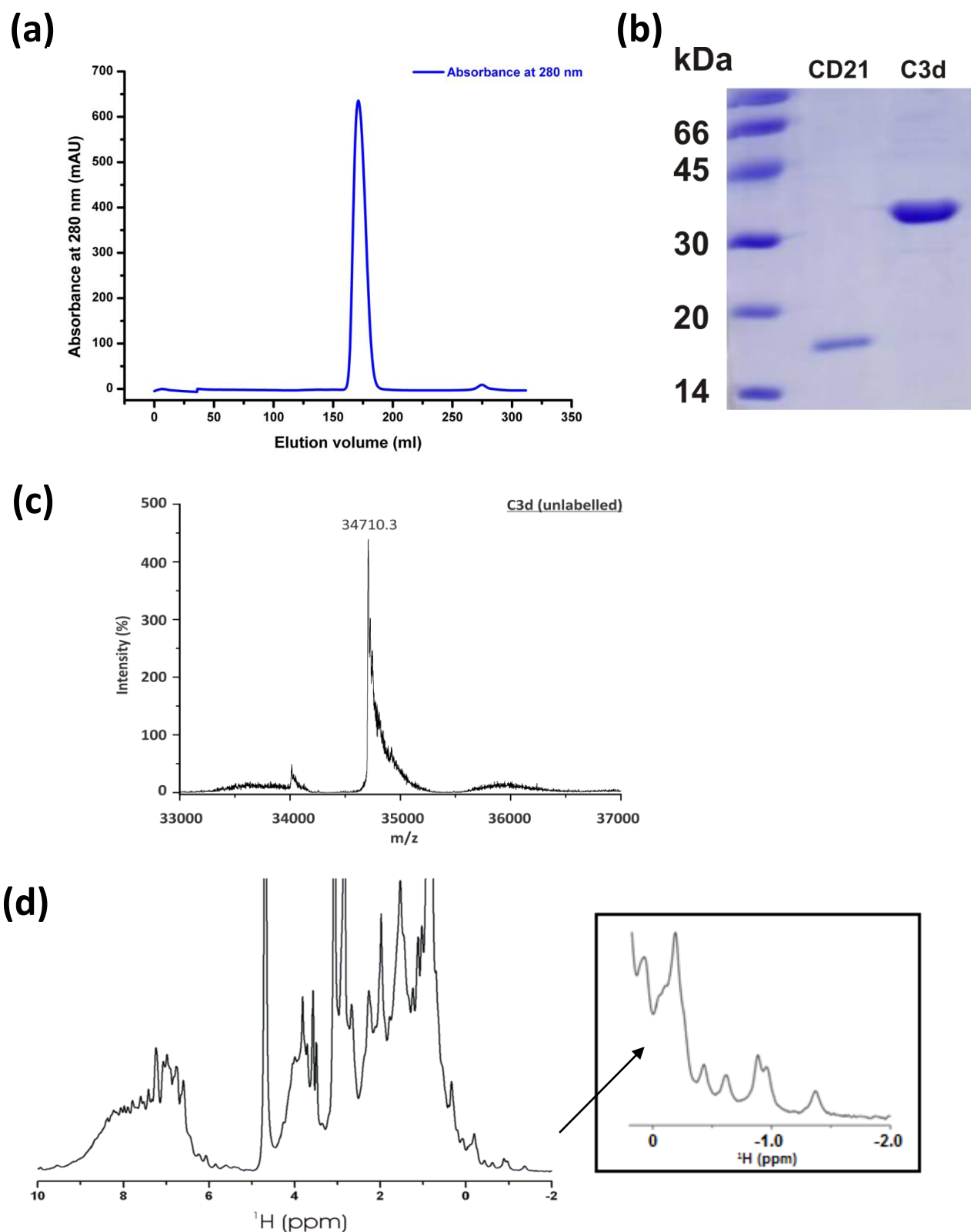


Figure 3.1 | Purification and characterization of wildtype unlabelled C3d

- A gel filtration elution profile of C3d showed a single peak corresponding to a molecular weight of ~ 35 kDa.
- A denaturing polyacrylamide gel showing a single band of the 15 kDa CD21 SCR1-2 and the 35 kDa C3d.
- Molecular mass profile of wild-type unlabeled C3d obtained by maximum entropy processing of the ESI-MS spectrum, indicating an observed molecular weight of 34,710 Da.
- A proton NMR spectrum of C3d collected on a Bruker Avance III 500 MHz spectrometer. The upfield shifted methyl region beyond 0 ppm is highlighted in a black square box.

3.1.2 CHARACTERIZATION OF MUTATED AND ISOTOPE-LABELED C3d

Various C3d mutants were designed to exploit structure-guided interaction mapping between CD21 SCR1-2. One mutant, C3d A17C, comprising an Ala 17 to Cys point mutation, was designed for thiol-coupling on a sensor chip to conduct SPR experiments. All mutations were generated successfully and 1D proton NMR experiments were carried out to confirm structural integrity of C3d mutants (Figure 3.2). Unique upfield resonances from methyl groups packing against aromatic residues in a protein core are indicative of a correctly folded state. None of the mutations in C3d resulted in misfolding of the protein. Isotope labeled C3d was produced for heteronuclear NMR experiments and efficiency of ^{15}N -labeling was checked by mass spectroscopy. For an ^{15}N -labeled sample, the theoretical mass of ^{15}N -labelled C3d is 34,710 Da + 412 (additional mass from ^{15}N) = 35,122 Da. The observed mass of ^{15}N -labelled C3d indicated in the mass spectrum is 35,117 Da (Figure 3.3), *i.e.* 5 Da less than the theoretical mass. Assuming the missing 5 Da comes from incomplete labeling of ^{15}N , the efficiency of ^{15}N incorporation is therefore $(407/412) \times 100 = 98.8\%$.

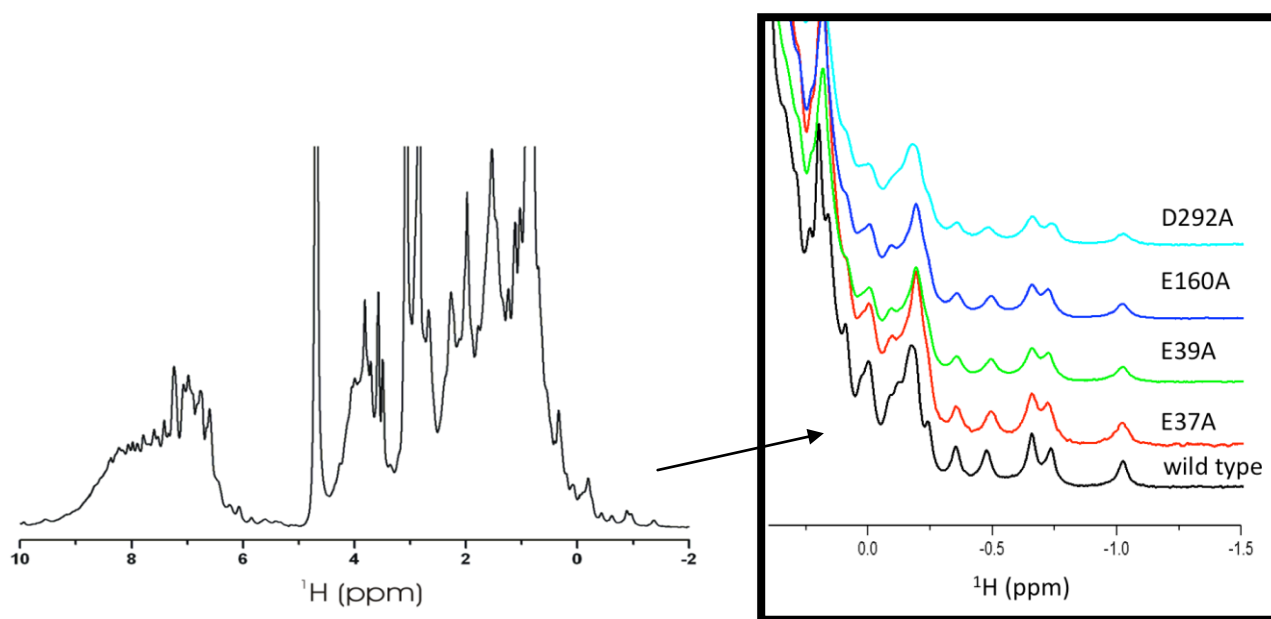


Figure 3.2 | 1D proton NMR spectra of C3d mutant

Examples of 1D proton NMR spectra of some of the C3d mutants used in the thesis. A highly shielded proton, such as a proton of a methyl group packed against an aromatic amino acid with heterocyclic ring systems, can resonate at a very low frequency and hence these protons are usually found upfield of 0 ppm.

The upfield shift methyl resonances are conserved in all C3d mutants, indicating there are no significant structural changes after point mutation and the proteins are well folded.

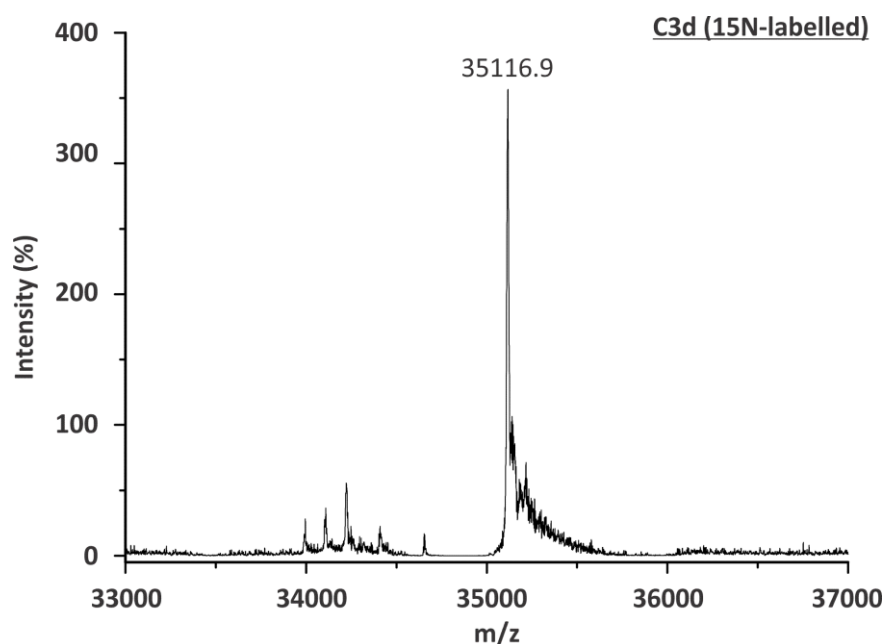


Figure 3.3 | Molecular mass profile of ^{15}N -labelled C3d obtained by maximum entropy processing of the ESI-MS spectrum

The theoretical mass of ^{15}N -labelled C3d is 35,122 Da and the observed mass is 35117 Da. Assuming the missing mass are due to incomplete ^{15}N incorporation, the ^{15}N labeling efficiency is 98.8%.

3.1.3 CHARACTERIZATION OF WILD-TYPE AND MUTATED CD21 SCR1-2

Each SCR domain of CD21 contains two disulfide bonds and the reducing environment of bacterial cytoplasm hinders expression of soluble CD21 SCR1-2 protein from bacterial expression systems. In this thesis, CD21 SCR1-2 has been successfully purified from *E. coli* derived inclusion bodies with a refolding protocol modified from other proteins that are structurally related to CD21 SCR1-2 (Dodd I *et al.*, 1995; Pettigre DM *et al.*, 2006; White J *et al.*, 2004). A recombinant construct of CD21 SCR1-2 fused with a C-terminal His₆ tag allowed efficient purification with Ni-NTA affinity chromatography after refolding. Protein sample was further purified with size exclusion chromatography using a Superdex 75 column to ensure sample purity and homogeneity (Figure 3.4a/Figure 3.1b). A typical yield of 7 mg of purified refolded protein per litre culture was achieved. A 1D proton NMR experiment was carried out to ensure that CD21 SCR1-2 existed as a well structured protein in solution after refolding (Figure 3.4c). Specific CD21 SCR1-2 mutants were generated for site-directed mutagenesis studies and 1D proton NMR experiments were carried out to confirm the structural integrity of these mutant proteins (Figure 3.4d). None of the mutations on CD21 SCR 1-2 used in this thesis resulted in misfolding of the protein.

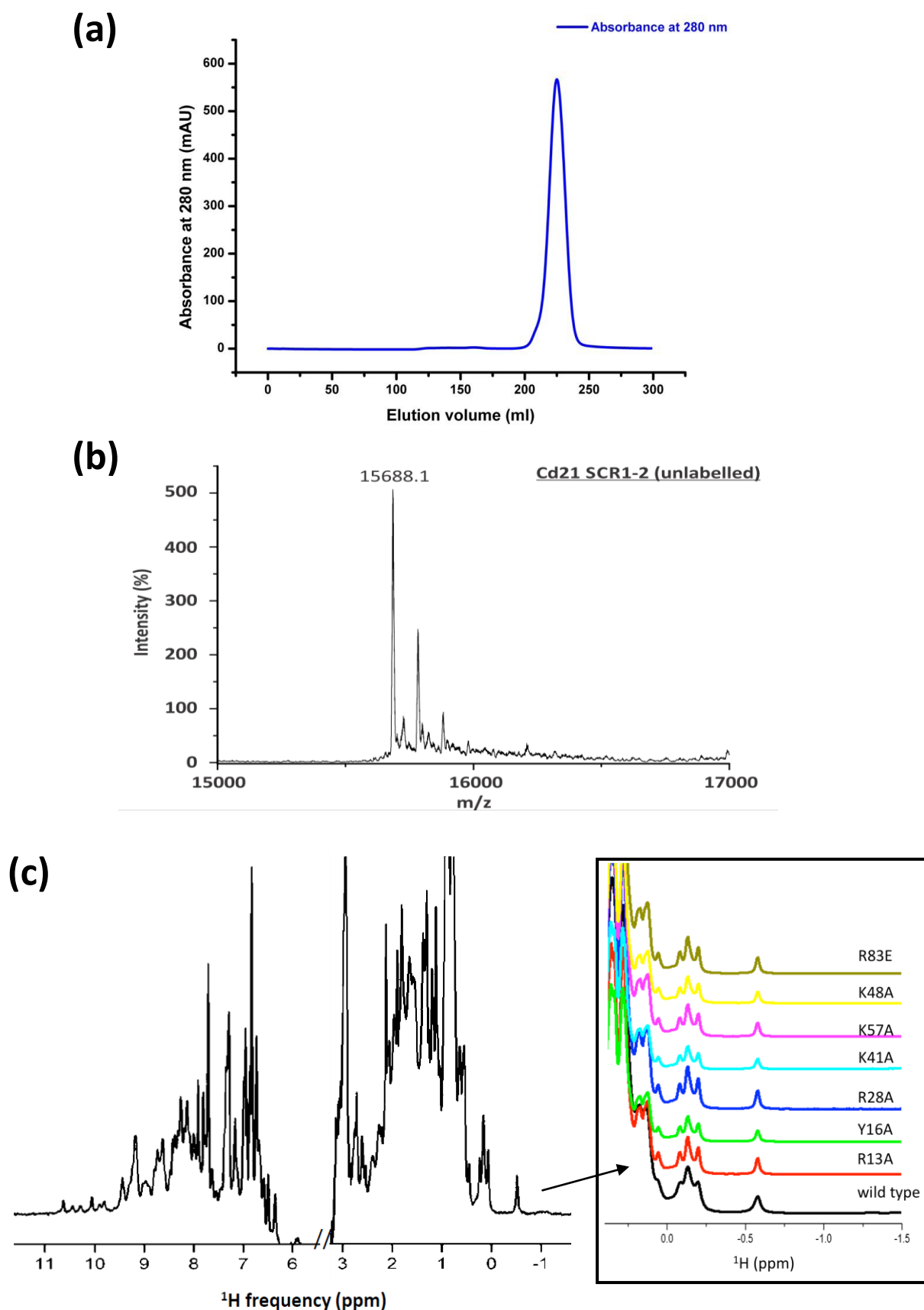


Figure 3.4 | Purification and characterization of unlabelled CD21 SCR1-2

- A gel filtration elution profile of CD21 SCR1-2 showed a single peak corresponding to a molecular weight of ~ 16 kDa.
- A mass spectrum of wild-type unlabeled CD21 SCR1-2 indicating its observed molecular weight of 15688 Da.
- A proton NMR spectrum of CD21 SCR1-2 collected on a Bruker Avance III 500 MHz spectrometer. The upfield shifted methyl region beyond 0 ppm of the wild type and mutant spectra is highlighted in a black square box.

3.2 SPR ANALYSIS ON THE WILD-TYPE C3d AND CD21 SCR1-2 INTERACTION

3.2.1 BRIEF INTRODUCTION TO SURFACE PLASMON RESONANCE (SPR)

SPR is a powerful label-free detection technique that exploits the change of the refractive index of a sensor surface as a function of sample concentration or surface density. The refractive index is dependent on the sample concentration at the interface, allowing binding interactions to be monitored in real time (Van der Merwe PA, 2001). The basic setup of SPR experiments is that one interaction partner is tethered to the sensor surface, which consists of a thin gold film on the flat surface of a prism, while the other is studied in the solute phase. Figure 3.5 shows a schematic representation of an SPR system.

When a beam of plane polarized light passes from material with a high refractive index into a material with a low refractive index, some light is reflected from the interface. If the angle at which the beam strikes the gold film on the surface of the prism is greater than the critical angle, the light will be totally internally reflected. However, the reflection will not be total as some of the energy is lost through interaction with oscillating plasma waves at the surface (surface plasmons) of the metal film. When the wave vector of the incident light matches the wavelength of the surface plasmons, the electrons resonate, resulting in a loss of energy and therefore a reduction in the intensity of the reflected light. Since the amplitude of the wave vector in the

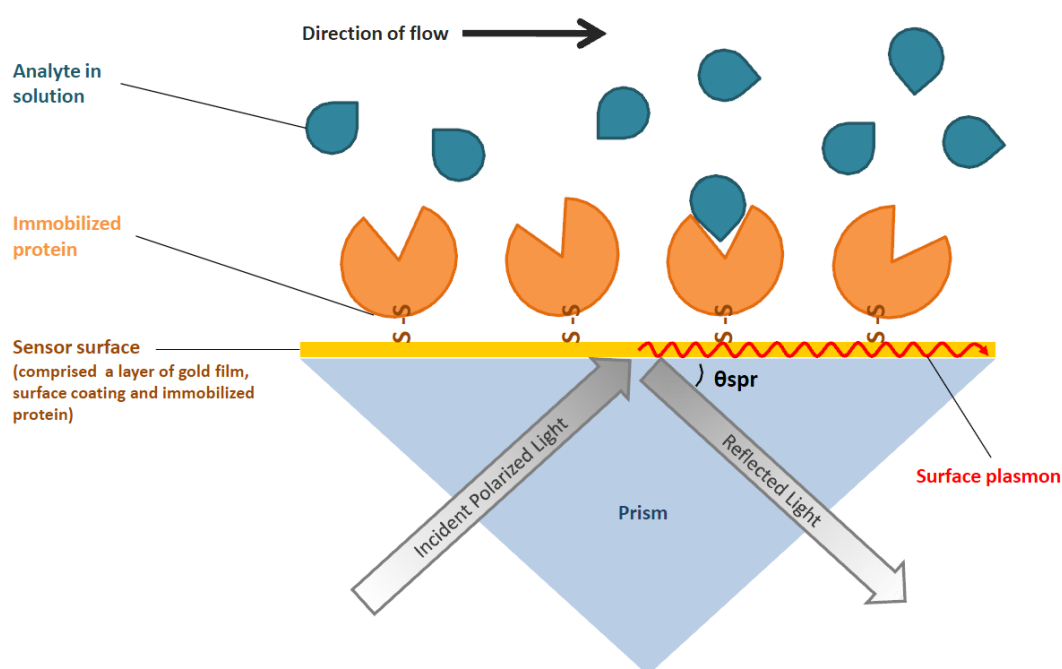


Figure 3.5 | Schematic diagram of an SPR system

Protein target of interest is immobilized on the sensor surface and the ligand is flowed over the sensor surface in solution. The binding of ligand to the immobilized protein results in an increase in the refractive index. The change in refractive index is measured in real time and converted to binding response.

plane of the metallic film depends on the angle at which it strikes the interface, a surface plasmon resonance angle (θ_{spr}) exists at which the energy loss is greatest. This results in a dip in the intensity of reflected light and changes in the angle of the dip are measured in the SPR experiment. An evanescent electrical field associated with the plasma wave travels for a short distance (approximately 300 nm) into the medium from the metallic film. Therefore, the resonant frequency of the surface plasmon (and thus θ_{spr}) depends on the refractive index of the medium close to the sensor chip surface. If the surface is immersed in an aqueous buffer (refractive index of approximately 1.0) and protein (refractive index of approximately 1.33) binds to the surface, this results in an increase in refractive index which is detected by a change in the θ_{spr} . The instrument uses a photodetector array to measure very small changes in θ_{spr} . The change is quantified in resonance units (RUs) with 1 RU equivalent to a shift of 10^{-4} degrees (Van der Merwe, 2001).

3.2.2 BINDING AFFINITY AND KINETICS OF THE C3d/CD21 SCR1-2 INTERACTION

SPR experiments were carried out to investigate the interaction between C3d and CD21 SCR1-2 under low and physiological ionic strength at 25°C. A C3d construct, A17C, was engineered to have a free exposed thiol group spatially on the opposite surface of the CD21 SCR1-2 binding site (Clemenza L and Isenman DE, 2000; Bomb M, 2008). It was immobilized onto a CM4 sensor chip to a density of < 300 RU by thiol coupling (Schuck P, 1997). The interaction between C3d and CD21 SCR1-2 exhibited fast-on, fast-off kinetics irrespective of ionic strength tested (Figure 3.6). “Fast-on, fast-off” kinetics occurs when the association phase reaches

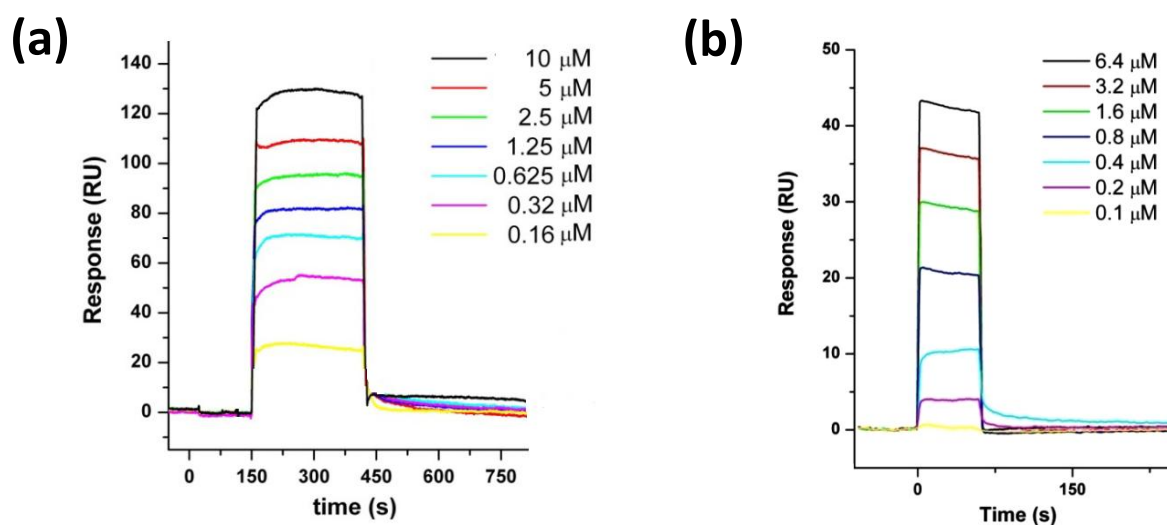


Figure 3.6 | Representative SPR sensorgrams of the wild type C3d/CD21 SCR1-2 interaction

SPR experiments were carried out at an ionic strength of **a)** 50 mM NaCl; **b)** 150 mM NaCl.

equilibrium and the bound complex dissociates at least as fast as the time required to transfer new buffer into the SPR flowwell. It makes equilibrium analysis straightforward but prevents precise measurement of association and dissociation rate constants. The interaction has an affinity of $0.3 \pm 0.06 \mu\text{M}$ at 50 mM NaCl. Increasing the NaCl concentration from 50 mM to 150 mM lowers the affinity approximately 5-fold, to $1.4 \pm 0.21 \mu\text{M}$. The K_D in both cases was estimated by plotting the binding response at equilibrium against the concentration of CD21 SCR1-2 and fitting a Langmuir binding isotherm (Figure 3.7). The Langmuir model, which applies in most binding events, assumes 1) both binding partners are homogenous; 2) the analyte is monovalent; 3) all binding events are independent (Van der Merwe, 2001). Ionic strength plays an important role in the binding affinity of the C3d/CD21 SCR1-2 interaction, as observed in ITC experiments in previous studies (Bomb M, 2008) as well as other molecular dynamic stimulation and solution experiments (Clemenza L *et al.*, 2000; Zhang L *et al.*, 2007; Guthridge JM *et al.*, 2001; Morikis D *et al.*, 2004).

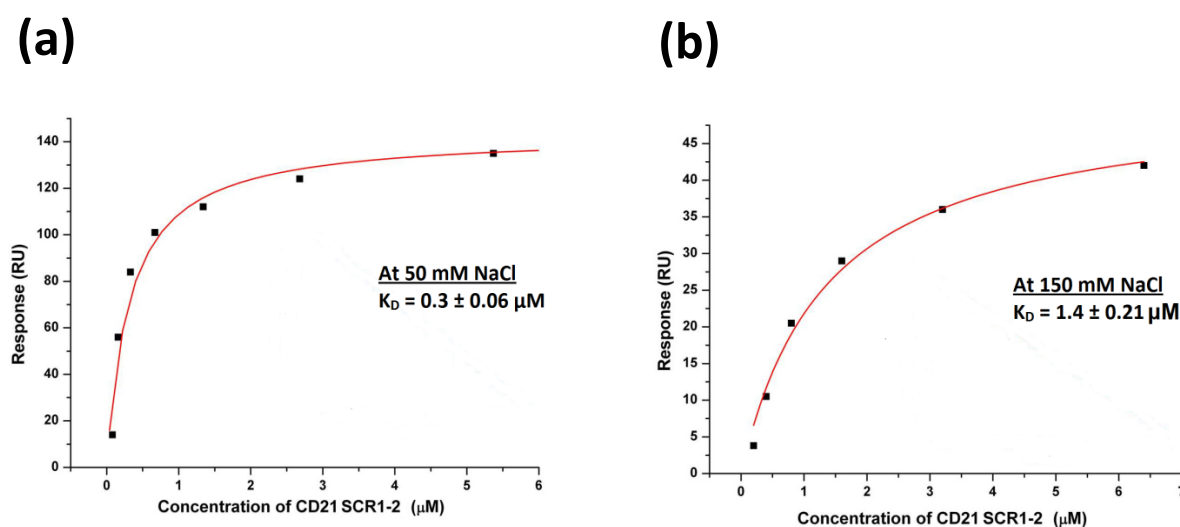


Figure 3.7 | Binding constant determination of wild type C3d/CD21 SCR1-2 interaction

Binding curves of wild type C3d-CD21 SCR1-2 interaction fitted with Langmuir binding isotherm at an ionic strength of **a)** 50 mM NaCl; **b)** 150 mM NaCl

3.3 ITC ANALYSIS ON THE C3d AND CD21 SCR1-2 INTERACTION

3.3.1 BRIEF INTRODUCTION TO ISOTHERMAL TITRATION CALORIMETRY (ITC)

ITC is a calorimetric technique which directly measures the heat evolved or absorbed during molecular association (O'Brien R *et al.*, 2001). The direct thermodynamic observable is the heat associated with a molecular interaction, for instance when a ligand is titrated into a solution containing the macromolecule of interest. The experimental setup is, in principal, dependent on the thermodynamic characteristics of the system of interest, namely the expected binding affinity and the heat effect of the interaction. The shape of experimental binding curve, c , is dependent on the product of the binding constant K_a , the stoichiometry of the interaction and the molar concentration of ligand $[M_T]$ being titrated (Wiseman T *et al.*, 1989):

$$c = K_A \cdot n[M_T]$$

The dimensionless parameter c is crucial for determination of the binding constant. At high C -values ($C > 500$), the affinity of binding partners is very strong and the transition from the unbound state to the saturated complex formed is too sharp, resulting in a rectangular binding isotherm (Figure 3.8). Under these conditions,

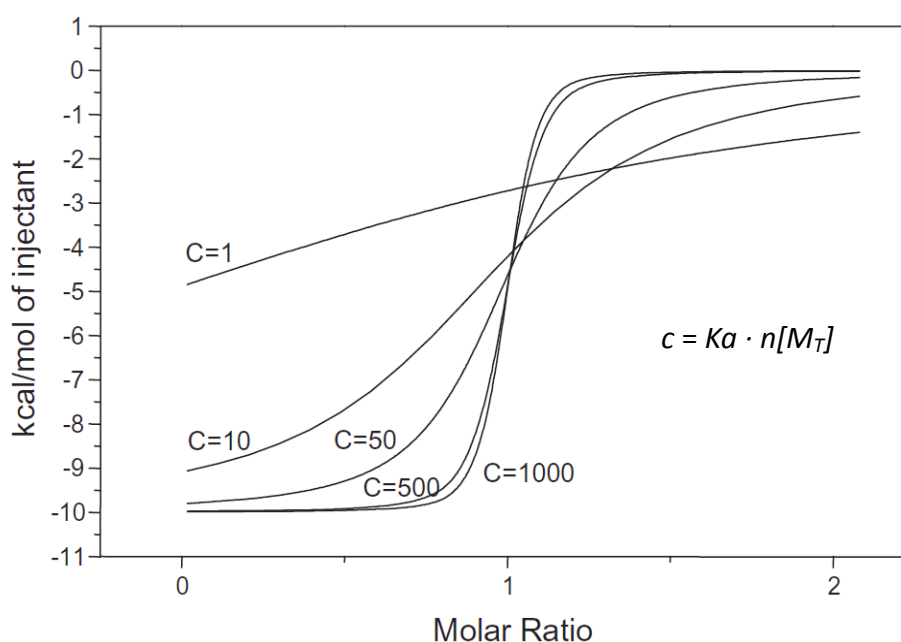


Figure 3.8 Simulated calorimetric binding curves adapted from *Thermodynamics of Protein-Ligand Interactions: History, Presence and Future Aspects* (Perozzo R *et al.*, 2004)

The curves are simulated for several C -values and illustrate the dependence of the shape of the curve on the product of the association constant K_a and the total ligand concentration $[M_T]$.

there are too few points in the transition region necessary to calculate K_A . It has been established that conditions under which $10 < c < 100$ must be satisfied in order to get reliable estimates of the binding affinity (Wiseman T *et al.*, 1989). A low C-value ($c < 10$) corresponds to weak binding and the binding curves appear as a horizontal trace that yields inadequate information for precise determination of K_A (Pierce MM *et al.*, 1999; Velazquez-Campoy A *et al.*, 2004). An ITC instrument consists of two identical cells, a sample and a reference cell, both composed of a highly efficient thermal conducting material. These are surrounded by an adiabatic jacket and cooled by a surrounding water bath (Figure 3.9). The signal monitored by ITC is the differential power applied to the sample cell in order to maintain an equal temperature as the reference cell. The total heat released or absorbed upon an injection of ligand into the cell corresponds to the area under the signal vs. time curve (Figure 3.10).

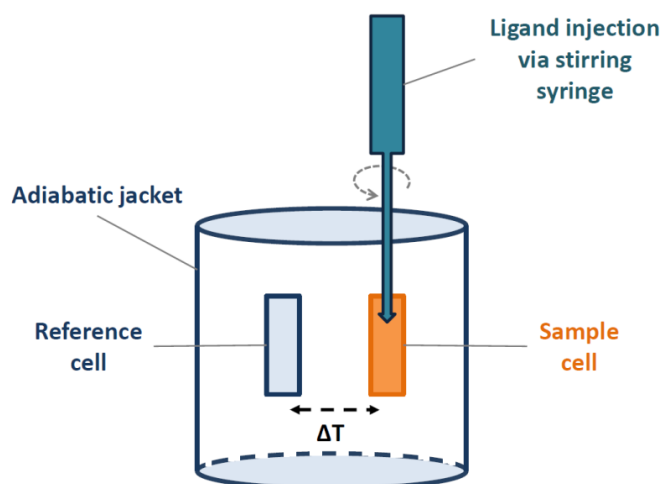


Figure 3.9 Schematic diagram of an ITC instrument

The reference cell is filled with degassed water and serves as a temperature reference to the sample cell for all experiments. The injection syringe contains a small paddle on end act as a stirrer to ensure continual mixing of the injected ligand and protein in the sample cell.

ITC is the most direct method to measure the heat change on macromolecular complex formation. In most cases, a simple reversible association between a macromolecule M and a ligand L is characterized by its binding constant K_A :

$$K_A = [ML] / [M] [L]$$

The observable response on an ITC experiment is the heat change associated with each injection of ligand, which is directly proportional to the total amount of formed complex and the molar enthalpy of binding ΔH (Leavitt S, 2001; Pierce MM *et al.*, 1999):

$$q = V_o \cdot \Delta H \cdot \Delta[ML]$$

where q is the heat associated with the change in complex concentration $\Delta[ML]$,
 ΔH is the molar enthalpy of binding and
 V_o is the reaction volume of the sample cell

Each aliquot of ligand injection in an ITC experiment gives rise to a heat change. However, as the experiment progresses the change in heat diminishes correspondingly as the concentration of unoccupied binding sites in the sample cell becomes depleted. In a single experiment, it is possible to obtain a very precise estimate of the association and dissociation constants (K_A and K_D) as well as the reaction enthalpy (ΔH), and thus the standard Gibbs free energy change (ΔG), the entropy change (ΔS) and the stoichiometry (n) of the interaction by the following thermodynamic expressions (Perozzo R *et al.*, 2004):

$$K_D = 1 / K_A \quad (1)$$

$$\Delta G = -R \cdot T \cdot \ln K_A \quad (2)$$

where R is the gas constant ($1.987 \text{ cal K}^{-1} \text{ mol}^{-1}$) and

T is the temperature of the reaction in Kelvin (K)

$$\Delta G = \Delta H - T \Delta S \quad (3)$$

These thermodynamic parameters can yield important information about a given protein-ligand interaction.

ΔH represents the non-covalent bond energy occurring during the interaction. The measured enthalpy is the resultant value of formation and breakage of many individual bonds in solution. The change in binding enthalpy reflects the loss of hydrogen bonds and van der Waals interactions between protein and solvent, formation of protein-ligand bonds, salt bridges and van der Waals contacts, as well as solvent reorganization at protein surfaces. These individual components may produce either favorable or unfavorable contributions. A negative value of ΔH is energetically favourable. ΔS is associated with the disorder in a system and is predominantly governed by changes in solvation entropy upon ligand interaction. The burial of water-accessible surface area upon binding liberates solvent molecules which contributes often large and positive

change to the total entropy of interaction. Nonetheless, other unfavorable contributions arise from the reduction of rotational degree of freedom around torsion angles of protein and ligand side-chains in the bound complex, as well as the reduction of the number of particles in solutions and their degrees of freedom (Murphy KP *et al.*, 1993; Murphy KP *et al.*, 1994; Murphy KP *et al.*, 1995). A positive value of ΔS is energetically favourable. ΔG determines the direction in which bimolecular binding equilibria will spontaneously proceed, with more negative values of ΔG favouring higher affinity of binding.

3.3.2 THERMODYNAMICS OF THE WILD-TYPE C3d-CD21 SCR1-2 INTERACTION

ITC experiments were carried out in 20 mM Na₂HPO₄, 150 mM NaCl, pH 6.8 at 25°C. A CD21 SCR1-2 sample of 200 μ M was titrated in small aliquots against C3d, until a >3-fold excess was achieved (Figure 3.10a). The interaction has a K_D of 0.92 ± 0.07 μ M and shows the expected 1:1 stoichiometry ($N = 0.96 \pm 0.05$). The binding is predominantly entropic with the following thermodynamic contributions: $\Delta G = -34.4 \pm 2.7$ kJ/mol, $\Delta H = +48.2 \pm 0.9$ kJ/mol, $T\Delta S = +82.6$ kJ/mol. The thermodynamic signature of the C3d/CD21 SCR1-2 interaction suggests hydrophobic interactions and/or strong electrostatic contributions play a dominant role in the formation of the protein-protein complex.

Hydrophobic interactions are characterized by large favourable entropy changes alongside small and frequently unfavourable enthalpy changes (Lin Z *et al.*, 1994; Srinivas V *et al.*, 1999). This phenomenon can be explained by the increase in entropy which arises from the release of highly ordered water molecules surrounding apolar surfaces of the two interacting macromolecules. Solvent-exposed apolar groups on free macromolecules do not fulfill the hydrogen bonding capacity of water. Consequently, the existing hydrogen bond network of bulk water is reorganized and results in the formation of hydration shells, in which water molecules surrounding apolar groups make stronger bonds with each other than those of bulk water. Upon formation of macromolecular complexes, these water molecules return to the bulk where they form weaker bonds, a process which is entropically favourable and slightly endothermic (Eftink M and Biltonen R, 1980). Similarly, association of charged groups to macromolecules often is entropically driven and enthalpically unfavourable (Matulis D *et al.*, 2000). Solvent-exposed polar groups interact strongly with water molecules to

form solvation shells, which stabilize charged groups found on protein surfaces. The entropic advantage of electrostatic interactions arises from the displacement of water molecules or counter ions solvating the free charged groups and thus increasing the freedom of particles in solution. Altogether a positive entropy change is a strong indication that water molecules have been released from the surface of macromolecules upon complex formation (Jelesarov I and Bosshard HR, 1999).

It has been reported that the interaction between C3d and CD21 SCR1-2 exhibits a strong ionic strength dependency (Clemenza L *et al.*, 2000; Zhang L *et al.*, 2007; Guthridge JM *et al.*, 2001; Morikis D *et al.*, 2004; Bomb M, 2008). The ITC experiment was repeated in 20 mM Na₂HPO₄, 400 mM NaCl, pH 6.8 at 25°C (Figure 3.10b). Table 3.1 shows a summary of the effect of ionic strength on the C3d-CD21 SCR1-2 interaction. By plotting log (K_A) versus log (ionic strength), the data results in a straight line ($R=0.98$) with a slope of -1.6 units (Figure 3.11). A strong negative linear regression of the graph reflects that electrostatics play a dominant role in the C3d-CD21SCR1-2 interaction. The result is in excellent agreement with the SPR-derived log-log plot (Bomb M, 2008), in which the gradient of the slope was also -1.6 units.

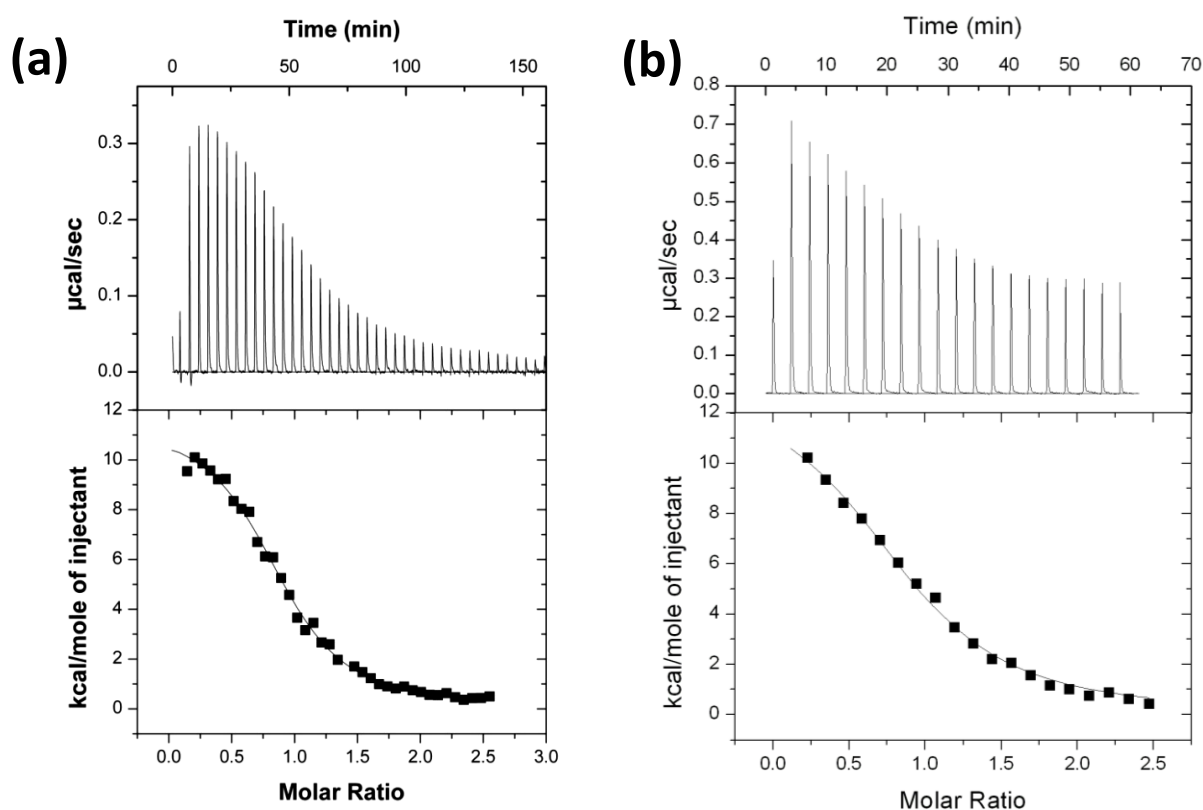


Figure 3.10 ITC analysis for the C3d/CD21 SCR1-2 interaction at physiological and high ionic strength

C3d titrated against CD21 SCR1-2 in a) 20 mM NaH₂PO₄, 150 mM NaCl, pH 6.8; b) 20 mM NaH₂PO₄, 400 mM NaCl, pH 6.8. The top part of the diagrams represents the raw data per injection and the bottom part represent the integrated areas of each injection plotted against the molar ratio of ligand titrated.

Table 3.1 | The effect of ionic strength on the C3d/CD21 SCR1-2 interaction

Ionic strength of NaCl (mM)	K_D (μ M)	ΔG (kJ/mol)	ΔH (kJ/mol)	$-T\Delta S$ (kJ/mol)
50 [‡]	0.19 $\pm$ 0.02	-38.4 $\pm$ 3.8	+58.5 $\pm$ 0.5	96.8
150	0.92 $\pm$ 0.07	-34.4 $\pm$ 2.7	+48.2 $\pm$ 0.9	82.6
400	4.89 $\pm$ 0.07	-30.3 $\pm$ 2.5	+55.1 $\pm$ 0.8	85.4

[‡] The experiment was performed by Bomb M, 2008.

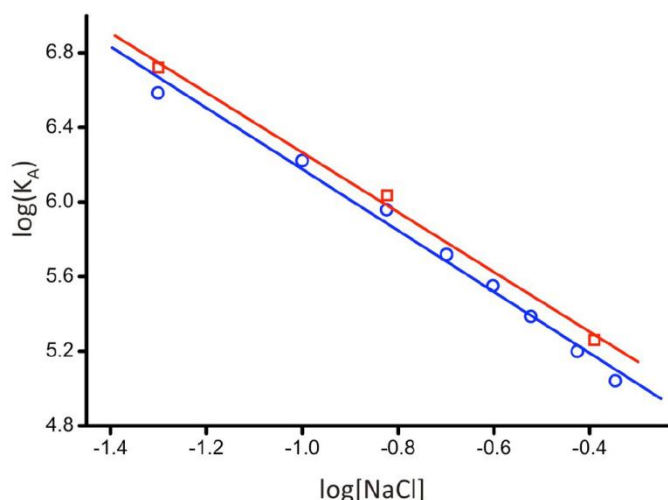


Figure 3.11 | The effect of ionic strength on the C3d/CD21 SCR1-2 interaction

Data from ITC measurements were plotted in red. Data from previously reported SPR measurements (Bomb M, 2008) were plotted in blue. Data from both measurements resulted in negative linear regression with a slope of -1.6.

3.3.3 MUTATIONAL STUDIES ON THE C3D-CD21 SCR1-2 INTERACTION

To gain a more complete picture of the C3d/CD21 SCR1-2 interaction, a series of C3d and CD21 SCR1-2 mutants was generated. These mutants were designed largely based on independent flow cytometric analysis of mutations targeting the C3d/CD21 SCR1-2 interaction (Clemenza L and Isenman DE, 2000; Hannan JP *et al.*, 2005), as well as interfacial residues in the crystal structure of the C3d/CD21 protein complex solved by Szakonyi G *et al.*, 2001. Because the NMR data described by Bomb M as well as Kovacs JM *et al.* clearly indicated a binding site different to that observed in the crystal structure of the C3d/CD21 SCR1-2 complex, an extensive site-directed mutagenesis studies of the two possible interaction surfaces were performed. Fourteen CD21 SCR1-2 mutants were produced and characterized. These include seven residues (R13, Y16, R28, S32, K41, K48 and K57) and a double mutant (R28/S32) from the NMR-defined binding site (Bomb M, 2008), two mutations (R83 and G84) and a double mutant (R83/G84) from the site of complex formation observed in the x-ray structure (Szakonyi G *et al.*, 2001) and one additional mutant (R122) to serve as a

negative control. From the C3d protein, seventeen residues were mutated: three amino acids from the crystallographic site (L116, E117 and N170), ten residues from an alternate site predicted from the earlier mutagenesis studies (D36, E37, E39, E42, E160, D163, E166, E167, E37/E39 and D163/E166) (Clemenza L and Isenman DE, 2000; Bomb M, 2008), and three residues as controls (K232, E265 and E292). The C3d mutagenesis defined a ring of acidic residues surrounding a deep hydrophobic groove as critical to CD21 SCR1-2 binding.

Through site-directed mutagenesis, a number of alanine substitutions were introduced to wild-type C3d and CD21 SCR1-2 to identify their functional ligand-binding surfaces and to determine the energetic contributions of specific residues. Changes in the binding affinities on mutation were measured using ITC and these results were presented all together in Table 3.2. The changes in binding energetics from the site-directed mutagenesis were mapped onto the C3d and CD21 SCR1-2 surface and summarized in Figure 3.12.

Table 3.2 | Affinities and thermodynamic parameters of C3d and CD21 SCR1-2 mutants

C3d	K_D (μ M)	ΔG (kJ/mol)	$\Delta\Delta G$ (kJ/mol)	ΔH (kJ/mol)	$-T\Delta S$ (kJ/mol)
wildtype	0.19±0.02	-38.3±0.2	-	+58.5±0.5	96.8
C17A	0.20±0.02	-38.2±0.1	+0.1	+53.0±0.3	91.2
D36A [‡]	0.59±0.03	-35.5±0.3	+2.8	+28.2±0.4	63.7
E37A	0.80±0.05	-34.8±0.2	+3.5	+46.3±0.3	81.1
E39A	0.59±0.06	-35.5±0.2	+2.8	+36.5±0.5	72.0
E37/E39A [‡]	2.72±0.10	-31.7±0.2	+6.6	+38.4±0.4	70.2
E42A [‡]	0.11±0.03	-39.7±0.2	-1.4	+49.8±0.6	89.5
L116A	0.19±0.03	-38.3±0.4	+0.0	+47.6±0.7	85.6
E117A [‡]	0.22±0.04	-38.0±0.4	+0.3	+48.1±0.6	86.1
E160A	0.75±0.05	-34.9±0.2	+3.4	+38.1±0.7	73.0
D163A	1.15±0.09	-33.8±0.2	+4.5	+31.3±0.6	65.2
E166A [‡]	0.96±0.03	-34.3±0.1	+4.0	+24.3±0.4	58.6
E167A [‡]	2.16±0.14	-32.3±0.2	+6.0	+35.7±0.2	68.0
D163/166A [‡]	4.03±0.13	-30.8±0.1	+7.5	+25.9±0.5	49.4
N170A	0.16±0.03	-38.7±0.4	-0.4	+44.8±0.8	83.6
K232A	0.24±0.02	-37.8±0.2	+0.5	+47.4±0.3	85.2
E265A	0.27±0.03	-37.5±0.3	+0.8	+58.8±0.7	96.2
D292A [‡]	1.73±0.08	-32.9±0.3	+5.4	+40.6±0.3	73.3
CD21 SCR1-2	K_D (μ M)	ΔG (kJ/mol)	$\Delta\Delta G$ (kJ/mol)	ΔH (kJ/mol)	$-T\Delta S$ (kJ/mol)
wildtype	0.19±0.02	-38.3±0.2	-	+58.5±0.5	96.8
R13A	3.60±0.30	-31.1±0.2	+7.2	+40.8±1.6	71.9
Y16A [‡]	3.90±0.35	-30.9±0.2	+7.4	+7.8±0.6	38.7
R28A [‡]	0.61±0.08	-35.5±0.3	+2.8	+24.1±0.3	60.6
S32E	0.50±0.07	-36.0±0.2	+2.3	+15.4±0.6	51.3
R28A/S32E	1.01±0.10	-34.2±0.2	+4.1	+21.9±0.8	56.1
K41A [‡]	3.75±0.05	-31.0±0.3	+7.3	+25.6±0.2	50.1
K41E	no binding	-	-	-	-
K48A [‡]	0.43±0.03	-36.3±0.2	+2.0	+13.6±0.2	49.9
K57A [‡]	1.50±0.20	-33.2±0.3	+5.1	+15.5±0.4	48.7
R83A [‡]	0.21±0.02	-38.1±0.2	+0.2	+51.2±0.7	89.3
R83E [‡]	0.57±0.06	-34.1±0.3	+4.2	+20.2±2.3	90.9
G84Y	0.32±0.03	-37.0±0.2	+1.3	+47.2±1.8	84.2
R83E/G84Y [‡]	0.28±0.04	-37.4±0.3	+0.9	+53.3±0.8	90.7
R122E [‡]	0.19±0.02	-38.0±0.2	+0.3	+56.8±0.7	94.8

[‡] New mutations that were characterized in this thesis.

Among the CD21 SCR1-2 mutants, independent or combinational mutations on R83 and G84 had no major effect towards the binding affinity and thermodynamics of the C3d/CD21 SCR1-2 interaction (Table 3.2). These two residues were shown to mediate binding between C3d and CD21 SCR1-2 in the crystal structure of the protein complex solved by Szakonyi G *et al.* (Figure 1.3, Chapter I). Replacing charged residue Arg 83 with a neutral alanine and likewise flexible Gly 84 with a bulky aromatic tyrosine showed no interruption to protein binding. This was further exemplified with combinational mutations of R83E/G84Y, whereby replacing original residue types to completely opposite types merely increased the K_D to 280 ± 40 nM. These two residues made no energetically important contributions to the free energy of binding as indicated by $\Delta\Delta G$ values of <1 kJ mol⁻¹. However, mutating residues from CD21 SCR1 located on the C3d binding surface indicated by previous NMR chemical perturbation studies (Bomb M, 2008; Kovacs *et al.*, 2009) had profound effects on the binding energetics to C3d. Five residues in CD21 SCR1, R13, Y16, R28, K41 and K57, made substantial energetic contributions to the free energy of binding as indicated by $\Delta\Delta G$ values of $+2.8$ to $+7.4$ kJ mol⁻¹ (Table 3.2). These residues were clustered in a hydrophobic core surrounded by positively charged residues in SCR1. A point mutation on CD21 SCR2, namely R122E, was designed to serve as a negative control and confirmed that only specific positively charged and hydrophobic residues located in SCR1 appear to mediate binding to C3d.

Correspondingly in the C3d protein, amino acids N170, L116 and E117, made no energetic contributions to the free energy of binding to CD21 SCR1-2 as evidenced with negligible $\Delta\Delta G$ values. These residues were located in the observed crystallographic protein-protein interface in the crystal structure solved by Szakonyi G *et al.*. A number of residues from the previously identified "acidic pocket" appeared to form the functional binding epitope for CD21 SCR1-2 (Figure 3.12). Mutation on these negatively charged residues, with the exception of E42, increased the K_D by 3 to 21-fold with $\Delta\Delta G$ values of $+2.8$ to $+7.5$ kJ mol⁻¹ (Table 3.2). Three point mutations located spatially away from the proposed CD21 SCR1-2 binding site, namely A17C, K232A and E265A were used as negative controls. These control mutants confirmed that only specific negatively charged residues in the acidic pocket on C3d appeared to make electrostatic contributions towards binding to CD21 SCR1-2.

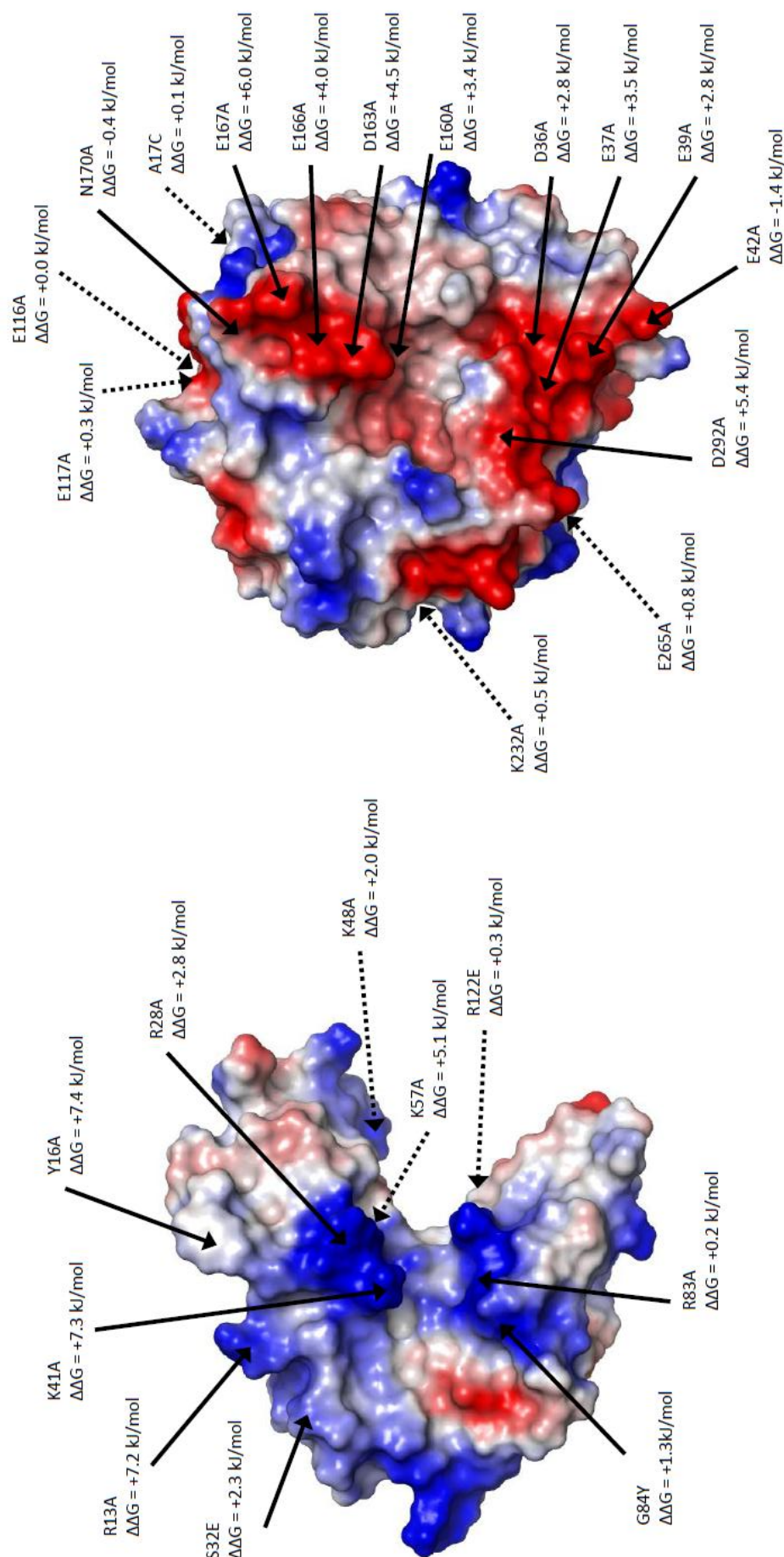


Figure 3.12 | A summary of microcalorimetric analysis on the binding energetics from site-directed mutagenesis studies on C3d and CD21 SCR1-2.

Left = CD21 SCR1-2; Right = C3d. The proteins were represented in molecular surfaces which were coloured according to electrostatic charge. Positively charged region were blue and negatively charged region were red.

3.3.4 IDENTIFICATION OF SALT BRIDGE INTERACTIONS BY DOUBLE MUTANT CYCLES

In the case of protein-protein interactions where electrostatics contribute strongly to binding energy, construction of double mutant cycles is another mutational approach that can be used to establish whether the mutated residues interact with each other (Wells JA, 1990; Schreiber G *et al.*, 1995; Frisch C *et al.*, 1997). The concept of double mutant cycle is illustrated in Figure 3.13. If the effects of mutating two residues are additive then the mutated residues are considered to act independently of each other. However, if the change in ΔG values of the two mutants ($\Delta\Delta G$) is significantly smaller than the sum of the two components then the mutated residues can be predicted to interact with each other.

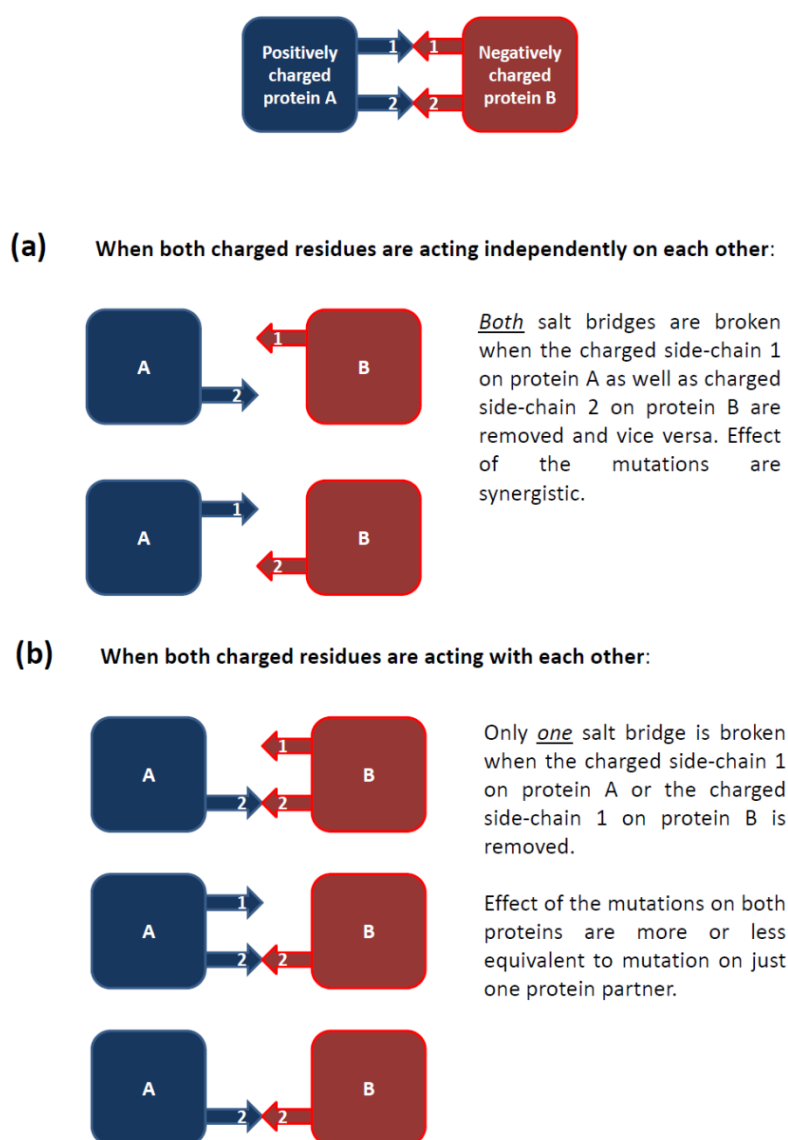


Figure 3.13 | Diagram illustrating the double mutant cycle

Double mutant cycles were constructed between a subset of electrostatic mutations from CD21 SCR1-2 (wildtype, R13A, R28A and K57A) and C3d (wildtype, E37A, E39A, D163A, E166A, N170A and D292A). The results are summarized in Table 3.3. A number of pairs of mutants appear to abrogate the interaction, indicating that the effect of these residue pairs are additive. In contrast, CD21 SCR1-2 R13A showed no additive effect with either C3d E37A or C3d E39A (Table 3.3), suggesting that E37 and E39 interact with each other, most likely by forming a salt bridge in the bound complex. Similarly, CD21 SCR1-2 K41A also shows no additive effect with C3d D163A, suggesting that these residues also form a salt bridge in the bound complex. As expected, in all cases the C3d N170A construct behaves very much like wildtype C3d in these double mutant interactions, confirming that this residue is not involved in CD21 SCR1-2 binding.

By integrating all of the data from our solution studies, a model of the C3d/CD21 SCR1-2 protein complex representing physiologically relevant binding interfaces is generated in combination with molecular docking studies. Results from the microcalorimetric analysis provided binding restraints to the online protein-protein docking server ClusPro (Comeau SR *et al.*, 2004; Kozakov D *et al.*, 2010) and crystal structures of C3d (PDB:1C3D) and CD21 SCR1-2 (PDB:1LY2) were supplied for computing the docking of two protein structures. Docking of ligand and receptor computationally requires a precise representation of the molecular energetic and a powerful algorithm to search for potential binding modes. ClusPro's algorithm is especially engineered to address protein-protein docking and was registered as the best docking sever for protein complexes based on structures of the component proteins in the EMBL-EBI Critical Assessment of Predicted Interactions (CAPRI) experiments in 2009 (<http://structure.bu.edu/news/2010/01/03/both-our-group-and-our-cluspro-server-were-best-performers-latest-rounds-capri>). ClusPro utilizes rigid body global search based on the Fast Fourier Transfor (FTT) correlation approach that evaluates the energies of billions of docked modes on a virtual grid. It samples by rotating the ligand structure with 70,000 rotations and translate each rotation in x,y,z directions relative to the receptor on a virtual grid. The first one thousand rotation and translation combinations with low binding energy score were then chosen. A bound ligand-receptor complex typically has a lower energy than its constituent parts. The aim of molecular docking studies is to identify low energy binding modes of a ligand within an active site of a receptor using known structural information. ClusPro then

clusters the top one thousand ligand positions within a 9Å Cα rmsd radius to find a region of attraction around local energy minima. The docking server typically generates thirty low energy binding modes in the end. Each computationally suggested binding mode is manually inspected to identify one that either resembles the experimentally observed binding mode, or is in agreement with experimental data defining the binding interfaces. Based on the results from microcalorimetric analysis of the double mutant cycle, a low energy binding mode of the docked C3d/CD21 SCR1-2 structure has been chosen and shows a striking shape and charge complementarity (Figure 3.14). The predicted salt bridges formed between CD21 SCR1-2 R13A and C3d E37A/E39A as well as CD21 SCR1-2 K41 and C3d D163 provides two anchor points of the interaction surface. The interaction between the oppositely charged surfaces also allows the hydrophobic residues between the basic clusters of CD21 SCR1-2 to come together with the hydrophobic cavity on C3d. The hydrophobic interaction surface is thus shielded from the surrounding solvent. Whereas previous binding studies indicated a number of disparate residues to be involved in the binding interaction (Hannan JP *et al.*, 2005; Kovacs *et al.*, 2009), our current model indicates the two proteins interact over a single contiguous site. Finally, the site of antigen attachment to C3d (formed by Cys17 and Gln20) is on the opposite face of C3d, exposed to the solvent and easily accessible to antigen as required for functional activity of C3d (Fearon DT *et al.*, 1995; Nagar B *et al.*, 1998).

Table 3.3 | A summary of double mutant cycle analysis

CD21 SCR1-2	C3d	K_D (μ M)	ΔG (kJ/mol)	$\Delta\Delta G$ (kJ/mol)	ΔH (kJ/mol)	$-T\Delta S$ (kJ/mol)
wildtype	wildtype	0.19	-38.3	-	58.5	96.8
wildtype	E37A	0.78	-34.8	3.5	46.3	81.1
wildtype	E39A	0.59	-35.5	2.8	36.5	72.0
wildtype	D163A	1.15	-33.8	4.5	31.3	65.2
wildtype	E166A	0.96	-34.3	4.0	15.1	49.4
wildtype	N170A	0.16	-38.7	-0.4	44.8	83.6
wildtype	D292A	1.73	-32.9	5.4	40.6	73.3
wildtype	E160A	0.75	-34.9	3.4	38.1	73.0
R13A	wildtype	3.60	-31.1	7.2	40.8	71.9
R13A	E166A	No binding	-	-	-	-
R13A	N170A	3.10	-31.4	6.9	14.9	46.3
R13A	E37A	3.47	-31.1	7.2	11.3	42.5
R13A	E39A	3.51	-31.1	7.2	10.7	41.8
R28A	wildtype	0.61	-35.5	2.8	24.6	60.6
R28A	E37A	1.17	-33.8	4.5	23.8	57.6
R28A	E166A	No binding	-	-	-	-
R28A	N170A	0.53	-35.8	2.5	22.4	58.2
R28A	D292A	6.90	-29.4	8.9	25.8	55.2
K57A	wildtype	1.50	-33.2	5.1	15.5	48.7
K57A	E37A	No binding	-	-	-	-
K57A	E166A	No binding	-	-	-	-
K57A	N170A	1.73	-32.8	5.4	5.9	38.7
K57A	D292A	No binding	-	-	-	-
K41A	wildtype	3.75	-31.1	7.2	31.5	62.6
K41A	E160A	No binding	-	-	-	-
K41A	D163A	3.22	-31.4	7.0	3.22	96.1

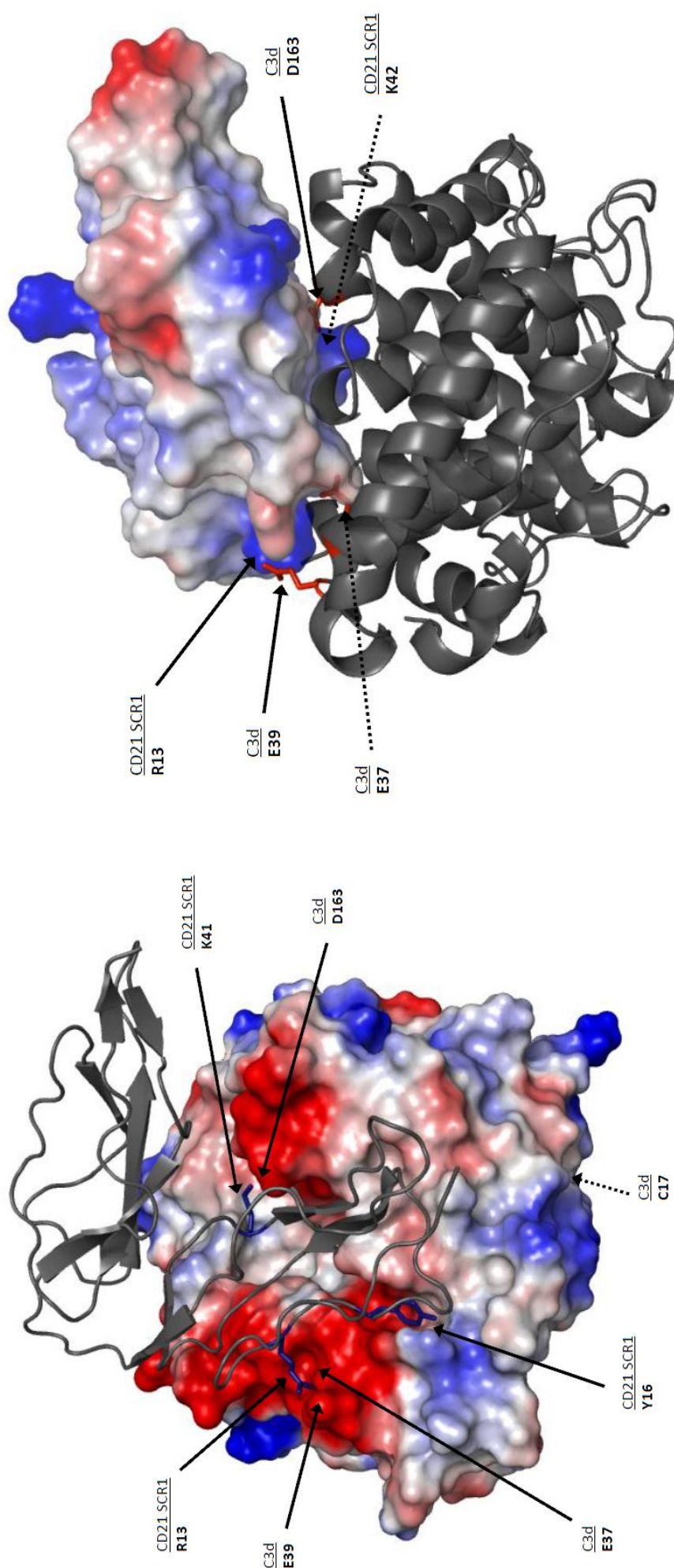


Figure 3.14 A docking model of the C3d/CD21 SCR1-2 protein complex generated by ClusPro

Results from the microcalorimetric analysis were used as binding restraints. *Left* = CD21 SCR1-2 is represented in cartoon ribbons and C3d is represented in molecular surfaces which are coloured according to electrostatic charge; *Right* = CD21 SCR1-2 is represented in surfaces and C3d in cartoon ribbons. Key residues that were predicted to form salt bridges in the bound complex are indicated, as well as other residues to help illustrate the binding orientation of the proteins.

3.4 DISCUSSION

SPR and ITC were used as complementary approaches in characterizing the protein-protein interaction between C3d and CD21 SCR1-2. The association of C3d and CD21 SCR1-2 has a thermodynamic signature which is atypical for protein-protein interactions. It is entropically driven and the favourable entropy term evolved is likely derived from increased freedom of water and/or counter ions upon formation of the protein-protein complex (Jelesarov I and Bosshard HR, 1999). The interaction thermodynamics are consistent with a recognition event driven by complementary electrostatics and hydrophobicity. Supporting this hypothesis for the mechanism of interaction are the thermodynamic analyses of the site-directed mutants. The mutation of either charged or hydrophobic residues at the binding interface decreases both the overall affinity of the interaction and the favourable contribution of the entropic term to binding energy (Table 3.2). A graph of entropy ($-T\Delta S$) versus enthalpy (ΔH) was plotted based on the thermodynamics data from forty three protein-protein interactions gathered by Stites WE, 1997 (Figure 3.15). Approximately 59% of the protein-protein interactions were found to be enthalpically favoured. There were eighteen cases where protein complex association were driven by entropy and enthalpically opposed. By plotting the thermodynamic parameters of the C3d/CD21 SCR1-2 interaction on to the same graph, it was clear that the interaction sits on the far end of the entropy favoured region, suggesting an unusual mechanism of interaction for C3d and CD21 SCR1-2. A recent study seemed to suggest a two-site binding model for the CR2/C3d interaction, consisting of a high affinity site on CD21 SCR1 and a low affinity site predominantly formed by amino acids on SCR2 (Kovacs *et al.*, 2009). While that NMR-based study identified some of the amino acids on CD21 SCR1-2 crucial for its interaction with C3d, notably R13, Y16, A22, R28, Y29 and C31, it also implicates a number of other amino acids. The present study in this chapter clearly indicates that these additional residues do not form part of the binding interface but were artefacts arising from non-specific ionic interactions that occur at low ionic strength between the two oppositely charged proteins. Furthermore, the ITC data presented in this study indicate a 1:1 stoichiometry of the interaction and gives no suggestion of a second, low affinity binding site.

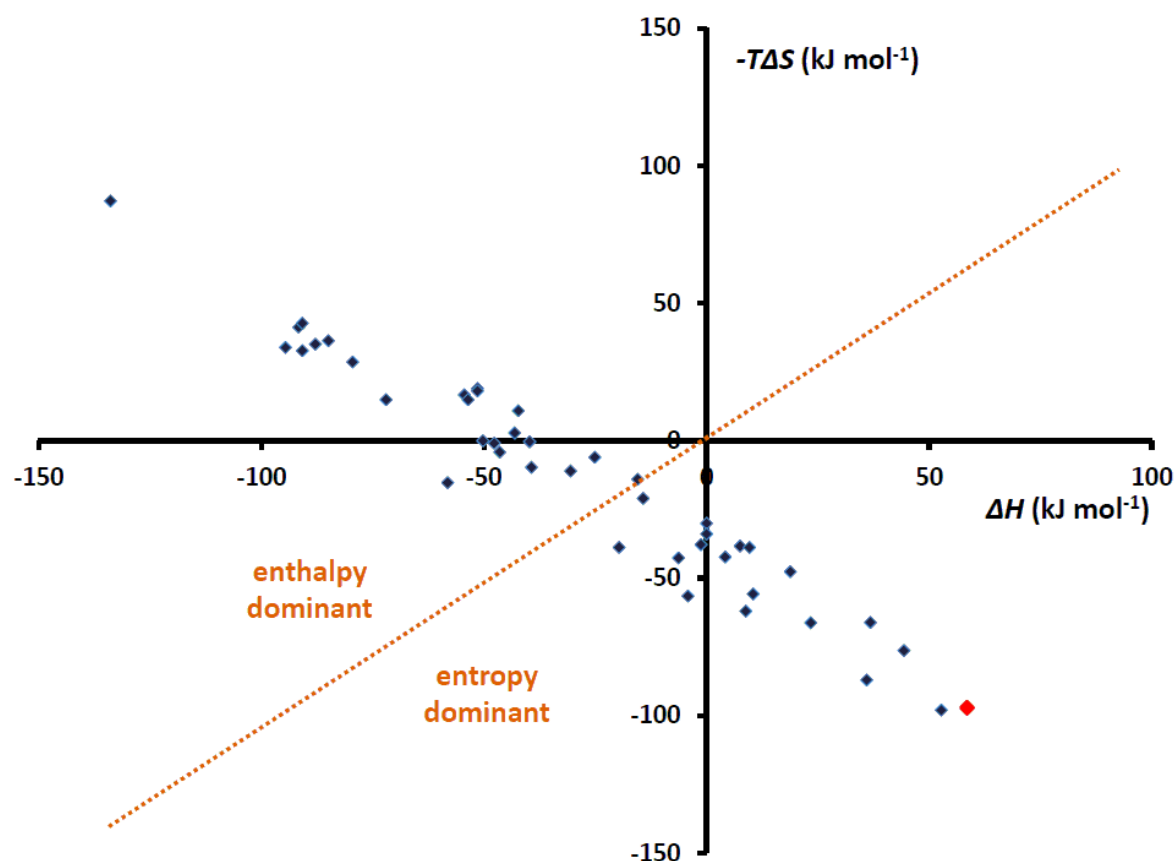


Figure 3.15 | A summary of the thermodynamics on enthalpy and entropy data from forty three protein-protein interactions gathered by Stites WE, 1997

Diagonal orange dotted line represents $\Delta H = -T\Delta S$. Interactions above the line have a dominant, favourable, enthalpic contribution to ΔG , whereas those below the line have a dominant entropic contribution. The red dot represents the C3d/CD21 SCR1-2 interaction.

The work in this chapter has given insight into the unusual binding mechanism of the C3d/CD21 SCR1-2 interaction and delineated the binding interfaces on both proteins. It was hoped that these findings might guide the design of therapeutic small molecule inhibitors that could interfere with the interaction between CD21 SCR1-2 and C3d-antigen complexes. The surface identified on C3d is a particularly promising site for the development of small molecule inhibitors. The binding site presents a deep hydrophobic groove, and docking studies suggest that the hydrophobic residues on the CD21 SCR1-2 surface could not insert deeply enough to make full use of the potential binding surface. Consequently, a well designed small molecule could make optimal use of the groove, achieving better affinity and specificity than the native protein ligand.

CHAPTER IV

BACKBONE ASSIGNMENT OF C3d BY NMR

4.1 INTRODUCTION TO NMR SPECTROSCOPY OF LARGE PROTEINS

C3d is the final degradation product of complement component C3 and represents the minimal full binding activity ligand for CD21 SCR1-2 (Kalli KR *et al.*, 1991; Clemenza L and Isenman DE, 2000). The molecular weight of C3d is 35 kDa and the crystal structure of human C3d shows a topology of twelve α -helices spiralling from outside to inside (Nagar B *et al.*, 1998; PDB:1C3D). There are six parallel helices in the core of C3d surrounded by another six anti-parallel helices to the core. Since the crystal structure of the C3d-CD21 SCR1-2 protein complex solved by Szakonyi G *et al.* appears not physiologically relevant (PDB:1GHQ), solution NMR is used as an allied biophysical method to reassess the C3d-CD21 SCR1-2 interaction as described in Chapter III. Our own laboratory have successfully assigned the NMR spectra of CD21 SCR1-2 (15 kDa) and demonstrated the C3d binding site is a contiguous surface on the SCR1 domain of CD21, which agrees with mutagenesis studies (Clemenza L and Isenman DE, 2000; Hannan JP *et al.*, 2005; Bomb M, 2008; Chapter I, Figure 1.4). It was

hoped that NMR studies on C3d would offer complementary binding information to understand the protein-protein interaction between C3d and CD21 SCR1-2. The work should provide clarification on their precise binding modes from the perspective of both proteins with residue-specific information in solution.

Historically the application of protein NMR spectroscopy is limited to study of small proteins of less than 25 kDa due to technical issues (Güntert P, 1998; Gerhard W and Wüthrich K, 1999). High quality NMR spectra with good sensitivity and spectral resolution are the absolute fundamentals in protein NMR studies. These requirements are more difficult to achieve with increasing molecular weight. Limiting factors of the amenability of NMR studies to proteins above 25 kDa are broad peaks and spectral overcrowding and low sensitivity of high-dimensional NMR experiments for large proteins (Wagner G, 1993; Kay LE and Gardner KH, 1997). Nevertheless, instrument and experimental developments in the last decade have facilitated a number of NMR studies of larger proteins up to 183 kDa (Religa TL *et al.*, 2010). The RCSB Protein Data Bank contains 8042 solution NMR structures of proteins as of January 2012 (<http://www.pdb.org>). The size distribution of these deposited structures shows the typical size of proteins amenable to NMR studies is still predominantly below 30 kDa (Table 4.1). Hence, C3d (35 kDa) is considered a challenging target for routine NMR studies even with state-of-the-art technology and instruments.

Table 4.1 | Molecular weight distribution of solution NMR structures deposited in the RCSB Protein Data Bank as of January 2012

Molecular weight of protein (kDa)	Number of structure deposited
< 10.0	3844
10.0 - 19.9	3577
20.0 - 29.9	475
30.0 - 39.9	65
40.0 - 49.9	42
50.0 - 59.9	6
60.0 - 69.9	2
70.0 - 79.9	4
80.0 - 89.9	3
90.0 - 99.9	1
> 100	5

In NMR experiments, a specially-developed pulse sequence of electromagnetic wave at radiofrequency is applied to a protein sample and magnetization energy is transferred between NMR-active nuclei, and observed using free induction decay (FID). Pulse sequences can be optimized to select and record NMR signals evolved from specific types of connections between nuclei, such as magnetization transfer through chemical bonds or through space (within about 5Å) irrespective of the bonding network. Through-bond experiments are used to assign unique chemical shifts to specific nuclei (*i.e.* fingerprints) of a protein sample, whereas through-space experiments are usually used to generate distance restraints needed for protein structure calculation. The magnetization energy transfer efficiency is dependent on molecular weight and the rate of signal decay scales with the size of the protein molecule. Large proteins have slower tumbling rates in solution and a more rapid decay of NMR signals, termed spin relaxation. The relaxation of NMR-active nuclei is caused by the local magnetic fields that each NMR-active nucleus (a spinning nucleus act like a bar magnet) helps to generate. These local magnetic fields fluctuate as the NMR-active nuclei are attached to protein molecules that tumble in solution. The fluctuating magnetic fields restore the equilibrium distribution of magnetization and therefore obliterate the magnetization energy that is meant to be transferred along the protein chain. In addition, NMR pulse sequences often use long decays for the necessary coherence magnetization transfer steps in experiments and therefore have much lower sensitivity to signal evolved from NMR-active nuclei of large proteins. As a result, NMR spectra of large proteins suffer line broadening effect, where chemical shifts (or peaks) appear less intense and more spread out. Furthermore, increased molecular weight introduces more NMR-active nuclei and therefore higher complexity is seen in a given spectrum due to extensive chemical shift overlap. The range of proteins suitable for NMR studies is usually limited to 25 kDa. Fortunately, modern NMR spectrometers with higher magnetic fields (*e.g.* ≥ 700 MHz) and cryogenic probes have significantly improved intrinsic sensitivity and peak separation for the spectra of large proteins (Wüthrich K, 1998). The development of multinuclear and multidimensional pulse sequences for NMR experiments, coupled with uniform or specific isotopic labelling of proteins, are becoming vital strategies to extend the size limit of protein NMR beyond 30 kDa. Some of these approaches were used to aid NMR studies of C3d and are discussed in this chapter accordingly.

4.2 NMR BACKBONE RESONANCE ASSIGNMENT OF C3d

4.2.1 SAMPLE OPTIMIZATION

Heteronuclear single-quantum correlation (HSQC) experiment is frequently used to monitor ligand binding and to map binding sites between molecules in a complex. In a ^1H , ^{15}N -HSQC experiment, the NMR signals are spread effectively into a second dimension to reduce spectral degeneracy. When a pulse sequence is applied to a protein sample, backbone amides of the protein generate chemical shifts according to their proton and nitrogen frequencies in a given chemical environment (Bodenhausen D and Ruben DJ, 1980; Kay LE, Keifer P and Saarinen T, 1992). Hence, a ^1H , ^{15}N -HSQC spectrum is unique to a particular protein and the chemical shifts evolved from backbone amides are extremely sensitive to the surrounding chemical environment. All amino acids, except proline, are associated with a set of chemical shifts (*i.e.* peak) in the HSQC spectrum. In addition, certain amino acids with nitrogen-containing side-chains also generate additional peaks that are commonly found in their respective defined regions in the spectrum. For instance, the sidechain of tryptophan contains an additional NH bond and thus generates an additional peak that is usually located in a region downfield of 10 ppm in the HSQC spectrum. Similarly, sidechain of asparagine and glutamine each contain an NH_2 group that generates two additional peaks which can be easily identified based on their identical nitrogen shifts around 112 – 115 ppm. Although the sidechain of lysine also contains an NH_3^+ group, additional peaks are not expected in the HSQC spectrum when the experiment is carried out at pH 6.8 as the protons are in fast exchange.

Figure 4.1 shows a set of ^1H , ^{15}N -HSQC spectra of C3d collected from a uniformly ^{15}N -labeled C3d sample on a Bruker Avance III 500 MHz spectrometer equipped with a cryoprobe at 25 °C. The theoretical number of peaks, including sidechains of all tryptophan, asparagine and glutamine, expected in the ^1H , ^{15}N -HSQC spectra of C3d is 371. Inherent problems of NMR studies on large proteins, such as extensive chemical shift overlaps and low spectral resolution, are manifested in the ^1H , ^{15}N -HSQC spectra of C3d recorded on a low field strength instrument. Nonetheless these low resolution spectra are still be able to provide information on the optimization of experimental conditions (*e.g.* protein concentration and salt content of NMR sample) for

subsequent experiments. Phosphate buffer without salt, or with as little salt content as possible, is typically used for NMR samples as the presence of salts significantly diminishes sensitivity for most NMR experiments (Kovacs H *et al.*, 2005; Flynn PF *et al.*, 2000, Kelly AE *et al.*, 2002; Voehler MW *et al.*, 2006). Low concentration of NaCl (*e.g.* 50 mM and no more than 100 mM) is often used to improve sample solubility, especially for proteins that are prone to aggregation. High salt buffer increases the ionic strength and the dielectric conductivity of the NMR sample, which leads to dissipation of the radiofrequency power used in NMR pulse sequences and emergence of ring currents in the sample tube. The sensitivity gain offered by instrument's hardware like a cryoprobe can be lost when the salt concentration of the sample becomes too high, and the loss of sensitivity will become more severe at high field strengths, *e.g.* 900 MHz or above (Horiuchi T *et al.*, 2005). The initial experimental condition used was a ^{15}N -labeled C3d sample at 400 μM in 20 mM NaH_2PO_4 , 50 mM NaCl, pH6.8 (Figure 4.1a). The ^1H , ^{15}N -HSQC spectrum of C3d collected under these conditions display extremely poor resolution, with few resolvable peaks in non-crowded regions of the spectrum. Increasing salt concentration from 50 mM to 100 mM improves spectra resolution moderately by reducing chemical shifts overlap (Figure 4.1b). Correspondingly, increased ionic strength also enhances spectra resolution at lower protein concentration (Figure 4.1c-d). The ^1H , ^{15}N -HSQC spectrum obtained from a sample of 200 μM ^{15}N -labeled C3d in 20 mM NaH_2PO_4 , 100 mM NaCl, pH6.8 shows a good number of individually resolvable chemical shifts. Additional chemical shifts also emerge, which are otherwise not observable under conditions with low salt content or high protein concentration. Observations from these spectra indicates that C3d is prone to aggregation at low ionic strength. Increasing protein concentration above 200 μM offers no gain in spectra resolution for ^1H , ^{15}N -HSQC experiments. Nonspecific protein aggregation in homogeneous sample increases the apparent molecular size of C3d and causes severe line broadening effects in HSQC spectra. Ionic strength and protein concentration dependent aggregation of C3d has been reported in a separate analytical ultracentrifugation studies, in which C3d has been shown to exist in monomer-dimer and monomer-trimer equilibria in 50 mM NaCl, but as a monomer only in 137 mM NaCl (Li K *et al.*, 2008).

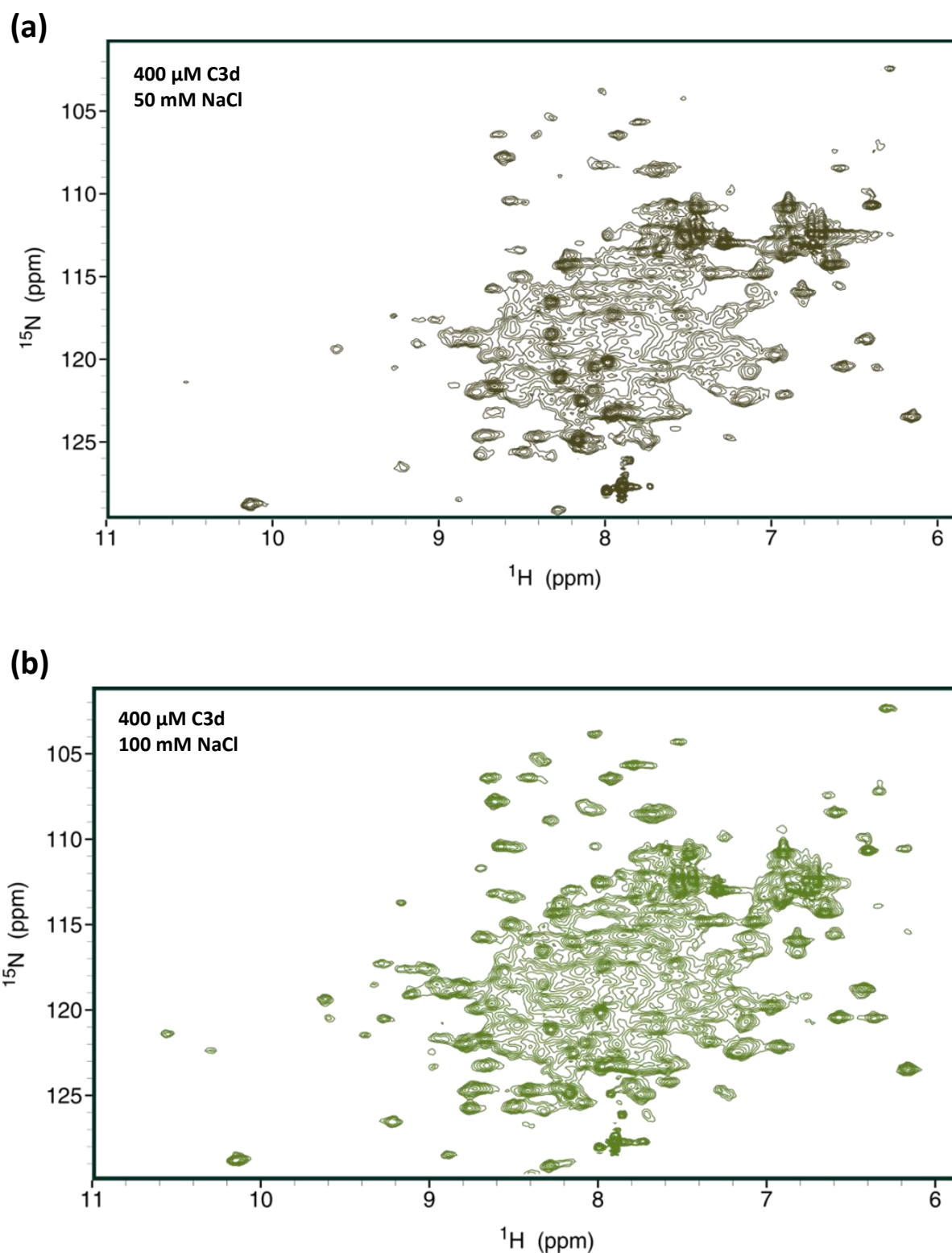


Figure 4.1 ^1H , ^{15}N -HSQC spectra of C3d collected on a Bruker Avance III 500 MHz spectrometer at 25 $^{\circ}\text{C}$

- (a) The spectrum was collected with a uniformly ^{15}N -labeled C3d sample of 400 μM in 20 mM NaH_2PO_4 , 50 mM NaCl, pH 6.8. The overall spectral resolution is very low largely due to the inherent high molecular weight of C3d.
- (b) The spectrum was collected with a uniformly ^{15}N -labeled C3d sample of 400 μM in 20 mM NaH_2PO_4 , 100 mM NaCl, pH 6.8. Doubling the salt concentration improves overall spectral resolution.

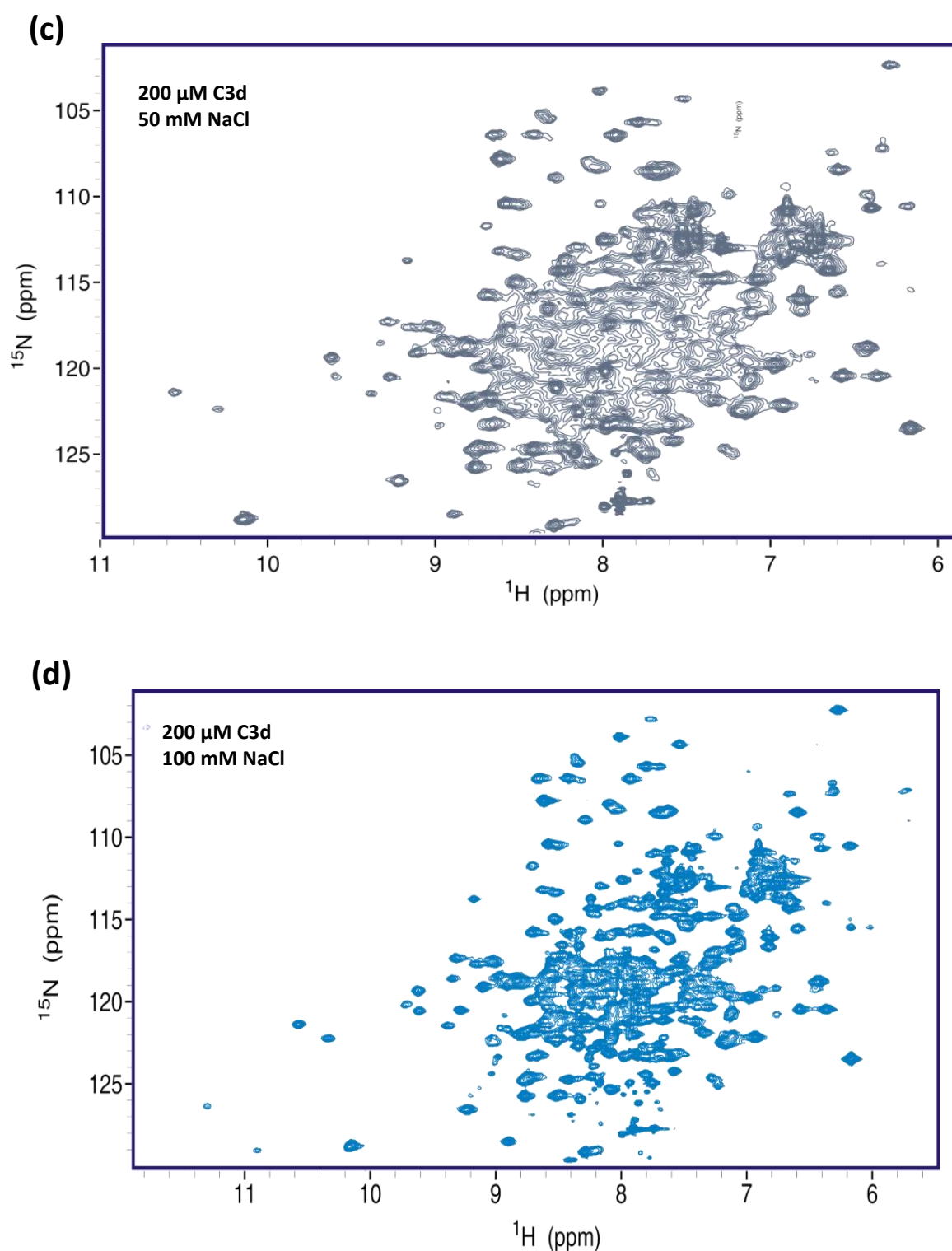


Figure 4.1 ^1H , ^{15}N -HSQC spectra of C3d collected on a Bruker Avance III 500 MHz spectrometer at 25 °C

- (c) The spectrum was collected with a uniformly ^{15}N -labeled C3d sample of 200 μM in 20 mM NaH_2PO_4 , 50 mM NaCl, pH 6.8. Reducing protein concentration by a half also improves overall spectral resolution.
- (d) The spectrum was collected with a uniformly ^{15}N -labeled C3d sample of 200 μM in 20 mM NaH_2PO_4 , 100 mM NaCl, pH 6.8. All together, lowering protein concentration and raising the salt content greatly improves spectral resolution so that a great number of peaks are observed even at 500 MHz.

Moderate ionic strength coupled with lower protein concentration appear to minimize aggregation of C3d and offer adequate sensitivity to HSQC experiments. The tested experimental conditions provide a good foundation to explore triple-resonance (*e.g.* three dimensional) experiments, which are commonly used for the sequential backbone amide assignment of large proteins. Constructed from two-dimensional (*e.g.* HSQC) experiments, the triple-resonance experiments effectively split the NMR signals into an additional ^{13}C dimension to split infinitive spectral degeneracy. The incorporation of deuterium (^2H) in ^{15}N , ^{13}C -labeled proteins for triple-resonance experiments has been established to improve sensitivity and resolution of NMR spectra (Gardner KH and Kay LE, 1998). As mentioned in the previous paragraph, local fluctuating magnetic fields generated by NMR-active nuclei have a deteriorative effect in magnetization energy transfer during NMR experiments. The size of these local magnetic fields generated by NMR-active nuclei is proportional to the gyromagnetic ratios (γ) of the nuclei that produce them. The local magnetic fields generated by protons are four-fold larger than those generated by carbons and thus triple resonance experiments often fail for large proteins of 25 kDa or greater. The gyromagnetic ratio of a deuteron (^2H) is approximately 6.5 times smaller than that of a proton (Gardner KH and Kay LE, 1998) and therefore substituting deuterons for carbon-bound protons remove a large amount of the fluctuating local magnetic fields. The ^2H substitution slows down the relaxation of ^{13}C and in turn increases sensitivity and resolution of NMR experiments that rely on magnetization transfer through ^{13}C nuclei, such as backbone amide triple-resonance experiments. Figure 4.2 shows a ^1H , ^{15}N -TROSY-HSQC spectrum of C3d collected from a uniformly ^2H , ^{15}N , ^{13}C -labeled C3d sample on a Bruker Avance III 700 MHz spectrometer equipped with a cryoprobe at 25 °C using a transverse relaxation-optimized spectroscopy (TROSY) pulse sequence. The protein concentration was at 400 μM in 20 mM NaH_2PO_4 , 150 mM NaCl, pH6.8. At this point spectral resolution and sensitivity improves enough to observe chemical shifts from individual backbone amides. The number of chemical shifts expected in a ^1H , ^{15}N -TROSY-HSQC spectrum of C3d is 301* and the number of chemical shifts observed in the spectrum is 300.

*371 peaks expected in a standard HSQC of C3d – chemical shifts from sidechains of 35 Asn and Gln (35×2) = 301

TROSY-based experiments are developed to extend NMR studies of large proteins beyond 25 kDa at high magnetic field strengths (Pervushin K *et al.*, 1997). The implementation of TROSY pulse sequence into triple-resonance experiments introduce several-fold sensitivity gain for ^2H , ^{15}N , ^{13}C -labeled proteins and approximately two-fold sensitivity gain for ^{15}N , ^{13}C -labeled proteins (Salzmann M *et al.*, 1998). In heteronuclear correlation experiments such as HSQC, the NMR signal of each nucleus of the ^{15}N - ^1H in backbone amides is originally split into two components and therefore a total of four components are observed. This quadruplet signal is routinely averaged into a single signal by broad-band decoupling techniques to simplify the spectrum (Wider G, 1998; Bax A and Grzesiek S, 1993; Freeman R and Kupce E, 1997). Individual components of this quadruplet signal possesses different relaxation rates (*i.e.* some components experienced a more rapid relaxation than the others) and the differential relaxation effects are very prominent to large proteins studied at high magnetic field strengths. Averaging the quadruplet signal in standard HSQC experiments at high fields thus intensifies line broadening effects that are deteriorating to overall spectral quality. In TROSY-based experiments, the quadruplet signal is not decoupled and only the most slowly relaxing component of each quadruplet signal is recorded. A ^1H , ^{15}N -TROSY-HSQC spectrum of a large protein thus shows narrower linewidth and higher sensitivity than a standard HSQC spectrum. As mentioned above, the rate of NMR-active nucleus relaxation is decreased by the substitution of deuterons for protons to minimize the amplitude of local fluctuating magnetic fields. These local fields also affect the quadruplet signal and TROSY-based experiments further benefit from deuteron substitution in protein sample, as only NMR signals experienced the smaller net field derived from subtraction of the local fields are selected for detection. Furthermore, conventional TROSY pulse sequence schemes filter out the NH_2 chemical shifts of the Gln and Asn sidechain, which normally appear in the spectral region 112 - 115 ppm along the nitrogen frequency and 6.7 - 8.7 ppm along the proton frequency in HSQC spectrum (Pervushin K *et al.*, 1997). There are twenty-two glutamine and thirteen asparagines residues in the C3d construct and each would normally generate two additional peaks from the sidechain amides in ^1H , ^{15}N -HSQC spectra. The ^1H , ^{15}N -TROSY-HSQC spectrum of C3d therefore effectively eliminates seventy chemical shifts from the sidechains of Gln and Asn to reduce spectral complexity for backbone assignment.

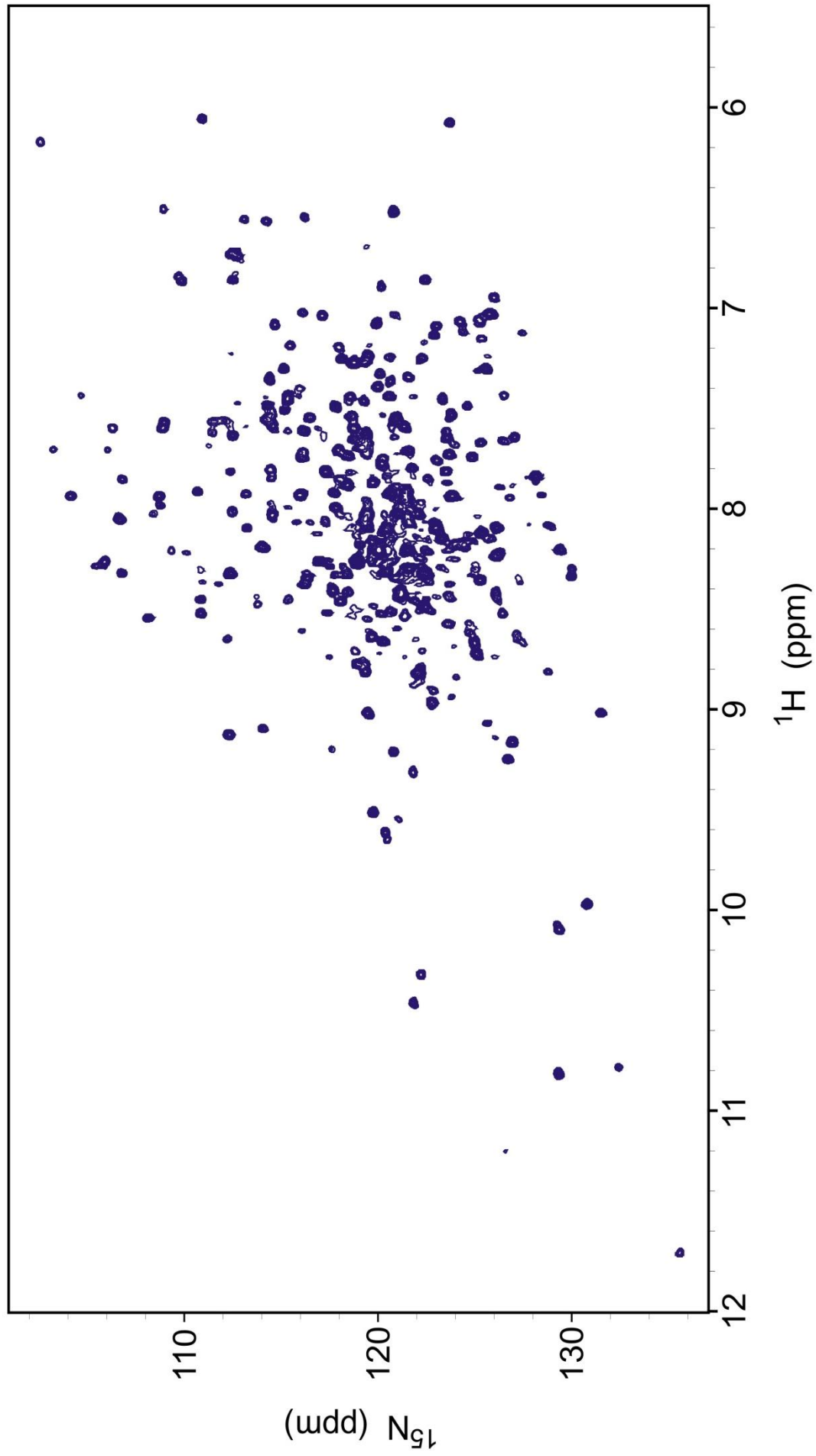


Figure 4.2 | ^1H , ^{15}N -TROSY-HSQC spectrum of C3d collected on a Bruker Avance III 700 MHz spectrometer at 25 °C

The spectrum was collected with a uniformly ^2H , ^{15}N , ^{13}C -labeled C3d sample of 400 μM in 20 mM NaH_2PO_4 , 150 mM NaCl, pH 6.8. At this point spectral resolution and sensitivity improve enough to observe chemical shifts from individual backbone amides. The number of expected chemical shifts is 301, of which 300 are observed in this spectrum.

4.2.2 TRIPLE-RESONANCE NMR EXPERIMENTS FOR SEQUENTIAL ASSIGNMENT

Heteronuclear multidimensional NMR experiments are used to establish sequential connectivity of the chemical shifts evolved from the backbone amide of a protein. Chemical shift assignment of a protein begins with an HSQC spectrum and triple- resonance spectra are used to correlate different backbone nuclei, such as ^{15}N and ^{13}C . The ^1H , ^{15}N -TROSY-HSQC spectrum of C3d (Figure 4.2) provides a starting point for the sequential assignment. Individual chemical shifts were manually picked from the spectrum and given arbitrary peak numbers until they were each assigned to the corresponding residues in C3d. In order to establish connectivity of backbone amides between amino acids, six TROSY-based triple-resonance experiments were collected from a uniformly ^2H , ^{15}N , ^{13}C -labeled C3d sample on a Bruker Avance III 700 MHz spectrometer equipped with a cryoprobe at 25 °C. NMR triple-resonance experiments are usually recorded and analyzed in pairs: HNCA and HN(CO)CA; HNCACB and HN(CO)CACB; and lastly HNCO and HN(CA)CO. The magnetization transfer pathways of these experiments are described in Chapter II. As a general rule, one of the experiments in the pair records both intra- and inter-residue correlations and the other one records only inter-residual correlations. All six spectra comprise a series of ^1H - ^{15}N planes (*i.e.* HSQC spectrum in slices) expanded with a carbon dimension. In the HNCA experiment, each H_N plane contains chemical shifts evolved from $^{13}\text{C}_\alpha$ from residues i and $i-1$ in the protein sequence. The HN(CO)CA contains only the $^{13}\text{C}_\alpha$ chemical shift from residue $i-1$. These experiments allow each backbone amide peak to be linked to the preceding $^{13}\text{C}_\alpha$ carbon, and sequential assignment can be done by matching the chemical shifts of each residue's own and previous carbons. The HNCACB and HN(CO)CACB experiments work in the same manner and contain both the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts. Chemical shifts of $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ are strongly correlated with amino acid types and can be used in primary residue typing when no extensive side chain information is available. The HNCO and HN(CA)CO works similarly, just with the ^{13}CO (carbonyl) chemical shifts rather than the $^{13}\text{C}_\alpha$. The increased dimensionality in triple-resonance experiments leads of a significant reduction in chemical shift degeneracy. For instance, the probability of a triplet of chemical shifts (three-dimensional NMR) that is unique to a residue at a particular site in the protein is higher than a doublet of chemical shifts (HSQC experiment). Therefore, a greater number of chemical shift assignments can be done unambiguously. The ^1HN , ^{15}N chemical shifts of an

amide form a spin system collectively with the correlated $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and ^{13}CO chemical shifts of residues i and $i-1$. In other words, a spin system represents an individual amino acid (except proline which does not give signal in NMR spectrum) in the protein and should contain chemical shifts evolved from its backbone amide, $\text{C}\alpha$, $\text{C}\beta$ and carbonyl. The total number of assignable spin system in C3d is 296. Table 4.2 shows a summary of the number of chemical shifts that were picked unambiguously in the triple-resonance spectra. A comprehensive list of the spin systems in C3d and their chemical shift values are detailed in Appendix 1. To visualize the information contained within the triple-resonance spectra, a strip plot was generated for each identified spin system using CcpNmr Analysis (CCPN, University of Cambridge). Figure 4.3 shows a specimen strip plot of a spin system correlating $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and ^{13}CO chemical shifts of residues i and $i-1$.

Table 4.2 Completeness of the NMR spectra			
TROSY-based spectrum	No. peaks expected	No. of peaks observed and linked to a spin system	Completeness
HSQC	301	300	99.7%
HN(CO)CA	296	288	97.3%
HNCA	592	516	87.2%
HN(CO)CACB	572	523	91.4%
HNCACB	1144	846	73.9%
HNCO	296	164	55.4%
HN(CA)CO	592	223	37.6%
<i>In summary</i>			
No. of spin system expected = 296			-
No. of spin system observed = 288			97.3%
No. of spin system with chemical shifts of $\text{C}\alpha$ from both residue i and $i-1$ = 261			88.2%
No. of spin system with chemical shifts of $\text{C}\alpha$ and $\text{C}\beta$ from both residue i and $i-1$ = 219			73.9%
No. of spin system with chemical shifts of $\text{C}\alpha$, $\text{C}\beta$ and CO from both residue i and $i-1$ = 131			44.3%

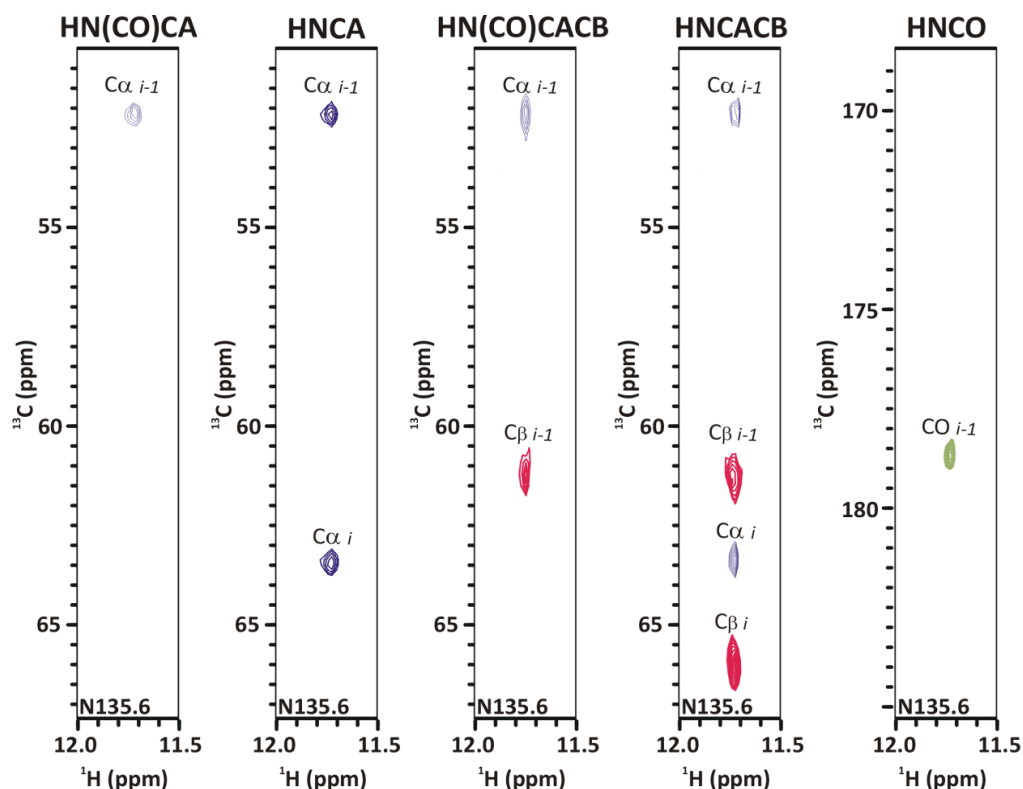
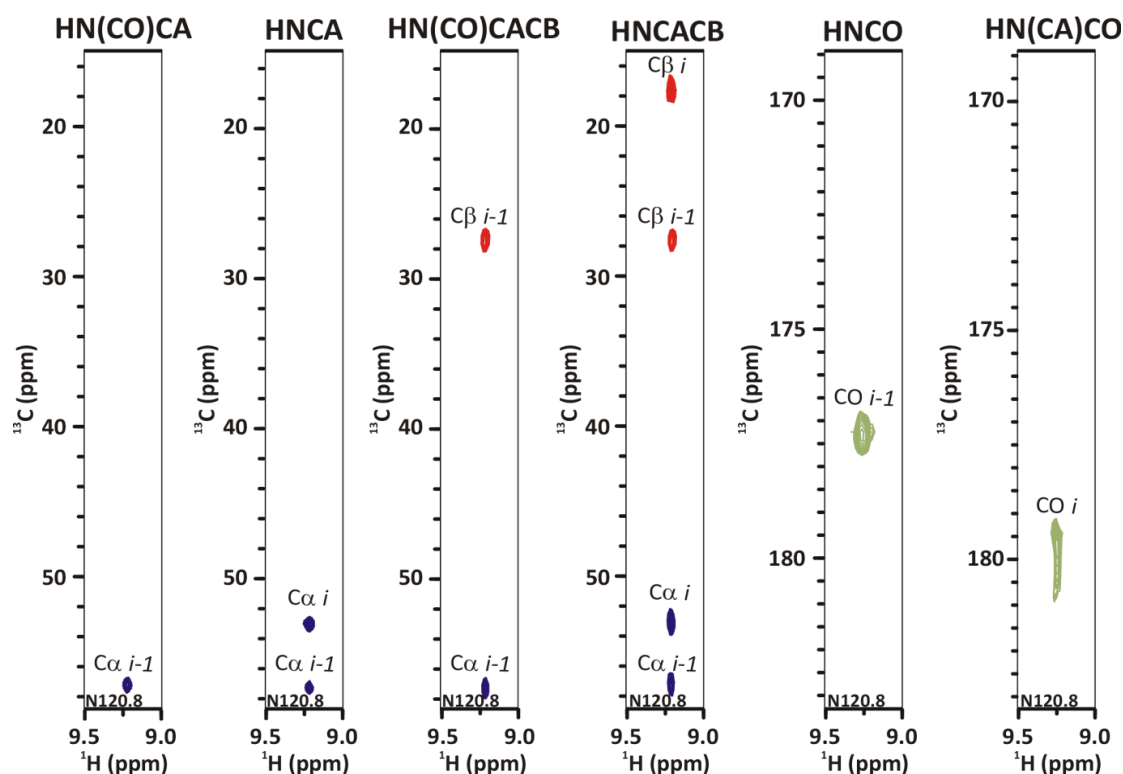
(a) Spin system 202**(b) Spin system 91**

Figure 4.3 | Specimen strip plot of selected spin systems from TROSY-based triple-resonance experiments

- (a) Strips from the triple resonance spectra of spin system 202. Individual carbons were coloured-coded and annotated. Each strip showed the expected number of peaks associated with the spin system, except in the HN(CA)CO where no peak was observed due to inherent low sensitivity of the experiment. The $C\alpha$ and $C\beta$ chemical shifts of residue i (*i.e.* own resonance) indicates a residue type of threonine.
- (b) Strips from the triple resonance spectra of spin system 91. The $C\alpha$ and $C\beta$ chemical shifts of residue i (*i.e.* own resonance) indicates a residue type of alanine.

4.2.3 COMPUTER-ASSISTED BACKBONE RESONANCE ASSIGNMENT

Connectivity of the spin systems and assigning them correctly to residues in the protein sequence can be established by traditional triple-resonance assignment process (Gardner KH and Kay LE, 1998; Grzesiek S and Bax A, 1992). Identification of the amino acid type is usually carried out by analysis of their $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ position using average carbon chemical shifts for individual amino acids deposited on the Biological Magnetic Resonance Data Bank (BMRB) website (<http://www.bmrwisc.edu>). Some amino acids, namely glycine, alanine, valine, isoleucine, serine and threonine have characteristic $^{13}\text{C}\alpha$ or $^{13}\text{C}\beta$ shifts, making them straight forward to allocate to a spin system. Carbon chemical shifts of some amino acid types, such as glutamic acid, glutamine, arginine, lysine and histidine, are not typically unique and clusters of these amino acid types often lead to ambiguity during the assignment process. C3d contains thirteen proline residues which are not assignable as they do not generate backbone chemical shifts due to the lack of amide group in a peptide bond. Although no NMR sidechain experiment was performed, the sidechain resonances of all five tryptophan in C3d are observed in the $^1\text{H}, ^{15}\text{N}$ -TROSY-HSQC spectrum. The sidechains of tryptophan can generally be easily visualized as they have an average proton chemical shift of about 10 ppm and a nitrogen chemical shift of 129 ppm, placing them in the bottom left hand corner in the $^1\text{H}, ^{15}\text{N}$ -TROSY-HSQC spectrum (Figure 4.2). A set of optimized TROSY triple-resonance pulse sequences were applied to suppress chemical shifts evolved from the sidechains of tryptophan in the triple resonance experiment so that they can be easily identified.

While the amino acid type could not be established for every single isolated spin system, it was possible to assign 48% of these spin systems to specific residues in the C3d protein sequence by looking at a series of sequential backbone chemical shifts connectivity. In general, there is only one position in the protein sequence where the chemical shifts of sequentially connected spin systems are consistent with a specific stretch of amino acid in the protein sequence. The extent of the backbone assignment is summarised in Figure 4.4 and the assignment of spin systems to specific residues in the C3d sequence is listed in Appendix III.

The connectivity of spin systems and assigning them to specific residues in the protein sequence was assisted by two assignment program, PACES (Coggins BE and Zhou P, 2003) and UNIO-MATCH (Volk J *et al.*, 2008). By providing a list of spin systems with all the available chemical shifts for $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and ^{13}CO , PACES links the

spin systems according to Equation 4.1 and then assigns them to the user deposited protein sequence. PACES follows an semi-automated iterative search process and allows user intervention at all stages.

Equation 4.1 | Assembly of spin systems connectivity

For two spin systems i and j with observed chemical shifts:

$$i = \{C\alpha^i, C\beta^i, CO^i, C\alpha^{i-1}, C\beta^{i-1}, CO^{i-1}\}$$

$$j = \{C\alpha^j, C\beta^j, CO^j, C\alpha^{j-1}, C\beta^{j-1}, CO^{j-1}\}$$

a sequential connection ij will be established if:

$$|C\alpha^i - C\alpha^{j-1}| \leq \delta C\alpha,$$

$$|C\beta^i - C\beta^{j-1}| \leq \delta C\beta,$$

$$|CO^i - CO^{j-1}| \leq \delta CO,$$

where $\delta C\alpha$, $\delta C\beta$ and δCO represent user defined chemical shift thresholds for connectivity and were kept below 0.4 ppm.

Similarly, UNIO-MATCH is a newly developed memetic algorithm from the Torsten Herrmann's group for fully automated sequence-specific polypeptide backbone NMR assignment. The programme employs local optimization for tracing partial sequence-specific assignments within a global population-based search environment. By providing a list of spin systems with all the available chemical shifts, UNIO-MATCH's optimization heuristics delivers high efficiency and robustness for matching sequentially connected spin systems to stretches of specific residue sequences in a protein. Backbone residue assignments suggested by both computer programmes are in agreement with each other and the connectivity of spin systems were each inspected individually using strip plots. The NMR backbone assignment of C3d only achieved 48% completion with the acquired triple-resonance data owing to overwhelming spectral overlap. The assigned residues were visually confirmed by triple-resonance strip plots and two other orthogonal approaches, namely specific amino acid labelling and nuclear Overhauser effect spectroscopy (NOESY). Figure 4.5 shows a specimen strip plot to illustrate the sequential connectivity of spin systems and assignment to a specific

sequence of residues in C3d. Having completed partial assignment of the ^1H , ^{15}N -TROSY-HSQC spectrum of C3d, chemical shifts that have been assigned to specific residues in the protein sequence were annotated in Figure 4.6. Figure 4.7 illustrates the extent of NMR assignment when these residues were mapped onto the crystal structure of C3d. The use of specific amino acid labelling and NOESY experiments will be covered in the following sections 4.2.4 and 4.2.5.

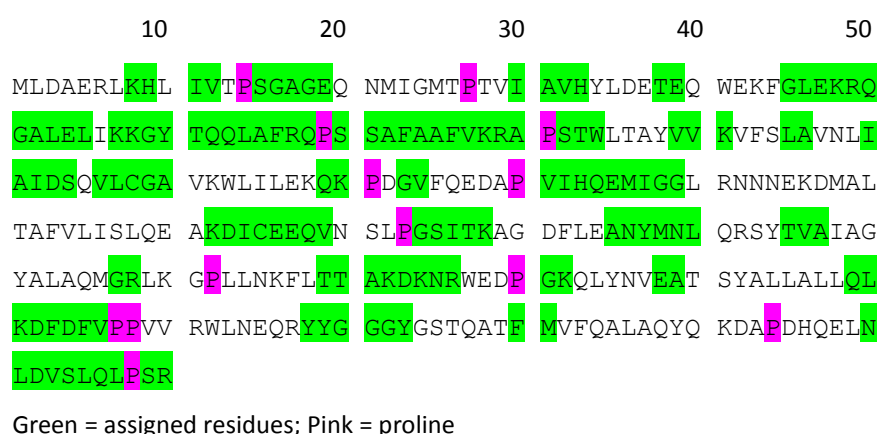


Figure 4.4 | A summary of residues that were unambiguously assigned in the C3d protein sequence

C3d comprises 310 amino acid residues and the N-terminal methionine is generally not observed in protein NMR spectra. In addition, there are 13 prolines in the C3d sequence which do not generate observable resonance in NMR spectra. Therefore, the total number of assignable residues (*i.e.* spin systems) is 310 residues – 13 Pro – 1 Met = 296. Using the traditional triple resonance assignment approach, as well as selective amino acid labelling and NOESY experiments as confirmation, it was able to unambiguously assign 142 NMR spins systems to corresponding residues in the C3d sequence. The NMR backbone assignment of C3d achieved (142/296) 48% completion.

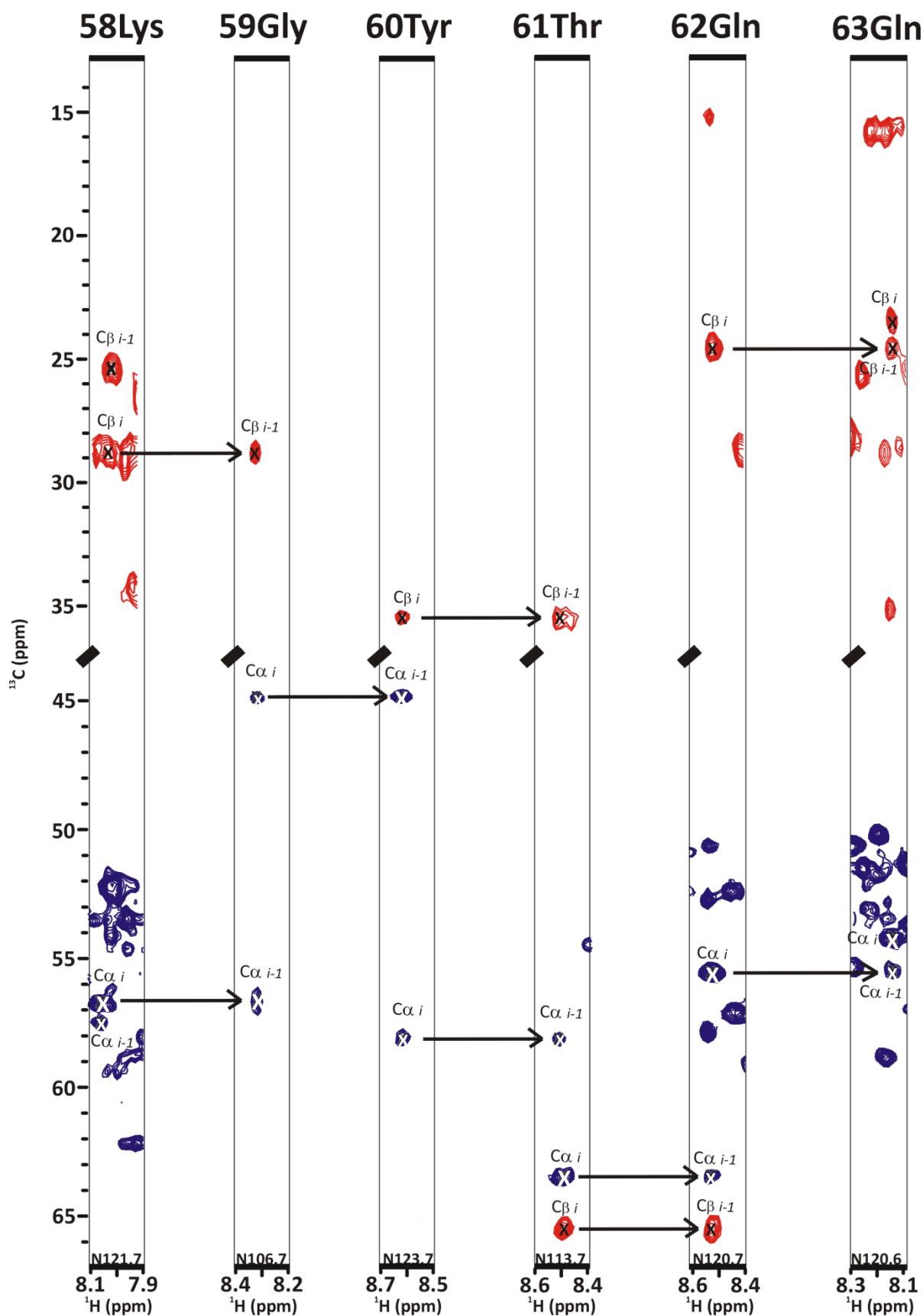


Figure 4.5 | Specimen strip plot to illustrate sequential connectivity of spin systems

Strips show the resonances assigned to 58 Lys – 63 Gln in the C3d protein sequence. Arrows between the strips show the connectivity between $\text{C}\alpha(i)$ and $\text{C}\alpha(i-1)$, as well as the connectivity between $\text{C}\beta(i)$ and $\text{C}\beta(i-1)$. The $\text{C}\alpha(i-1)$ and $\text{C}\beta(i-1)$ peaks were distinguished using the HN(CO)CA and HN(CO)CACB experiments, respectively.

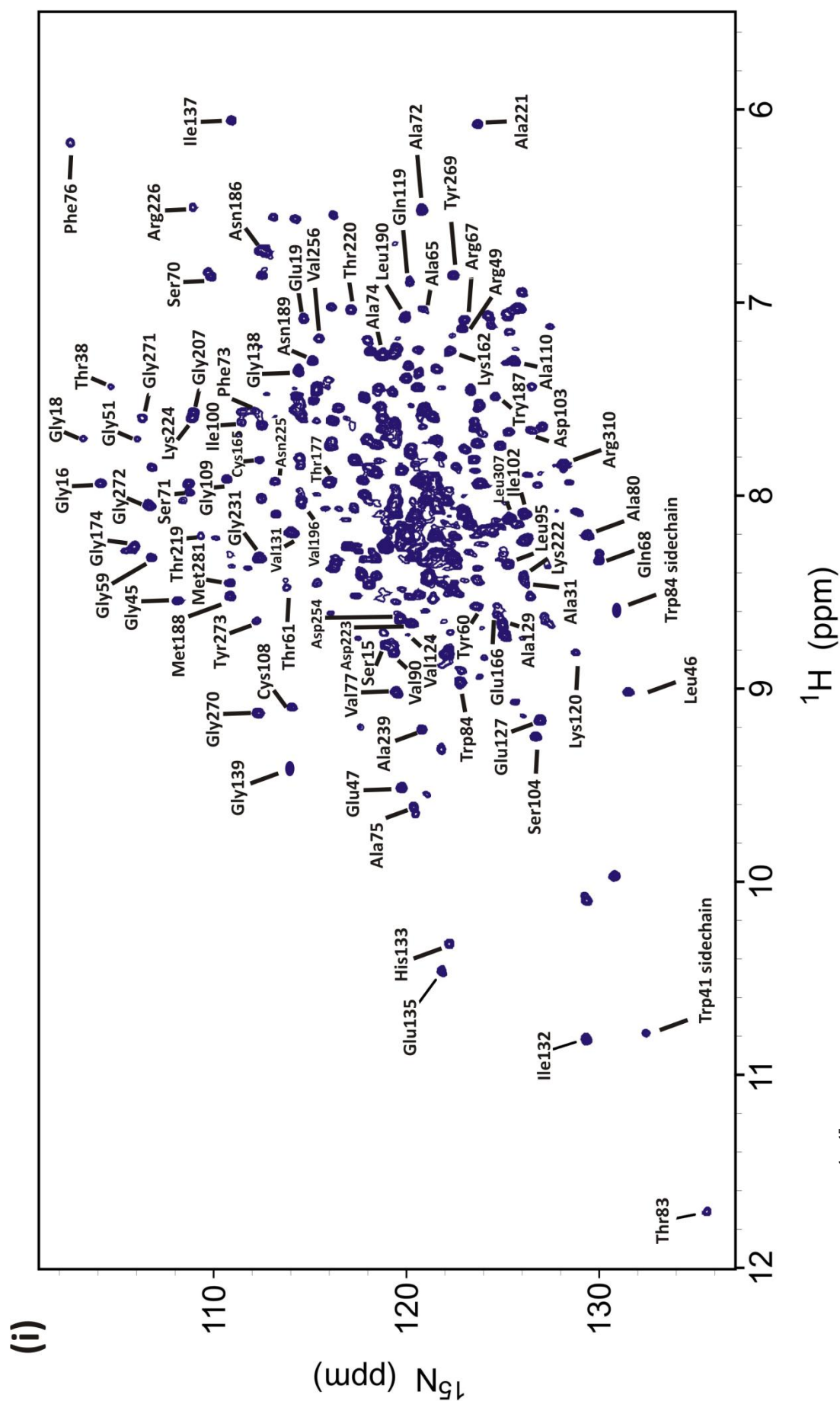


Figure 4.6 | Partially assigned ^1H , ^{15}N -TROSY-HSQC spectrum of C3d

(i) NMR backbone assignment of C3d achieved 48% completion with the acquired triple resonance data. Assigned chemical shifts are annotated in the spectrum to show their corresponding residues in the protein sequence. Annotation of the crowded central region (113 - 125 ppm in the nitrogen frequency and 7.0 - 8.5 ppm in the hydrogen frequency) are shown in the following zoom-in figures.

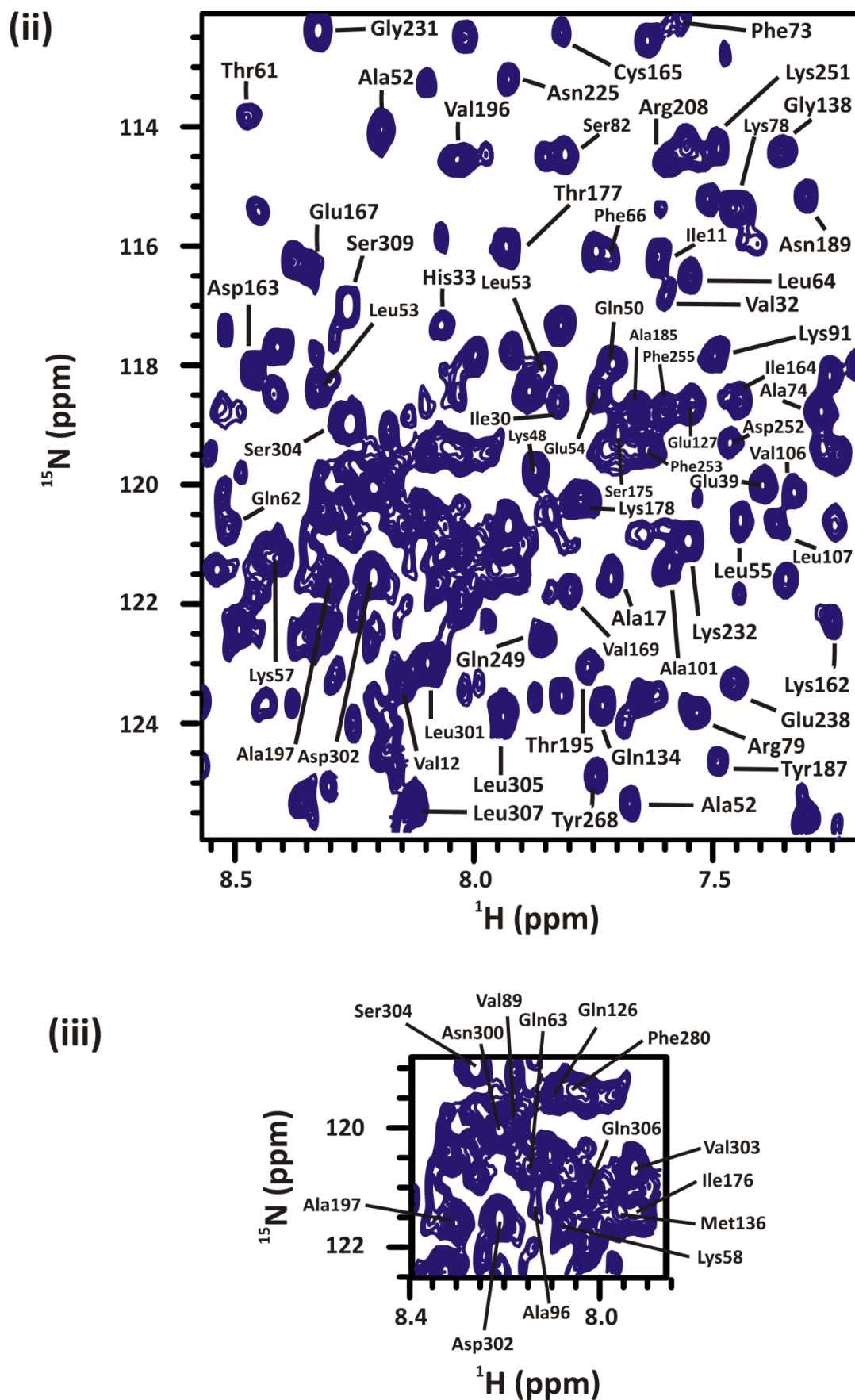


Figure 4.6 | Partially assigned ^1H , ^{15}N -TROSY-HSQC spectrum of C3d

- (ii) Zoom-in of the crowded central region (113 - 125 ppm in the nitrogen frequency and 7.0 - 8.5 ppm in the hydrogen frequency) with residue annotations.
- (iii) Zoom-in of the most crowded central region (119 - 122 ppm in the nitrogen frequency and 7.9 - 8.4 ppm in the hydrogen frequency) with residue annotations.

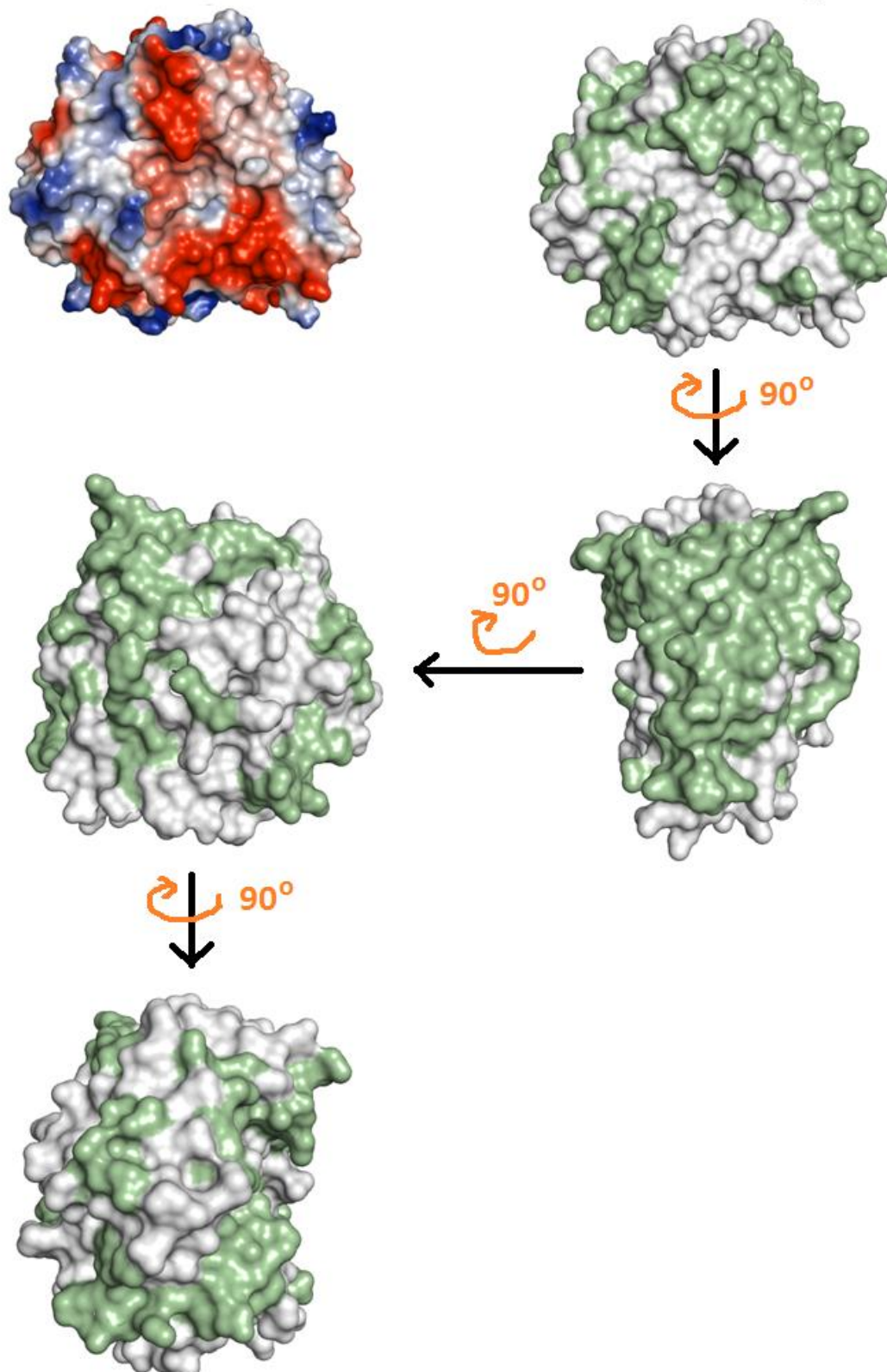
electrostatic representation**NMR backbone assignment**

Figure 4.7 | Mapping of NMR assigned residues to crystal structure of C3d

C3d is shown as an electrostatic surface representation in the top left corner to illustrate the negatively charged CD21 SCR1-2 binding surface. NMR assigned residues, coloured in green, are mapped to the adjacent C3d molecule showing the same CD21 SCR1-2 binding surface. The C3d molecule mapped with NMR assignment is shown in 3x 90° anticlockwise rotations to illustrate the extent of NMR assigned residues that are solvent accessible.

4.2.4 SPECIFIC AMINO ACID LABELLING TO CONFIRM IDENTITY OF SPIN SYSTEMS

Isotope labelling by residue type can be a useful tool for the backbone resonance assignment of challenging protein systems. Incorporation of specific ^{15}N -labeled amino acids in a protein followed by direct observation in an HSQC spectrum provides a means to confirm the chemical shifts from a given residue type. The strategy is often utilized to resolve ambiguities by locating a specific residue type in segments of sequentially connected spin systems in the primary protein structure (McIntosh LP and Dahlquist FW, 1990). However, bacterial expression of proteins with selective ^{15}N -labeled amino acid labelling is not straightforward and usually requires considerable effort to optimize expression systems to produce sufficient amounts of protein for NMR analysis. Incorporation of ^{15}N -labeled amino acids also has an disadvantage of line-broadening effect in NMR spectrum owing to dipole interaction between amide-proton resonance of the labelled residue and its own $\text{H}\alpha$. Furthermore, not all amino acid types are amenable to selective isotope labelling for NMR studies due to different *E. coli* metabolic pathways involved in amino acid biosynthesis (McIntosh LP and Dahlquist FW, 1990). For instance, glycine is metabolized to serine which is then subsequently converted to tryptophan by serine hydroxymethyltransferase. Glutamic acid is the precursor of arginine and proline, and glutamine and glutamic acid are inter-convertible by transaminases. Hence, transaminases or other metabolic enzymes that scramble ^{15}N render conventional selective isotope labelling difficult to apply for certain amino acids. Highly specific incorporation is observed for some amino acids that do not undergo isotope scrambling, such as leucine, isoleucine, valine and tyrosine. These residues present a clear advantage over the others by eliminating labelling ambiguity.

There are thirty-eight leucine residues in C3d, which makes it the most abundant amino acids in C3d and accounts for 12.3% of the total amino acid composition. ^{15}N -leucine was chosen for residue specific labelling to confirm partial NMR backbone assignment of C3d because of its resistance to isotope scrambling in bacterial metabolic pathways (Takeuchi K *et al.*, 2007). A ^{15}N -leucine labelled C3d sample was expressed in *E. coli* using a selective amino acid labelling protocol and purified as described in Chapter II. The final protein concentration yielded was 80 μM . A mass spectrum of the ^{15}N -leucine labelled C3d sample indicates full ^{15}N -leucine incorporation (Figure 4.8). The molecular weight of unlabeled C3d is 34,710 Da and thus the

expected molecular weight of ^{15}N -leucine labelled C3d is 34,748 Da, with each ^{15}N -leucine adding 1 Da to the unlabeled construct. The ^{15}N -leucine labelled sample appears to have been proteolytically cleaved at the C-terminus by seven amino acids (LQSLPSR, 910 Da) and gives a molecular species of 33,836 Da. An equivalent unlabeled C3d construct has an expected molecular weight of 33,800 Da and the molecular mass difference is consistent with the thirty-six leucines in the truncated C3d construct. Figure 4.9 shows a ^1H , ^{15}N -HSQC spectrum of ^{15}N -leucine labelled C3d collected on a Bruker Avance III 700 MHz spectrometer equipped with a cryoprobe at 25 °C. Only thirty chemical shifts are observed out of thirty-eight leucine residues in C3d. Low sensitivity of this spectrum is partly attributed to using a sample of high molecular weight protein at low protein concentration, as well as the incorporation of ^{15}N -labeled leucine which leads to line-broadening as mentioned previously. It appears some chemical shifts evolved from leucine broadened away under the experimental conditions tested and therefore are not observed in the spectrum. Overlaying the ^1H , ^{15}N -HSQC spectra of uniformly ^{15}N -labeled C3d and exclusively ^{15}N -leucine labelled C3d allows instant identification of chemical shifts evolved from leucine in the protein. It provides confirmation of residue identity in the sequential backbone assignment, namely Leu46, Leu53, Leu55 and Leu95, Leu107, Leu190, Leu250, Leu301 and Leu307.

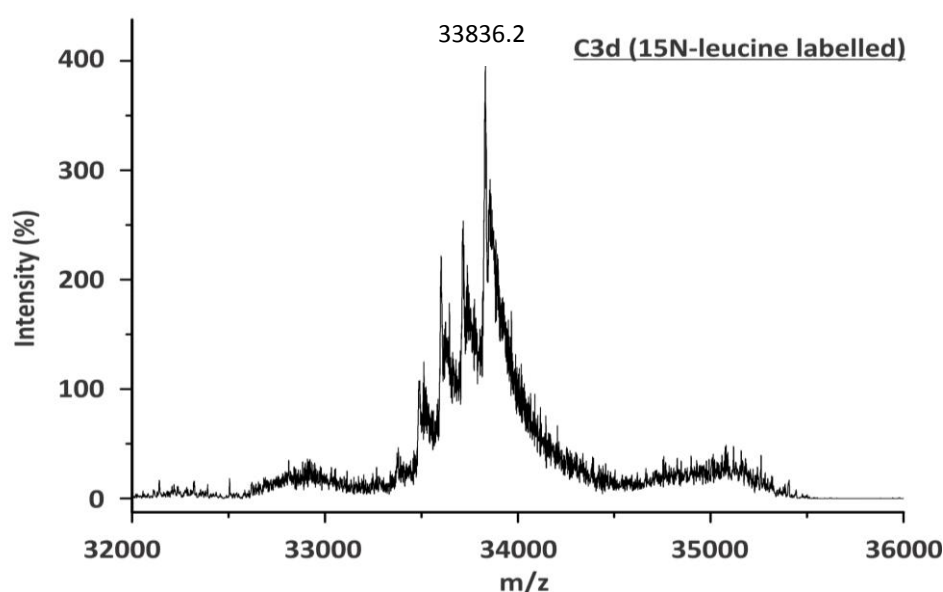


Figure 4.8 | Molecular mass profile of ^{15}N -leucine labelled C3d obtained by maximum entropy processing of the ESI-MS spectrum

The theoretical mass of ^{15}N -leucine labelled C3d sample is 34,748 Da. It appears the protein sample has undergone C-terminal truncation by seven amino acids (LQSLPSR, 910 Da) and gives a molecular species of 33,836 Da. Further C-terminally truncated C3d constructs are observed in the trailing peaks upfield of 33,836 Da, with each peak having a mass difference of approximately 110 Da (*i.e.* average molecular weight of one amino acid)

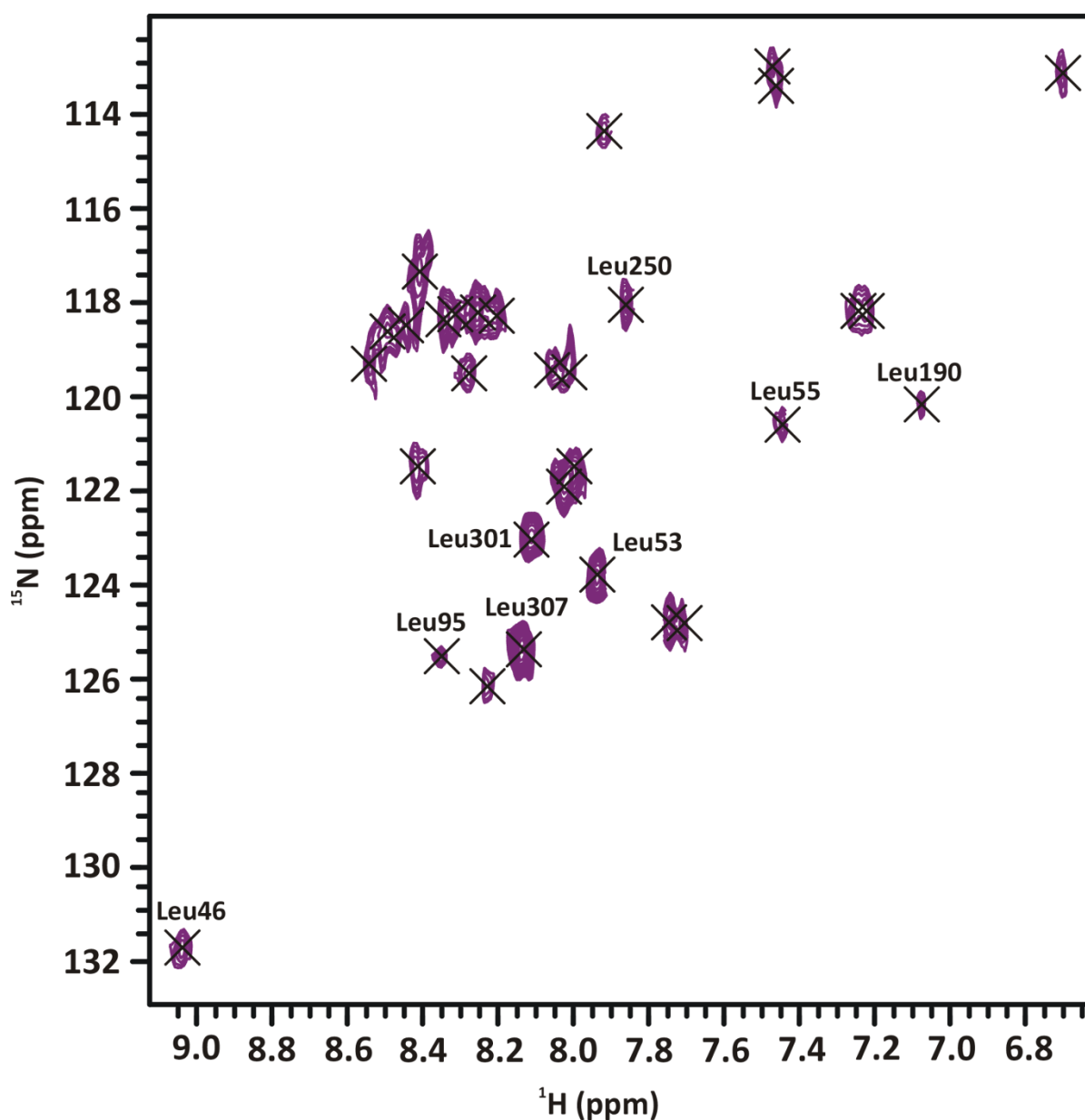


Figure 4.9 ^1H , ^{15}N -HSQC spectrum of ^{15}N -leucine labelled C3d collected on a Bruker Avance III 700 MHz at 25 °C

The spectrum was collected with a ^{15}N -leucine labelled C3d sample of 80 μM in 20 mM NaH_2PO_4 , 150 mM NaCl, pH 6.8. Only 30 chemical shifts are observed out of 38 leucine residues in C3d. The chemical shifts of Leu46, Leu53, Leu55 and Leu95, Leu107, Leu190, Leu250, Leu301 and Leu307 are annotated. It was not possible to assign the remaining 21 peaks to specific leucine residues in the C3d protein sequence.

4.2.5 NOESY EXPERIMENTS IN SEQUENTIAL BACKBONE RESONANCE ASSIGNMENT

Most of the backbone resonance assignments in C3d were carried out on the adaptation of triple-resonance experiments for large protein molecules. ^1H , ^{15}N -NOESY-HSQC experiments measure correlations between each amide proton to all nearby protons which are close in space (within 5 Å) and provide inter-proton distance information for solution NMR structure calculations (Marion D *et al.*, 1989). Nuclear Overhauser effects (NOEs) between spatially nearby protons in the protein are traditionally used to generate distance restraint to define the secondary and tertiary structure of a protein, but they can also support and complement the sequential assignment procedure. The crystal structure of C3d shows its secondary structure is predominately helical (PDB:1C3D) and helical conformations are characterized by short distances between amide protons on sequentially connected residues, such as backbone amide protons of residue i , residue $i-1$ and residue $i+1$. Pairs of protons that are closer in space give rise to more intense NOEs and thus the backbone HN – HN NOEs can be used to confirm the sequential assignment of backbone chemical shifts established by the traditional triple-resonance approach. A ^1H , ^{15}N -NOESY-HSQC spectrum of C3d was collected from a uniformly ^2H , ^{15}N , ^{13}C -labeled C3d sample on a Bruker Avance III 700 MHz spectrometer equipped with a cryoprobe at 25 °C. Obtaining NOESY chemical shift assignments is usually a laborious endeavour, particularly for high molecular weight proteins as some regions in the spectrum exhibit excessive chemical shift overlap. The ^1H , ^{15}N -NOESY-HSQC spectrum of C3d was analyzed as exhaustively as possible and Figure 4.10 shows a selective region of the ^1H , ^{15}N -NOESY-HSQC spectrum where the sequentially connected backbone NOE chemical shifts were used to verify sequential assignment utilizing the triple-resonance spectra. Each strip in the spectrum contains an amide chemical shift at the diagonal and NOEs from all other protons within 5 Å in space. Although NOEs to aliphatic protons can be observed in some strips, only NOEs to amide protons are needed to establish the sequential connectivity. Despite the fact that two neighbouring amide protons of residue $i\pm 1$ are visible to the amide proton of residue i in a helix, very often only correlations to one residue, either $i-1$ or $i+1$, are observed due to missing peak resulting from intrinsic line-broadening effect of large proteins or excessive chemical shift overlap in certain regions.

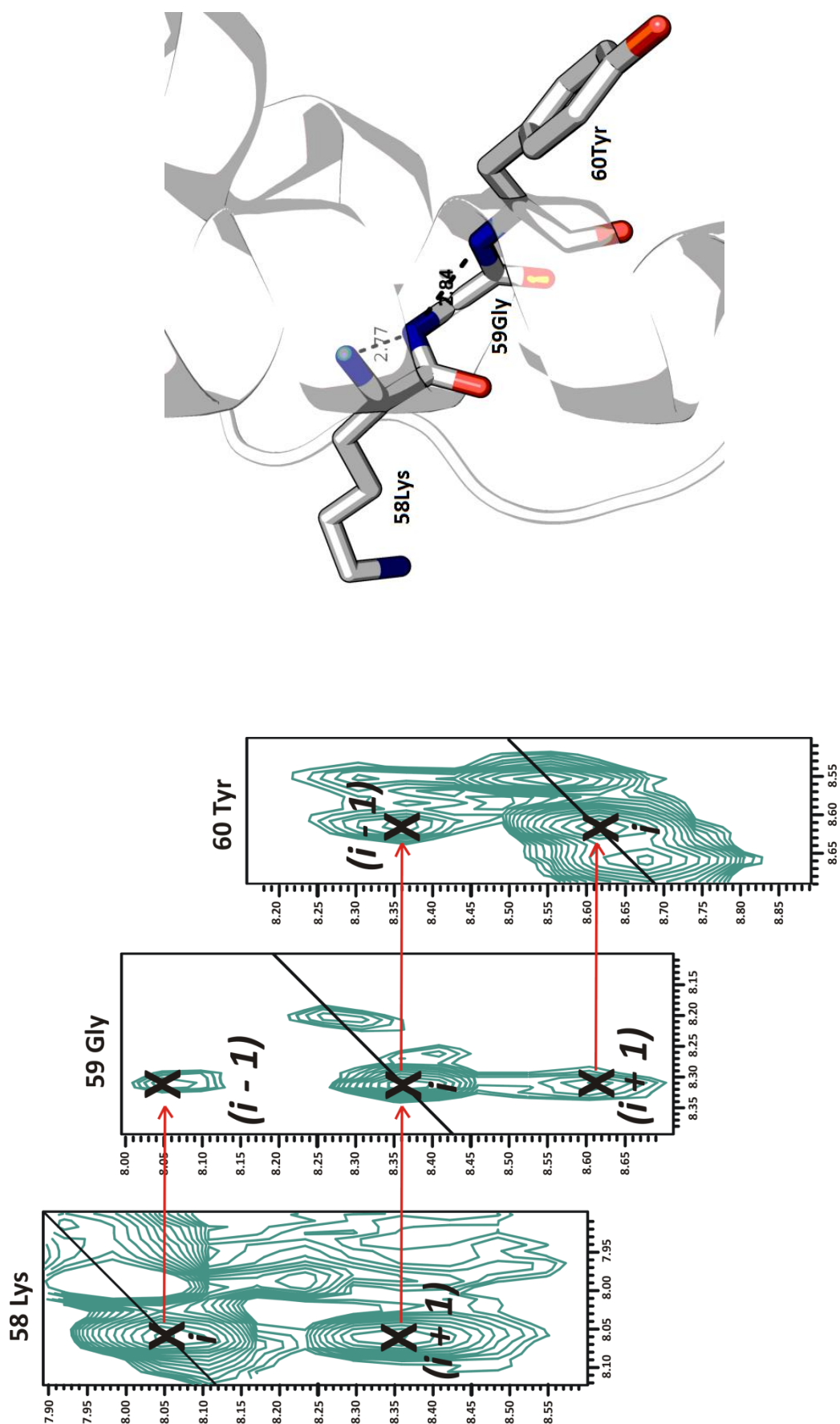


Figure 4.10 Specimen strip plot of the ^1H , ^{15}N -NOESY-HSQC spectrum of C3d collected on a Bruker 700 MHz spectrometer at 25 °C

Strips show the resonances assigned to 58 Lys – 60 Tyr in the C3d sequence. Helical conformations are characterized by short distances (<3 Å) between amide protons on sequentially connected residues, such as residue i , residue $i-1$ and residue $i+1$. Each strip contains an amide chemical shift at the diagonal and NOEs from all other protons within 5 Å in space. Arrows between the strips show the connectivity between backbone amides. The crystal structure of C3d indicates the distance between backbone amide of 58 Lys – 59 Gly and 59 Gly – 60 Tyr are 2.77 Å and 2.84 Å, respectively.

4.3 DISCUSSION

Inherent molecular characteristics of C3d, such as high molecular weight, amino acid composition and its predominant helical content, create an unfavourable protein system for NMR studies. All together it leads to substantial spectral degeneracy in certain regions of the NMR spectra. Helical proteins with molecular weight larger than 20 kDa have proven difficult to characterize by NMR because these proteins usually have strongly overlapping chemical shifts due to the similar chemical environment in the core of the protein. In the case of C3d, the reduction in spectral sensitivity also arises from increased molecular-weight-dependent relaxation rates. The spectral degeneracy is further complicated by ionic strength dependent aggregation of C3d. Aggregation increases the apparent molecular weight of C3d and intensifies the line-broadening effect of NMR signals observed in the spectra. Although raising the ionic strength effectively minimizes aggregation of C3d, there is a trade-off between sensitivity of NMR experiments and salt content of the protein sample. As mentioned previously, increasing the salt content diminishes sensitivity by dissipating the radiofrequency power used in NMR pulse sequences. Intermediate protein concentration coupled with moderate ionic strength, such as 150 mM NaCl, provides favourable experimental conditions to collect NMR triple-resonance data for sequential backbone chemical shift assignment of C3d.

C3d was labelled with ^2H , ^{15}N and ^{13}C isotopes and three pairs of complementary triple-resonance NMR experiments were recorded. Based on the chemical shift information derived from these spectra, 48 % of the assignable residues are identified and further validated by complementary assignment approaches such as selective amino acid labelling and ^1H , ^{15}N -NOESY-HSQC. The chemical shift dispersion of backbone resonance is inevitably poor in certain regions of the NMR spectra owing to high molecular weight and helical content of C3d, namely 7.5 – 8.6 ppm in the proton frequency and 117.6 – 125.0 ppm in the nitrogen frequency. Stretches of identical amino acids tend to result in similar inter-residues correlations and thus are usually difficult to establish sequential connectivity. For instance, the presence of three consecutive asparagine residues, Asn141 – Asn143, hinders efforts in sequential assignment and these residues remain unassigned. Alanine is the second most common amino acid type in C3d after leucine and there are thirty-one alanines

and thirty-eight leucines in the protein sequence. While the spin systems of alanine can be easily established owing to its characteristic carbon chemical shifts in the triple resonance spectra, a lot of them fall into a crowded region of the C3d spectra. It is not uncommon to find several alanine spin systems on a crowded nitrogen planes in the HN(CO)CACB and HNCACB spectra (data not shown). It is very difficult to isolate these alanine spin system, let alone establish sequential connectivity of these alanines correctly. Leucine is the most abundant amino acid type in C3d and has very similar carbon chemical shifts to asparagine and aspartic acid. There are thirteen asparagines, fifteen aspartic acids and thirty-eight leucines in C3d. All together, they give rise to sixty-six very similar spin systems in C3d. The assignment is further complicated by the presence of alanine-leucine or leucine-leucine pairs in the protein sequence, notably Ala149-Leu150, Leu213-Leu214 and Ala243-Leu244-Leu245-Ala246-Leu247-Leu248. There are also two identical stretches of Ala-Leu-Ala-Gln in the protein sequence from residues 202-205 and residues 285-288, respectively. Mapping these sequences of residues on the crystal structure of C3d shows some of these sequences are spatially adjacent to each other, thereby creating a very similar chemical environment for surrounding residues. It therefore leads to spectral overlap and ambiguity in chemical shift assignment. All together, high molecular weight, high helical content, unfavourable amino acid composition and ionic strength-dependent aggregation of C3d hinder full NMR backbone assignment. Fortunately, some amino acids that constitute the binding surface to CD21 SCR1-2 (Clemenza L and Isenman DE, 2000; Hannan JP *et al.*, 2005; Bomb M, 2008;) have been assigned and this assignment will form the basis to evaluate the C3d/CD21 SCR1-2 interaction by NMR in Chapter V.

Employing selective methyl group labelling and methyl-TROSY NMR experiments in future work would help to yield a complete backbone chemical shift assignment of C3d. The techniques have become a vital strategy for backbone chemical shift assignment in large protein systems where triple-resonance or NOESY experiments do not succeed in yielding full assignments (Ruschak AM and Kay LE, 2010). It usually involves using an isotope labelling scheme where isoleucine, leucine and valine methyls are protonated in an otherwise highly deuterated background in concert with methyl-TROSY NMR experiments that preserve the lifetimes of the resulting NMR signals. Methyl groups have more favourable relaxation characteristics and give more intense signal in NMR spectra than amides because they are located at the end of amino acid sidechains and rotate

more rapidly. They are frequently found in the hydrophobic core of protein structures and selective protonation of methyl groups makes it possible to measure NMR signals with high sensitivity and resolution for determination of solution structure of larger proteins. Methyl-TROSY routinely provides a rescue strategy to be used on large protein systems where amide TROSY fails and successful application includes proteins with molecular weights of several hundred kilodalton, notably p97 (Isaacson RL *et al.*, 2007), TET2 (Amereo C *et al.*, 2009), the CheA-CheW complex (Hamel DJ and Dahlquist FW, 2005), aspartate transcarbamoylase (Velyvis A *et al.*, 2007), Sec A (Gelís I *et al.*, 2007) and the 20S proteasome (Sprangers R and Kay LE, 2007). Selective methyl protonation of isoleucine, leucine and valine was developed by Kay *et al.* using 2-ketobutyrate ($3,3\text{-}^2\text{H}_2$) and 2-ketoisovalerate ($3\text{-}^2\text{H}_2$) as growth supplement during protein expression in minimal media comprised 100% D_2O (Gardner KH and Kay LE, 1997; Goto NK *et al.*, 1999). Using established methods in the literature, it has been shown that over 90% of methyl groups were labelled as CH_3 while other protons were substituted with deuterons (Tugarinov V and Kay LE, 2003). In addition, all the carbons or strictly the methyl carbons can be labelled with ^{13}C . A new isotope strategy for labelling only one methyl from Val or Leu while the other methyl will be NMR-inactive (*i.e.* $\text{U-}[^2\text{H}]\text{-Val/Leu-}[^{13}\text{CH}_3, ^{12}\text{CD}_3]$) has been developed by using 2-keto-3methyl- $\text{d}_3\text{-}3\text{-d}_1\text{-butyrate}$ (Tugarinov V and Kay LE, 2004). Despite the fact that this labelling strategy would more or less half the detectable methyl concentration, these methyls do not suffer from relaxation caused by intra-residue methyl protons. As a result, the methyl-TROSY experiments correlating methyl group with other sidechain ^{13}C resonances has have better sensitivity as the magnetization transfer is not diverted. Methyl-TROSY NMR experiments have been developed for the assignment of methyl ^{13}C and ^1H chemical shifts from Ile, Leu and Val residues in large proteins. The methyl chemical shifts are then correlated with backbone amide chemical shifts using two categories of pulse sequences – the first class of experiment transfer magnetization energy from the methyl group to the backbone amides for detection, while the second class uses an ‘out-and-back’ magnetization transfer where sidechain carbons or backbone carbonyls are correlated with methyl ^{13}C and ^1H (Tugarinov V and Kay LE, 2003). The method has proven successful for high molecular weight system as large as the 82 kDa malate synthase G (Tugarinov V *et al.*, 2005).

CHAPTER V

C3d INTERACTION ANALYSIS BY NMR SPECTROSCOPY

5.1 INTRODUCTION TO LIGAND INTERACTION ANALYSIS BY NMR

NMR spectroscopy is a powerful technique which allows determination of solution structure and dynamics of protein (Evans JNS, 1995; Rattle H, 1995). The technique relies on the absorption of energy by precessing nuclei in the presence of a magnetic field. The nuclei most frequently used in biological studies of NMR are ^1H , ^{15}N , ^{13}C and ^{31}P . These nuclei are magnetically active and their spins are split into two energy levels in a magnetic field. The population of nuclei in each energy level is mostly equal at equilibrium, nevertheless application of an appropriate radiofrequency pulse (*i.e.* pulse sequence) induces transition of nuclei spins between the two energy levels. It is both the excitation and the subsequent relaxation of nuclei spins to equilibrium that are exploited in NMR experiments. The transition of nuclei spins between separate energy levels is highly sensitive to the chemical environment of the nuclei. Changes in the chemical environment of a residue in a protein arising from ligand binding will lead to a change in the chemical shifts which can be

monitored by NMR. Chemical shift perturbation analysis is the most commonly used NMR method to map ligand binding interfaces on a protein (Zuiderweg ER, 2002). Although the NMR backbone assignment of C3d did not yield full completion as described in Chapter IV, the partially (48%) assigned ^1H , ^{15}N -HSQC spectrum of C3d can be monitored for changes in chemical shifts of specific residues in response to binding of unlabelled ligand. The chemical shifts of specific residues may either shift or broaden (*i.e.* disappear) upon ligand binding depending on their exchange characteristics. The exchange characteristic of chemical shifts is a reflection of the interaction affinity between a protein and its ligand. Hence, chemical shift perturbation analysis can yield residue specific information to allow elucidation of the location of the binding sites when mapped onto a known protein structure. In addition to mapping binding sites on a protein, NMR titration experiments can determine kinetics and affinities of protein – ligand interactions. The kinetics of an interaction can be broadly divided into three categories, namely fast, intermediate and slow exchange (Table 5.1).

Table 5.1 | Summary of NMR kinetics

Exchange characteristics	Typical affinity of interaction	Observed changes in chemical shifts
Fast	Weak ($K_D > 10\ \mu\text{M}$)	- Gradually shifting of chemical shifts with the same peak intensity throughout the course of titration. - Changes are easily trackable during titration.
Intermediate	Intermediate ($100\ \text{nM} < K_D < 10\ \mu\text{M}$)	- Extensive line broadening throughout the course of titration. - Chemical shifts usually disappear completely beyond half saturation point and reappear elsewhere in the spectra when the protein is in a fully bound state. - Very difficult to track.
Slow	Strong ($K_D < 100\ \text{nM}$)	- Chemical shifts are split between 'free' and 'bound' state during the course of titration. - Chemical shifts of the free state gradually disappear and are replaced by those of the bound state. - Not easily trackable.

5.2 CHEMICAL SHIFT PERTURBATION ANALYSIS OF C3d

5.2.1 THE PROTEIN-PROTEIN INTERACTION BETWEEN C3d AND CD21 SCR1-2

With the partially assigned ^1H , ^{15}N -HSQC spectrum of C3d and the availability of a crystal structure of C3d (Nagar B, Jones RG, Diefenbach RJ, Isenman DE and Rini JM, 1998; PDB:1C3D), NMR titration experiments with unlabelled CD21 SCR1-2 were performed to map its binding site on C3d. Two sets of NMR titration experiments were carried out to assess ionic strength dependency of the interaction – under physiologically relevant ionic strength (150 mM NaCl) and under very high ionic strength (400 mM NaCl). The protein sample used for spectra acquisition was 180 μM of ^{15}N -labelled C3d. Figure 5.1 shows a comparison of the titration experiments conducted in physiological and in very high ionic strengths. Upon addition of unlabelled CD21 SCR1-2 to ^{15}N -labelled C3d, a number of chemical shifts show line broadening effects (*i.e.* disappearance) regardless of the ionic strengths tested. All residues on C3d that mediate binding to CD21 SCR1-2 display intermediate exchange characteristics (Figure 5.2). The observation is in agreement with ITC and SPR determined binding affinity of $\sim 1\ \mu\text{M}$ in 150 mM NaCl buffer (Chapter III). Although residues on C3d that were perturbed appear to be largely the same in 150 mM NaCl and 400 mM NaCl buffer, the spectra collected in 20 mM NaH_2PO_4 , 150 mM NaCl, pH 6.8 show signs of protein aggregation and have poor signal-to-noise ratio. The reduction in spectral resolution is partly attributed to the lack of deuteration in protein sample which would otherwise reduce the line width of chemical shifts. Collecting HSQC spectra of ^{15}N -labelled C3d in 150 mM NaCl buffer appears suboptimal owing to ionic strength dependent aggregation characteristic of the protein. A sample buffer of 150 mM NaCl is regarded as ‘highly salty’ in the scope of NMR experiments. It is generally not advised to use high salt buffer for NMR experiments as high salt content would hamper sensitivity of NMR experiments as explained in Chapter IV. In the case of C3d, the set of titration experiments carried out in 20 mM NaH_2PO_4 , 400 mM NaCl, pH 6.8 nevertheless exhibits better signal-to-noise ratio. The use of very high salt buffer further diminishes the notorious line broadening effect inherent to the ^1H , ^{15}N -HSQC spectra of C3d. When collecting HSQC spectra of C3d on a 700 MHz spectrometer, it is only under very high salt conditions that ^{15}N -labelled C3d sample is capable to produce

spectral quality marginally comparable to the spectra collected with deuterated sample in physiological ionic strength buffer (*i.e.* ^2H , ^{13}C , ^{15}N -labelled C3d in 20 mM NaH_2PO_4 , 150 mM NaCl, pH 6.8 used for triple resonance experiments in Chapter IV). Hence, the set of titration experiments carried out in 400 mM NaCl buffer was used to illustrate residue specific information of the C3d/CD21 SCR1-2 interaction. Table 5.2 shows a summary of residues on C3d that are perturbed in the HSQC spectrum upon addition of CD21 SCR1-2.

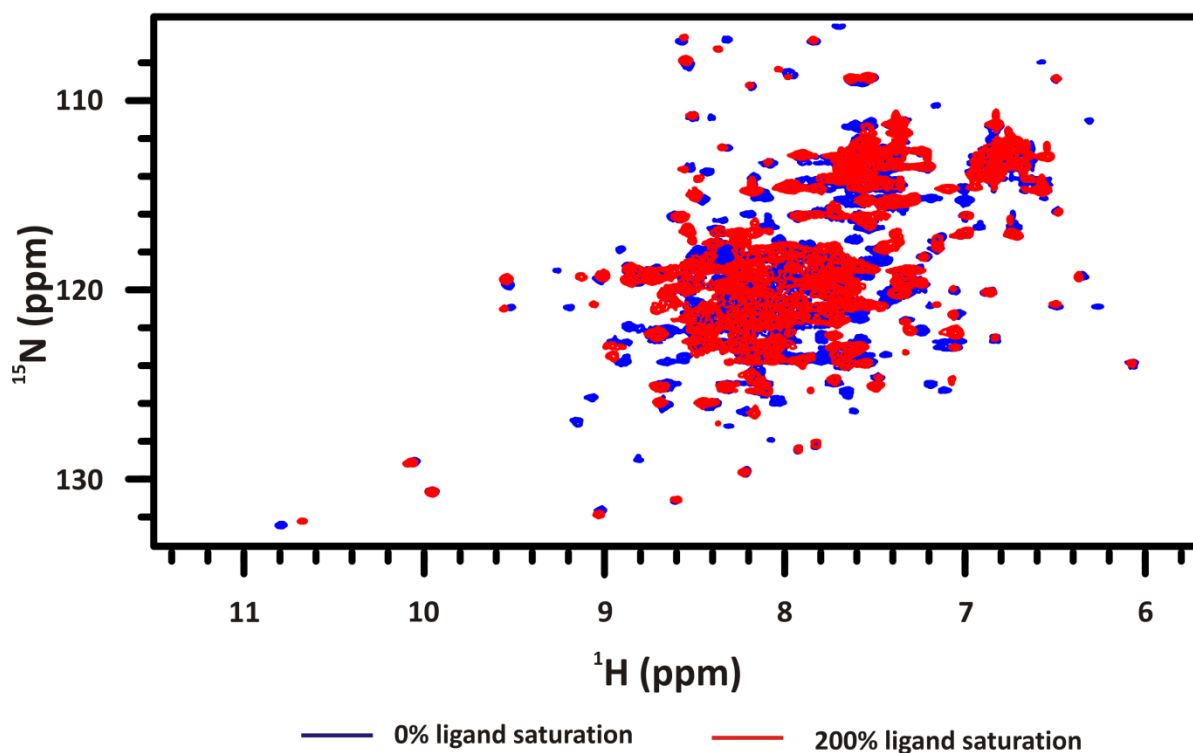
At early titration point (25% ligand saturation), only a few backbone chemical shifts show small signs of chemical shift perturbation. The chemical shift perturbation becomes apparent in the next titration point (50% ligand saturation) and the binding site for CD21 SCR1-2 is clearly delineated. Residues Thr38, Glu39, Ile100, Ala101, Ile102, Asp103, Ser104, Asp163, Ile164, Glu166, Glu167 and Gln168 (set 1) on C3d exhibit the strongest line broadening effect upon addition of CD21 SCR1-2 (Figure 5.2). These residues form a large contiguous binding surface for CD21 SCR1-2 when mapped to the crystal structure of C3d (Figure 5.3). A number of residues on C3d also exhibit notable line broadening effect at this point, namely Val32, His33, Trp41 sidechain, 59Gly, Tyr60, Thr61, Gln62, Gln63, Gln68 and Val131 and Ile137 (set 2). The set of chemical shifts that are most strongly perturbed at half saturation point (*i.e.* set 1) all broadened away in the next titration point (75% ligand saturation). While chemical shifts of set 2 residues show a significant reduction in chemical shift intensity at this point, a number of additional chemical shifts also exhibit line broadening effect. These additional chemical shifts include Gly51, Ala52, Leu53, 75Ala, 84Trp, 126Gln and 249Gln. Interestingly, a fair number of chemical shifts appear to broaden away in late titration points (100% and 200% ligand saturation). These residues are only affected at high concentration of CD21 SCR1-2. As expected with interactions showing intermediate exchange characteristics, a number of new chemical shifts appear in the ^1H , ^{15}N -HSQC spectrum of C3d in end titration point (200% ligand saturation). It is most likely due to the reappearance of chemical shifts of set 1 residues when C3d is in a fully bound state. However, it is not clear where chemical shifts that have broadened away in early titration points would reappear in the spectrum in end titration points.

The disappearance of additional chemical shifts at end titration points could be a result of the formation of C3d/CD21 SCR1-2 protein complex. The protein complex has a combined molecular weight of 50 kDa. It is possible that the additional molecular weight of CD21 SCR1-2 (15 kDa) becomes a burden and exerts an overall line broadening effect on C3d. The effect appears to have become increasingly significant when the protein complex approaches full saturation. It is speculated residues that only started to experience line broadening effect in late titration points (75% ligand saturation or greater) are not directly involved in binding to CD21 SCR1-2. These residues broaden away mostly due to an increase in apparent molecular weight of C3d, and thus increased line width in chemical shifts, when the protein is in a fully bound state.

Table 5.2 | NMR chemical shift perturbation analysis of the C3d/CD21 SCR1-2 interaction

Titration point	Observations in the ^1H , ^{15}N -HSQC spectrum of C3d
0% ligand saturation	Starting point
25% ligand saturation	Residues that exhibit minor line broadening effect: Thr38, Glu39, Ile100, Ala101, Ile102, Asp103, Ser104, Asp163, Ile164, Glu166, Glu167, Gln168 (set 1)
50% ligand saturation	- Residues from set 1 exhibit strong line broadening effect - Residues that exhibit notable line broadening effect: Val32, His33, Trp41 sidechain, 59Gly, Tyr60, Thr61, Gln62, Gln63, Gln68, 84Trp, Val131, Ile137 (set 2)
75% ligand saturation	- Residues from set 1 completely broadened away - Residues from set 2 almost broaden away - Residues starting to exhibit line broadening effect: Gly51, 52Ala, Leu53, 75Ala, 126Gln, 249Gln (set 3) and other unassigned residues
100% ligand saturation	- Residues from set 1 and set 2 completely broaden away. - Residues from set 3 and a fair number of additional unassigned residues exhibit significant line broadening effect
200% ligand saturation	- Increasing number of residues broaden away - New chemical shifts starting to reappear in the HSQC spectrum

(a) 20 mM NaH₂PO₄, 150 mM NaCl, pH 6.8



(b) 20 mM NaH₂PO₄, 400 mM NaCl, pH 6.8

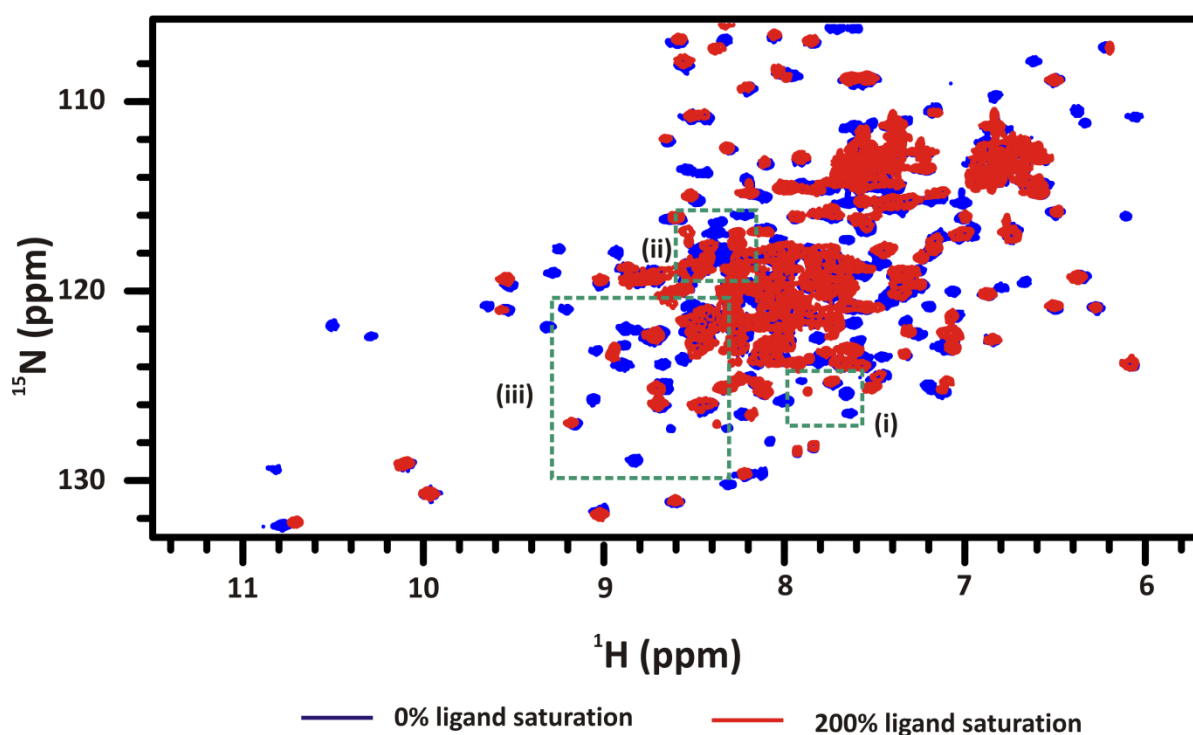


Figure 5.1 ^1H , ^{15}N -chemical shift perturbation experiments to map the interaction surfaces on C3d for CD21 SCR1-2

- (a) Titration experiments performed in 20 mM NaH₂PO₄, 150 mM NaCl, pH 6.8 are of poor spectral resolution. However, residues that experience changes in chemical shift upon ligand addition are the same as those observed in high salt condition.
- (b) Titration experiments performed in high salt buffer exhibit better spectral resolution and a number of residues clearly show intermediate exchange characteristics. Expanded views of the indicated area (i), (ii) and (iii) are shown in the following figure.

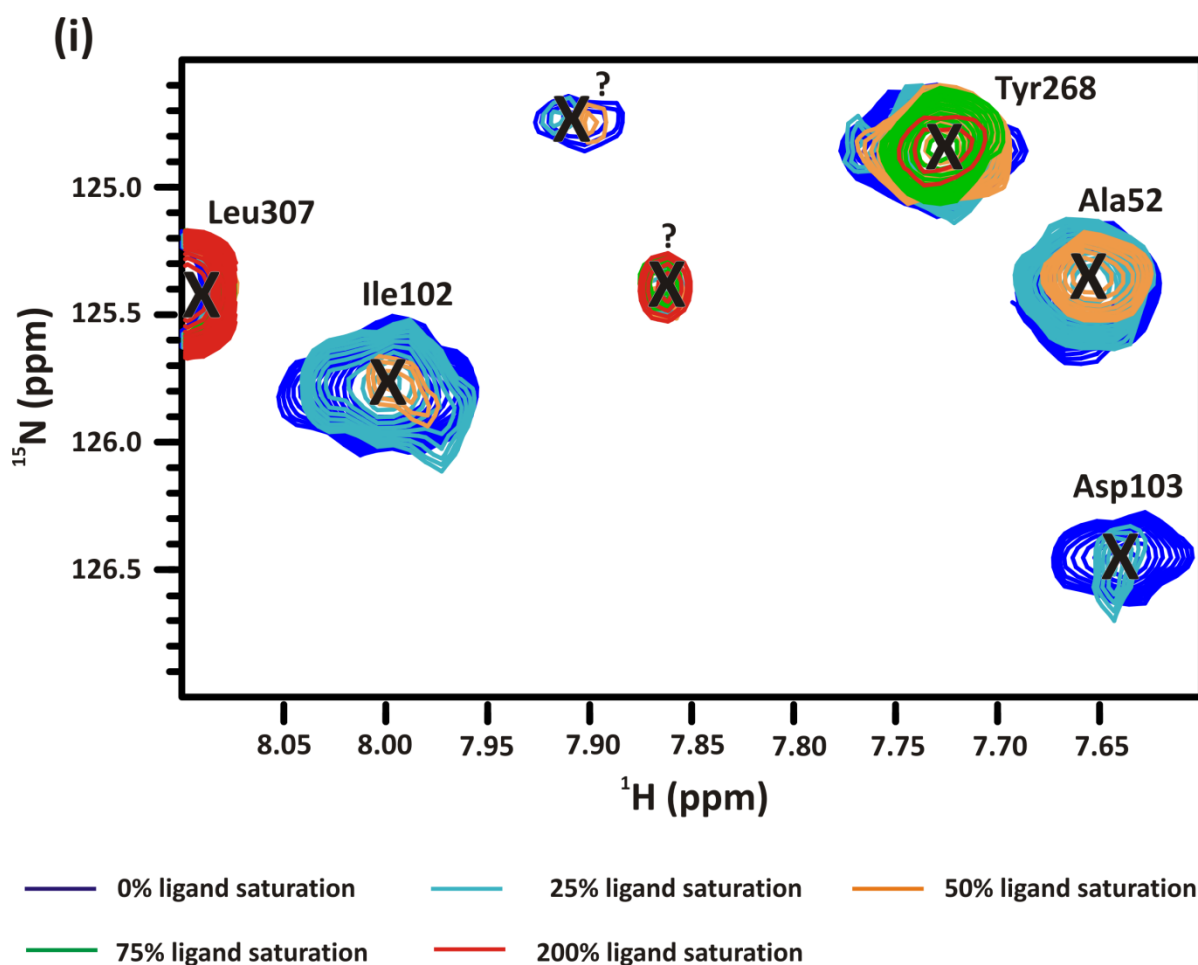


Figure 5.2 Inserts showing expanded views of the indicated areas in the ^1H , ^{15}N -chemical shift perturbation experiments

- (i) Different titration points were overlaid and chemical shifts that were able to be assigned to specific residues in C3d were annotated. The chemical shift corresponding to Leu307 is unaffected over the course of the titration experiments, suggesting this residue is not involved in the binding to CD21 SCR1-2. On the other hand, Ile102 and Asp103 show substantial chemical shift changes and broaden away in mid titration point. These two residues thus form a prominent binding site for CD21 SCR1-2. Ala52 and Tyr268 also show some changes in chemical shift towards late titration points. Residues began to broaden away in late titration points are thought to interaction with CD21 SCR1-2 through non-specific interaction, or due to high apparent molecular weight of the protein complex.

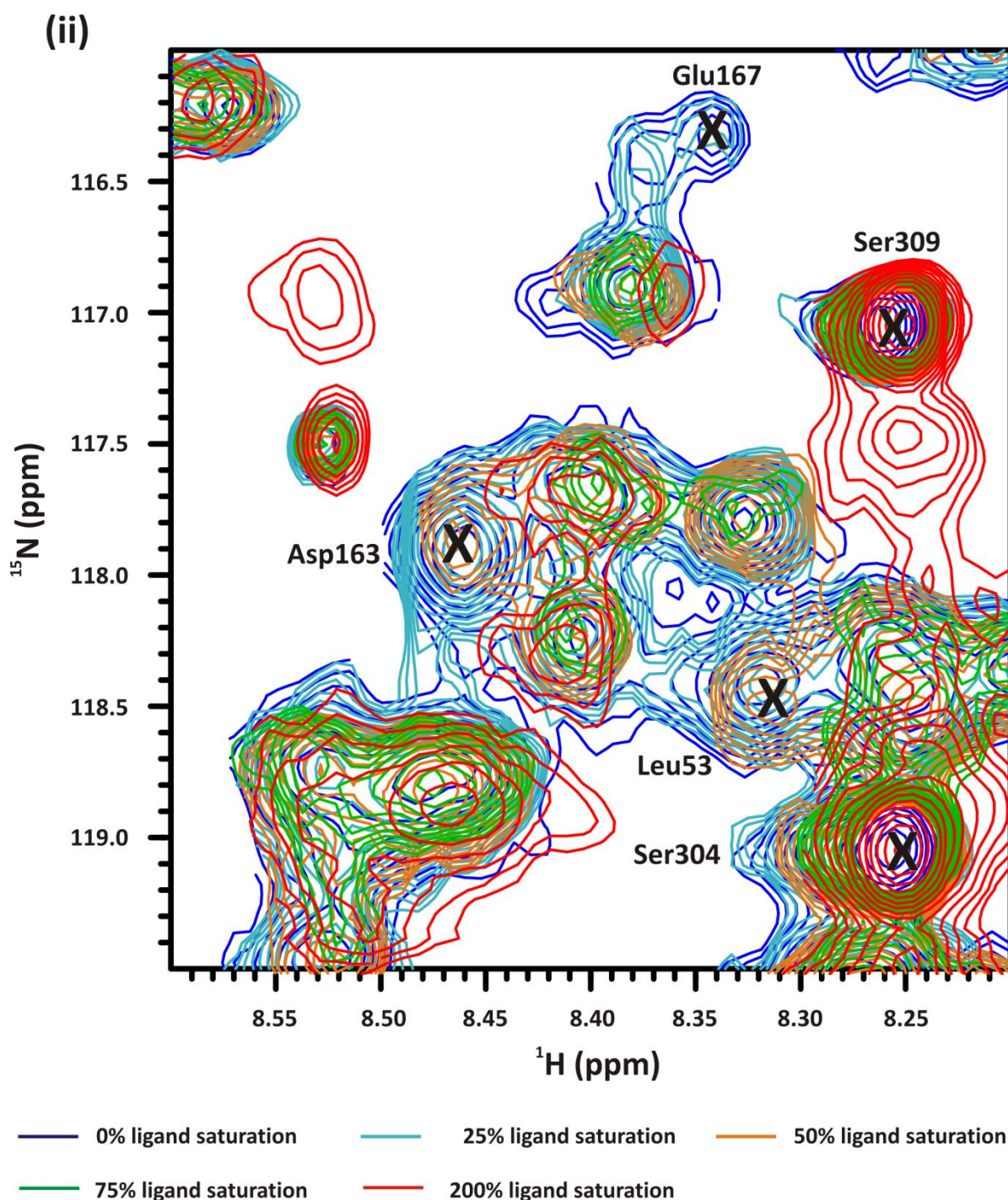


Figure 5.2 Inserts showing expanded views of the indicated areas in the ^1H , ^{15}N -chemical shift perturbation experiments

(ii) Different titration points were overlaid and chemical shifts that were able to be assigned to specific residues in C3d were annotated. Substantial chemical shift overlaps were observed in this area of the spectrum and it was not possible to monitor all the changes from individual chemical shifts.

Chemical shifts corresponded to Ser 304 and Leu309 were unaffected over the course of the titration, whereas Asp163 and Glu167 showed substantial chemical shift changes and broadened away in mid titration point. Leu53 only showed signs of chemical shift perturbation in late titration points and therefore was not considered to be part of the prominent binding site for CD21 SCR1-2. Two new chemical shifts arose in end titration points (two peaks only in red) which were characteristic of residues undergoing intermediate exchange. Chemical shifts that broadened away usually reappear elsewhere in the spectra when the protein is in a fully bound state.

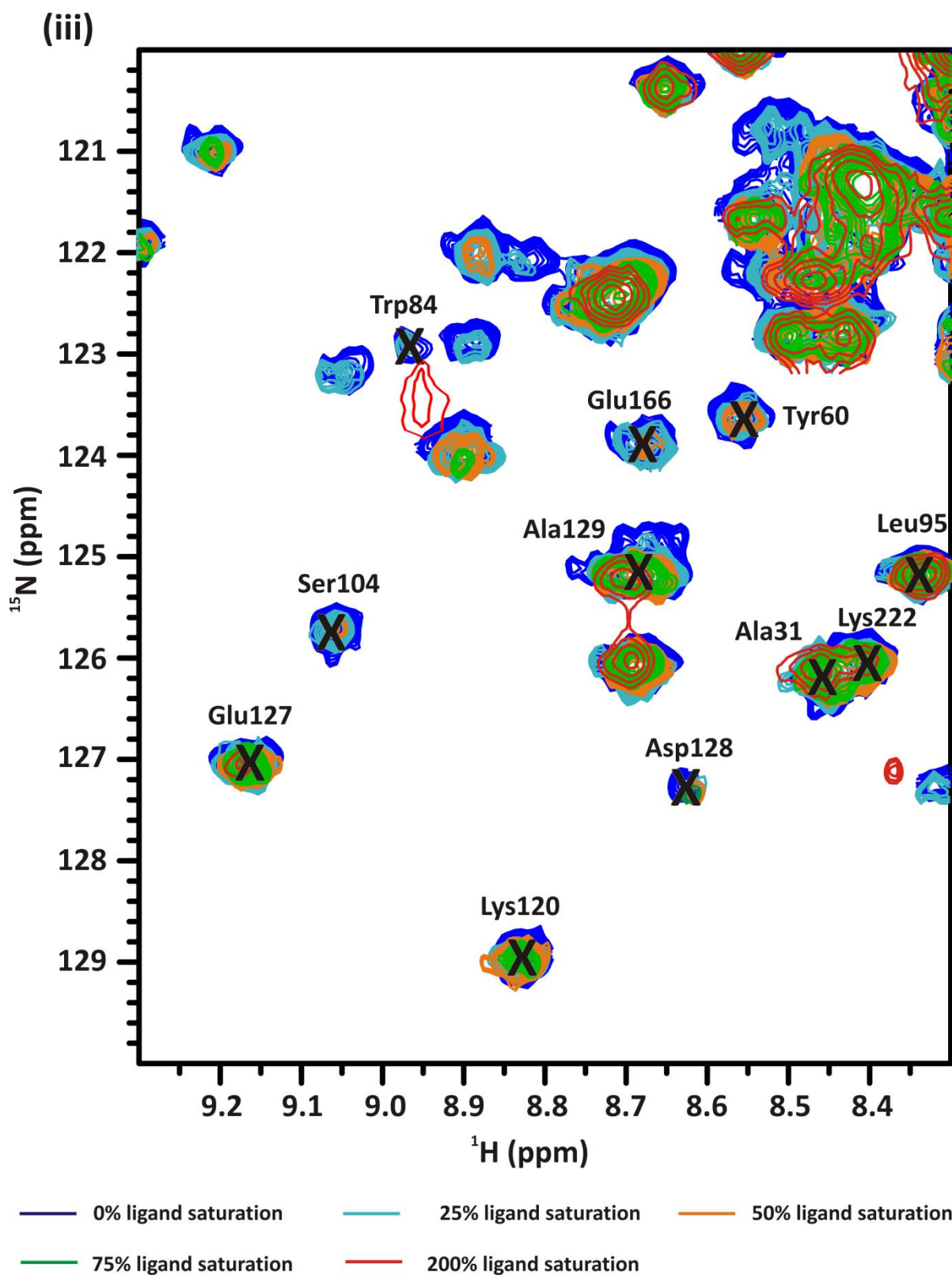


Figure 5.2 | Inserts showing expanded views of the indicated areas in the ^1H , ^{15}N -chemical shift perturbation experiments

- (iii) Different titration points were overlaid and chemical shifts that were able to be assigned to specific residues in C3d were annotated. Chemical shifts corresponded to Leu95, Lys222, Ala31 and Glu127 were unaffected over the course of the titration, whereas Trp84, Ser104 and Glu166 showed substantial chemical shift changes and broadened away in mid titration point. Lys120, Asp128 and Tyr60 only showed signs of chemical shift perturbation in late titration points and therefore were not considered to be part of the prominent binding site for CD21 SCR1-2.

Figure 5.3 illustrated the binding surfaces for CD21 SCR1-2 on C3d defined by NMR chemical shift perturbation analysis under physiologically relevant ionic strength. The partially assigned ^1H , ^{15}N -HSQC spectrum of C3d can only provide limited information with regard to assessment of residues that mediate the interaction observed in crystal structure of the C3d/CD21 SCR1-2 complex solved by Szakonyi G *et al.* (Szakonyi G, Guthridge JM, Li D, Youn, K, Holers VM and Chen XS, 2001; PDB: 1GHQ). In this crystal structure, residues Ser70, Ile 115, Leu116, Glu117, Lys118, Asn170 and Lys178 on C3d are involved in binding to CD21 SCR1-2. Only Ser70 and Lys178 were assigned in the ^1H , ^{15}N -HSQC spectrum of C3d and both residues do not show any signs of chemical shift perturbation in the NMR titration experiments. The chemical shift perturbation analysis reported in this chapter supports the interaction observed in the crystal structure of the C3d/CD21 SCR1-2 complex recently solved by van den Elsen JM and Isenman DE. (van den Elsen JM and Isenman DE, 2011; PDB: 3OED). Residues on C3d that exhibited the strongest line broadening effect (*i.e.* set 1) form a binding surface to CD21 SCR1-2 in the new crystal structure of the protein complex (PDB: 3OED). These residues include Thr38, Glu39, Asp163, Ile164, Glu166, Glu167, Gln168 which form part of the negatively charged CD21 SCR1-2 binding surface as defined by mutagenesis studies (Clemenza L and Isenman DE, 2000; Hannan JP *et al.*, 2005; Bomb M, 2008; Chapter III), as well as Ile100, Ala101, Ile102, Asp103, Ser104 which form a conserved binding pocket that mediates binding to CD21 SCR1-2 and *S. aureus*' inhibitory protein Efb-C, Ehp and Sbi-IV (Hammel M *et al.*, 2007a; PDB: 2GOX; Hammel M *et al.*, 2007b; PDB: 2NOJ ; Clark EA *et al.*, 2011; PDB: 2WY8). Mutagenesis studies also indicate residues Asp36, Glu37, Glu160 and Asp292 on C3d are involved in the binding to CD21 SCR1-2, nevertheless it was not possible to assign these residues in the ^1H , ^{15}N -HSQC spectrum of C3d. The majority of residues that exhibited notable line broadening effect in mid titration point (*i.e.* set 2) form an extension of the contiguous binding surface for CD21 SCR1-2 which is not observed in neither of the crystal structure of C3d/CD21 SCR1-2 protein complex. These residues include Val32, His33, Trp41 sidechain, Gly59, Tyr60, Thr61, Gln62, Gln63 and Gln68. Interestingly, residues Trp84, Val131 and Ile137 also show notable line broadening effect in mid titration point (50% ligand saturation). These three residues constitute a small hydrophobic pocket spatially away from the CD21 SCR1-2 binding surface (Figure 5.3).

The binding event between C3d and CD21 SCR1-2 becomes more complex throughout the remainder of the NMR titration (75%, 100% and 200% ligand saturation). In addition to residues that already showed significant chemical shift perturbation in mid titration points (set 1 and set 2), residues Gly51, Ala52, Leu53, Ala75, Gln126, Gln249 on C3d (set 3) also experience line broadening effect. All residues from set 3 do not form a contiguous binding surface when mapped to crystal structure of C3d. Other chemical shifts in the spectrum also broaden away at this point, nevertheless it is not possible to map these residues on the crystal structure of C3d due to the lack of full backbone assignment of the ^1H , ^{15}N -HSQC spectrum. Based on these observations, it is speculated that the complexity of the binding interface between C3d and CD21 SCR1-2 arises due to non-specific interactions between the highly oppositely charged binding partners. C3d carries an overall negative charge (Clemenza and Isenman, 2000; Nagar B *et al.*, 1998) whereas CD21 SCR1-2 carries an overall positive charge (Guthridge JM *et al.*, 2001; Hannan JP *et al.*, 2005). As a result, an increase in protein concentration would favour charge driven non-specific interactions and hence affect residues that are outside of the contiguous binding surfaces between C3d and CD21 SCR1-2. Residues on C3d that began to show line broadening effect only in late titration points are not considered to form part of the physiologically relevant binding site to CD21 SCR1-2. Rather, these residues either appear to bind to CD21 SCR1-2 through non-specific interaction at high protein concentration, or broaden away due to an increase in apparent molecular weight of C3d towards formation of the full C3d/CD21 SCR1-2 protein complex.

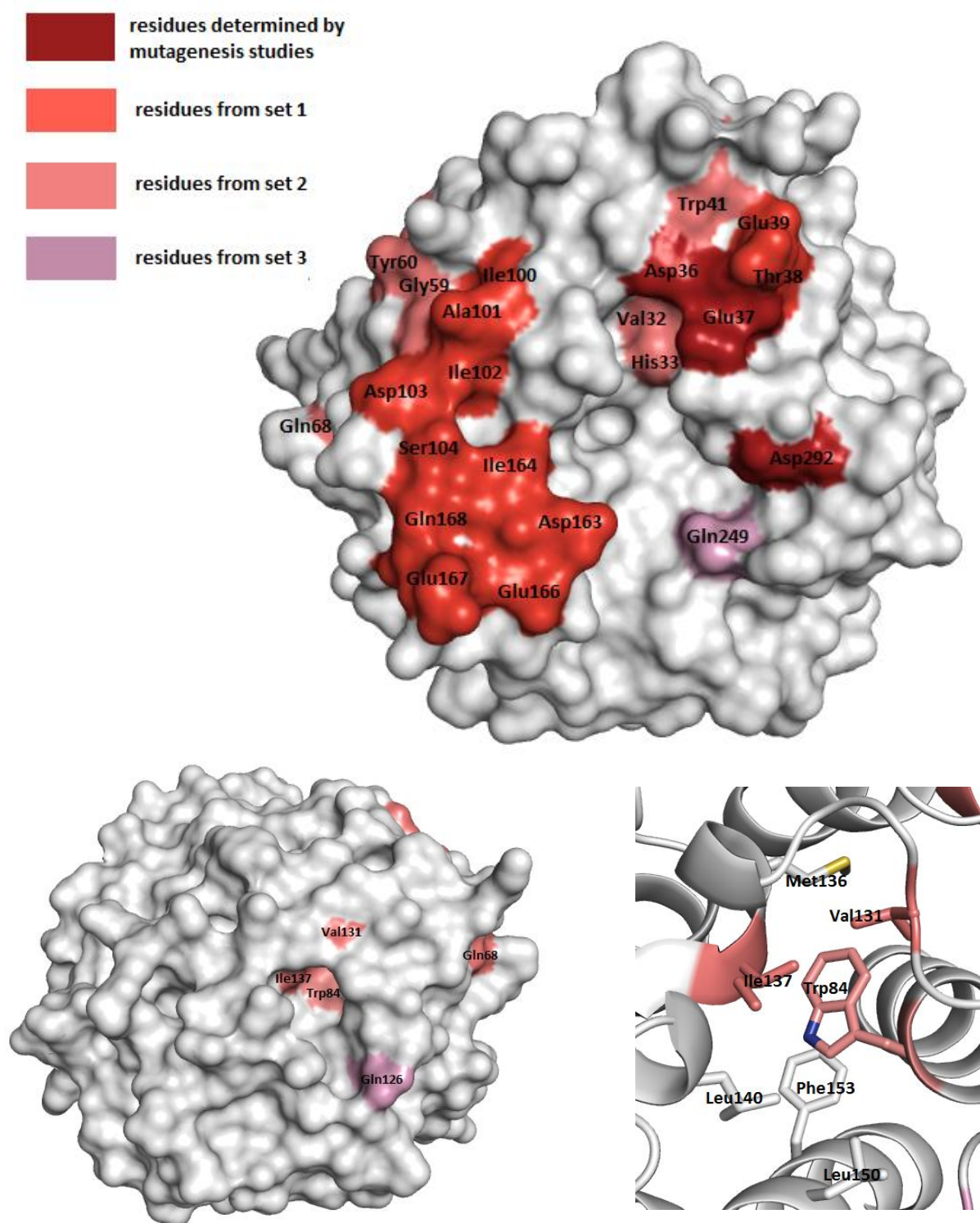


Figure 5.3 | The binding surfaces for CD21 SCR1-2 on C3d defined by ^1H , ^{15}N -chemical shift perturbation analysis

Chemical shifts of residues from set 1 are the most strongly perturbed during the course of titrations and they form a prominent binding site for CD21 SCR1-2. Residues from set 2 also show substantial changes in chemical shifts and a number of these residues form an extended binding surface next to the residues from set 1. Residues from set 3 show some changes in chemical shifts in late titration points and are probably not directly involved in the binding between C3d and CD21 SCR1-2. Asp36, Glu37 and Asp292 have been shown to be one of the key residues mediating the interaction between C3d and CD21 SCR1-2 in mutagenesis studies. However, it was not possible to assign them in the HSQC spectra of C3d.

CD21 SCR1-2 also appeared to interact with a small hydrophobic pocket on C3d which is on the opposite surface of the prominent binding site. Residues from this pocket that showed changes in chemical shifts were Trp84, Gln126, Val131 and Ile137 (set 2 and set 3).

5.2.2 THE INTERACTION BETWEEN C3d AND FACTOR H-DERIVED PEPTIDE

In the quest of low molecular weight inhibitors to disrupt the C3d/CD21 SCR1-2 interaction, a nonamer peptide was designed based on crystal structure of C3d/factor H SCR19-20 protein complex (Morgan HP *et al.*, 2011; PDB:3OXU). As reviewed in Chapter I, the development of complement-specific therapeutics proved to be a lot more challenging than initially anticipated. The complement activation cascade is mediated almost exclusively by large protein-protein interactions. Therefore, it is not surprising that complement therapeutics in the market and in the pipelines encompass biopharmaceuticals like antibodies, peptides and other soluble proteins (Ricklin D and Lambris JD, 2007). Therapeutic development of inhibitors of protein-protein interactions has proven difficult. Not only there is a lack of small molecule starting points for drug design, but also the target area on the surface of the proteins generally lacks definitive cleft for small molecules to bind. Typically, the contact surfaces involved in protein-protein interaction are approximately 1500 – 3000 Å² (Jones S and Thornton JM, 1996; Conte LL *et al.*, 1999), which are substantially larger than the typical range of 300 – 1000 Å² observed in protein – small molecule interaction (Smith RD *et al.*, 2006; Cheng AC *et al.*, 2007). Contacts in protein-protein interfaces therefore involve a large number of non-contiguous amino acid residues in the protein chains owing to broad interaction surfaces. As a result, short peptides derived from contiguous sequences at the binding interface generally failed to generate suitable chemical templates to initiate structure-based drug design programmes. However, availability of a crystal structure of C3d/factor H SCR19-20 protein complex sparks the idea of developing C3d-targeted peptide inhibitors. Much of the interaction between C3d and factor H was reviewed in Chapter I. It has been demonstrated that the SCR19 domain in factor H is responsible for binding to C3d, whereas the SCR20 domain is speculated to mediate contacts with the negatively charged concave surface on C3d via electrostatic steering effects.

Figure 5.4 shows the binding interface between C3d and factor H SCR20. Although the interaction between C3d and factor H SCR20 is not supported by an extensive network of molecular bonds, the negatively charged concave surface on C3d (*i.e.* binding surface for CD21 SCR1-2 and *S. aureus*' inhibitory proteins) is in close contact with a flexible loop in factor H SCR20, namely Cys1201 – Arg1206. It was speculated that a peptide derived from the loop may provide a starting point for development of peptide inhibitors to disrupt the

C3d/CD21 SCR1-2 interaction. This loop region in factor H SCR20 forms a negatively charged surface with residue R1203 penetrates deep in the positively charged concave surface on C3d. Hence, a nonamer peptide, VRFVAKRGY, derived from V1197 – Y1205 in factor H SCR20 was ordered (GenScript, NJ, USA) and tested for binding to C3d. Two amino acid substitutions were made to the original protein sequence, namely E1198 and C1201. The E1198 to arginine substitution was made to increase an overall positive charge of the peptide in hopes of increasing the electrostatic interaction between C3d and the peptide. The C1201 to alanine substitution was made to remove dimerization of the peptide and any potential thiol reaction mediated by free cysteine.

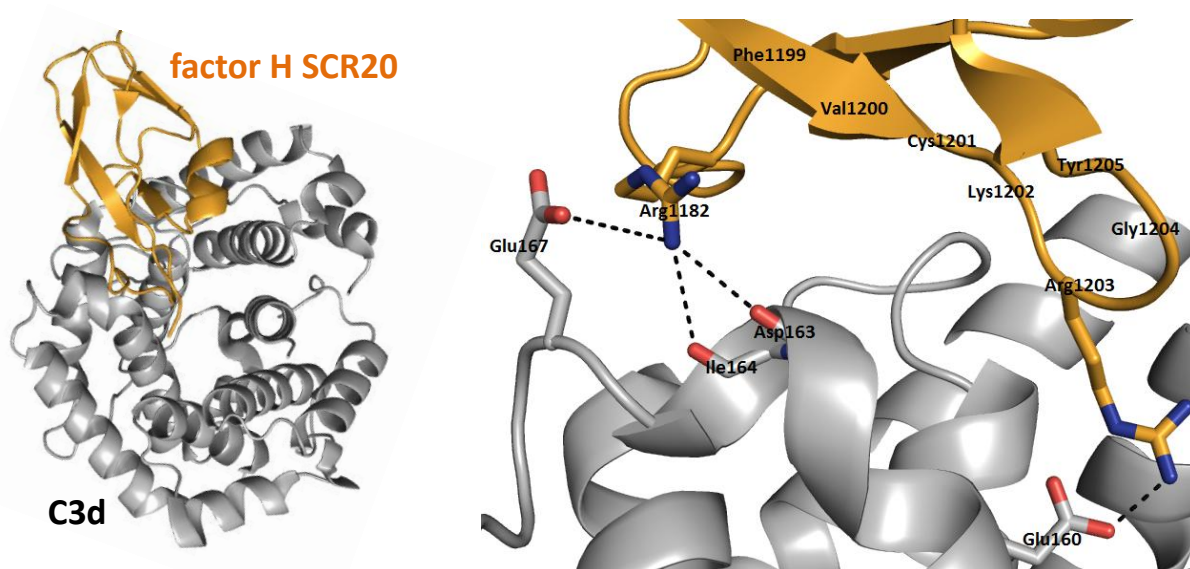


Figure 5.4 | Rationale behind the design of a potential C3d-binding factor H-derived peptide

C3d has been shown to form crystal contacts with the SCR20 domain of factor H (left; PDB:3OXU). A zoom-in view of the interaction interface (right) shows that Arg1203 on factor H SCR20 penetrates deeply into the CD21 SCR1-2 binding pocket on C3d. It was postulated that a peptide derived from a linear sequence on factor H SCR20, V1197 – Y1205, could be developed as a C3d-binding peptide to disrupt the interaction between C3d and CD21 SCR1-2.

(i) SPR AND ITC ANALYSIS

SPR experiments were carried out to test for binding between C3d and factor H-derived peptide in 20 mM NaH_2PO_4 , 150 mM NaCl, 0.005% P20, pH6.8 at 25°C. A C3d construct, A17C, was immobilized onto a CM4 sensor chip to a density of approximately 6000 RU by thiol coupling (Schuck P, 1997). An SPR sensorgram of the interaction (Figure 5.5) clearly shows that the peptide is binding to the surface of C3d in a saturable manner. The binding of the peptide displays fast-on, fast-off kinetics, which is also characteristic of the interaction between C3d and CD21 SCR1-2 (Chapter III). The K_D was estimated by plotting the binding response at equilibrium against the concentration of factor H-derived peptide and fitting a Langmuir binding isotherm (Figure 5.5). The interaction has an affinity of 0.74 mM and a 1:1 stoichiometry.

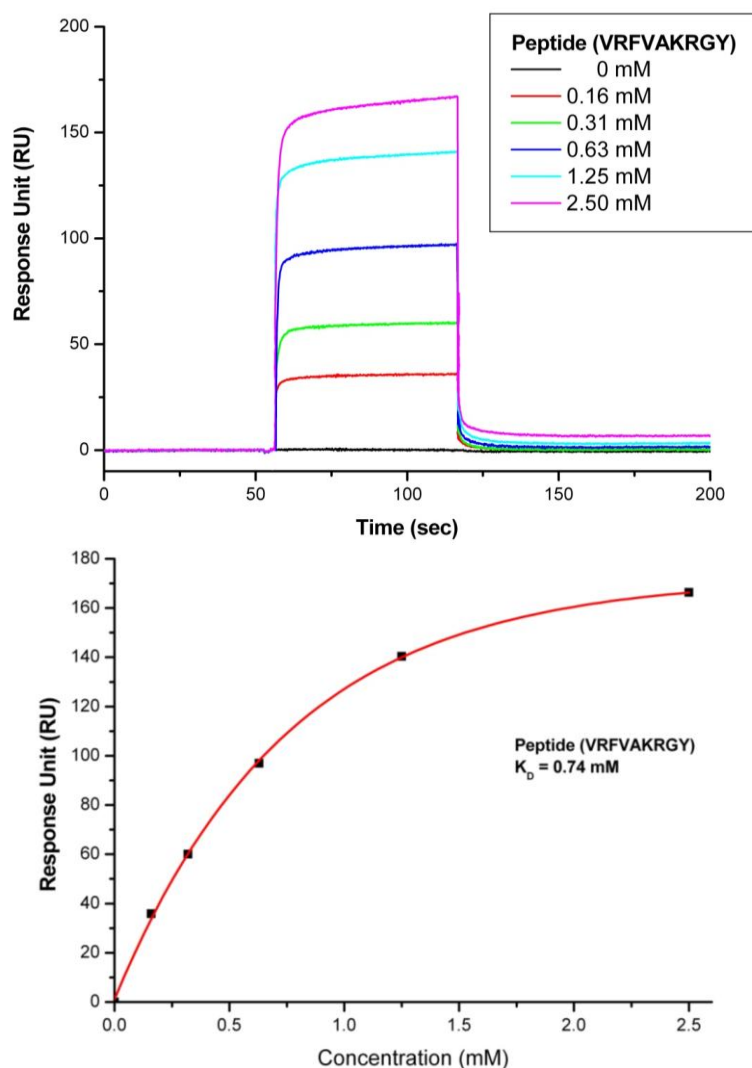


Figure 5.5 | SPR analysis of the interaction between C3d and factor H-derived peptide

The binding of factor H-derived peptide to immobilized C3d was determined over a range of ligand concentrations. Representative sensorgrams are shown at the top and the interaction displays a fast-on, fast-off kinetics. Plotting the binding response against ligand concentration (bottom) shows that the binding is saturable and the K_D for the interaction between C3d and factor H-derived peptide is 0.74 mM.

ITC experiment was carried out to further investigate the interaction between C3d and factor H-derived peptide. As mentioned in Chapter III, ITC experiments are most favourable when the dimensionless parameter c is in the range of 10–100 (Wiseman T *et al.*, 1989). There has been growing interest in evaluating binding events that fall in a low- c regime which is characteristic of low affinity system, typically with $c < 1$ (Turnbull WB and Daranas AH, 2003). Examples of low affinity systems that have been successfully characterized by ITC include carbohydrate-protein interactions, fragment-based drug design, protein-ligand and protein-protein interactions (Rekharsky MV and Inoue Y, 2002; Dam TK and Brewer CF, 2002; Zhang YL and Zhang ZY, 1998; Weber PC and Salemme FR, 2003; Erlanson DA, 2006). There is a strong correlation between ΔH and stoichiometry number n in the analysis of low- c data (Tellinghuisen J, 2005). Hence, it is not possible to determine both of these simultaneously. The usual solution is to fix n in the analysis as it is often known from prior knowledge of the binding interaction and precision in the prepared sample concentrations. On a typical ITC binding curve, ΔH influences the intercept on the y-axis and the binding stoichiometry n governs the position of the inflection point. The overall shape of the curve is dependent on both the binding constant K_a , as well as the concentration of binding sites $n[M_T]$. For $c > 10$, the curve is clearly sigmoidal (Chapter III, Figure 3.8) and allows simultaneous determination of ΔH , K_a and n . For a low affinity system with $K_D \sim 1$ mM, a c -value of 10 would require the initial receptor concentration of 10 mM and the ligand concentration in the syringe usually require ten times the receptor concentration in the sample cell. As a result, protein and ligand solubility becomes the principle limitations for evaluating low affinity systems using conventional ITC experimental design. Direct low c -value titrations can be performed as long as care is taken in designing the experiment to record an adequate portion of the binding curve with good levels of signal-to-noise. When an interaction falls under the low c regime, a ligand concentration five times higher than K_D is required to approach receptor saturation (Turnbull WB, 2011). The binding curve does become poorly defined as the c -value decreases below one, nonetheless the curve is not entirely flat. Figure 5.6 shows a direct ITC titration experiment of factor H-derived peptide against C3d. The experiment was performed in 20 mM NaH_2PO_4 , 150 mM NaCl, pH 6.8 at 20 °C. A hyperbolic curve was obtained as expected with low c value and the K_D of the interaction is 1.0 ± 0.3 mM, which is in agreement with the SPR experiment.

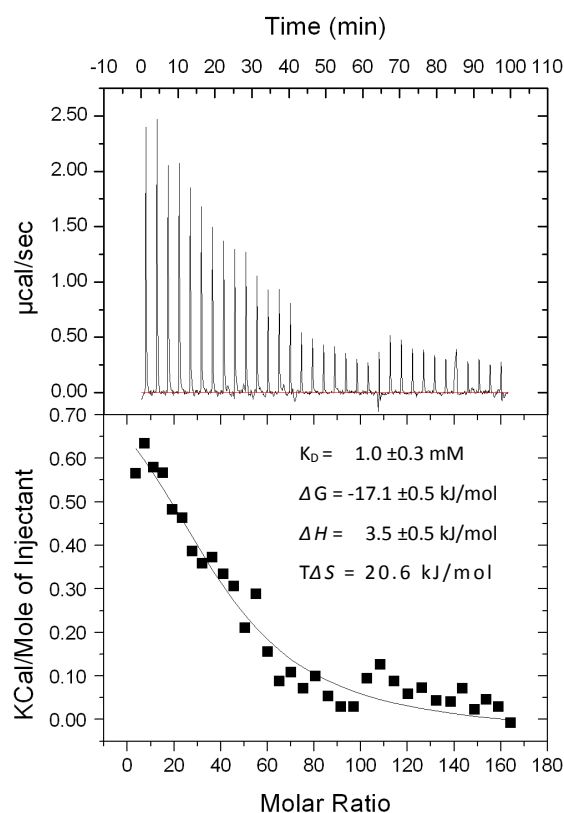


Figure 5.6 | ITC analysis of the interaction between C3d and factor H-derived peptide

The top part of the diagram shows the measured experimental data and the bottom part represents the integrated areas of each injection plotted against the molar ratio of ligand titrated. Response from the control 'heat of dilution' experiment (*i.e.* titration of ligand into plain buffer) has been subtracted.

Table 5.3 | Thermodynamic parameters of the interaction between C3d/CD21 SCR1-2 and C3d/factor H-derived peptide in 20 mM NaH₂PO₄, 150 mM NaCl, pH 6.8

Protein - Ligand	K_D (μ M)	ΔG (kJ/mol)	ΔH (kJ/mol)	$-T\Delta S$ (kJ/mol)
C3d/factor H-derived peptide	1000 $\pm$ 300	-17.1 $\pm$ 0.5	+3.5 $\pm$ 0.5	20.6
C3d/CD21 SCR1-2	0.92 $\pm$ 0.07	-34.4 $\pm$ 2.7	+48.2 $\pm$ 0.9	82.6

The thermodynamic signature of the C3d/factor H-derived peptide interaction shows a small degree of similarity to the interaction between C3d and CD21 SCR1-2. The binding event is endothermic as seen in the raw ITC data in Figure 5.6, which indicates an uptake of heat upon injection of ligand. The interaction is enthalpically unfavourable as indicated by a positive ΔH value and is predominately entropically driven by a favourable entropy change of ~ 20 kJ/mol. Table 5.3 showed a comparison of the thermodynamic signature of the interaction between C3d/CD21 SCR1-2 and C3d/factor H-derived peptide recorded in physiological relevant ionic strength. Both interactions are strongly entropically-driven, implying the C3d/peptide interaction is likely to be dominated by either electrostatic or hydrophobic binding events just like the C3d/CD21SCR1-2 interaction.

(ii) CHEMICAL SHIFT PERTURBATION ANALYSIS

NMR based chemical perturbation studies were used to assess the binding interaction between C3d and factor H-derived peptide. Unlabelled factor H-derived peptide was titrated against uniformly ^{15}N -labelled C3d. Seven datasets were collected with peptide concentrations ranging from 0 mM to 4 mM. The experiments were performed in 20mM NaH_2PO_4 , 150 mM NaCl, pH6.8 As seen in Figure 5.7, there are a few detectable changes in the ^1H , ^{15}N -HSQC of C3d in the presence of the peptide ligand. Over the course of this titration, ten residues show significant chemical shift changes, namely the backbone amides for Trp84, Val90, Lys120, Gln127, Ala129, Val131, Ile132, Glu135, Ile137, Gln143 and the sidechain of Trp84. All residues on C3d that mediate binding to factor H-derived peptide display fast exchange characteristics. Figure 5.8 shows a selected zoom-in region of the ^1H , ^{15}N -HSQC spectrum of C3d to illustrate residues that exhibit significant changes in chemical shift ($\Delta\delta_{\text{HN}} \geq 0.08$ ppm) upon the addition of the peptide. Saturation binding curves based on change in chemical shift position for two residues, Gln127 and sidechain of Trp84, were plotted, giving a K_D of 0.75 mM and 1.00 mM, respectively (Figure 5.9). Determining the average and standard deviations of K_D from these two residues showed that the overall K_D of the interaction is 0.88 ± 0.18 mM. The result is in good agreement with SPR and ITC data.

Mapping residues that are perturbed upon ligand addition on the crystal structure of C3d showed the peptide is binding weakly to a small hydrophobic pocket spatially away from the CD21 SCR1-2 binding surface (Figure 5.10). A few residues forming this small hydrophobic pocket are also perturbed in the C3d/CD21 SCR1-2 NMR titration experiments, namely Trp84, Val131 and Ile137. Although both CD21 SCR1-2 and factor H-derived peptide appear to interact weakly with this hydrophobic pocket, residues found in this pocket do not form the prominent binding site for CD21 SCR1-2. Given the factor H-derived peptide does not bind to the prominent binding site for CD21 SCR1-2 as anticipated, it will not be exploited further as an inhibitor to disrupt the interaction between C3d and CD21 SCR1-2.

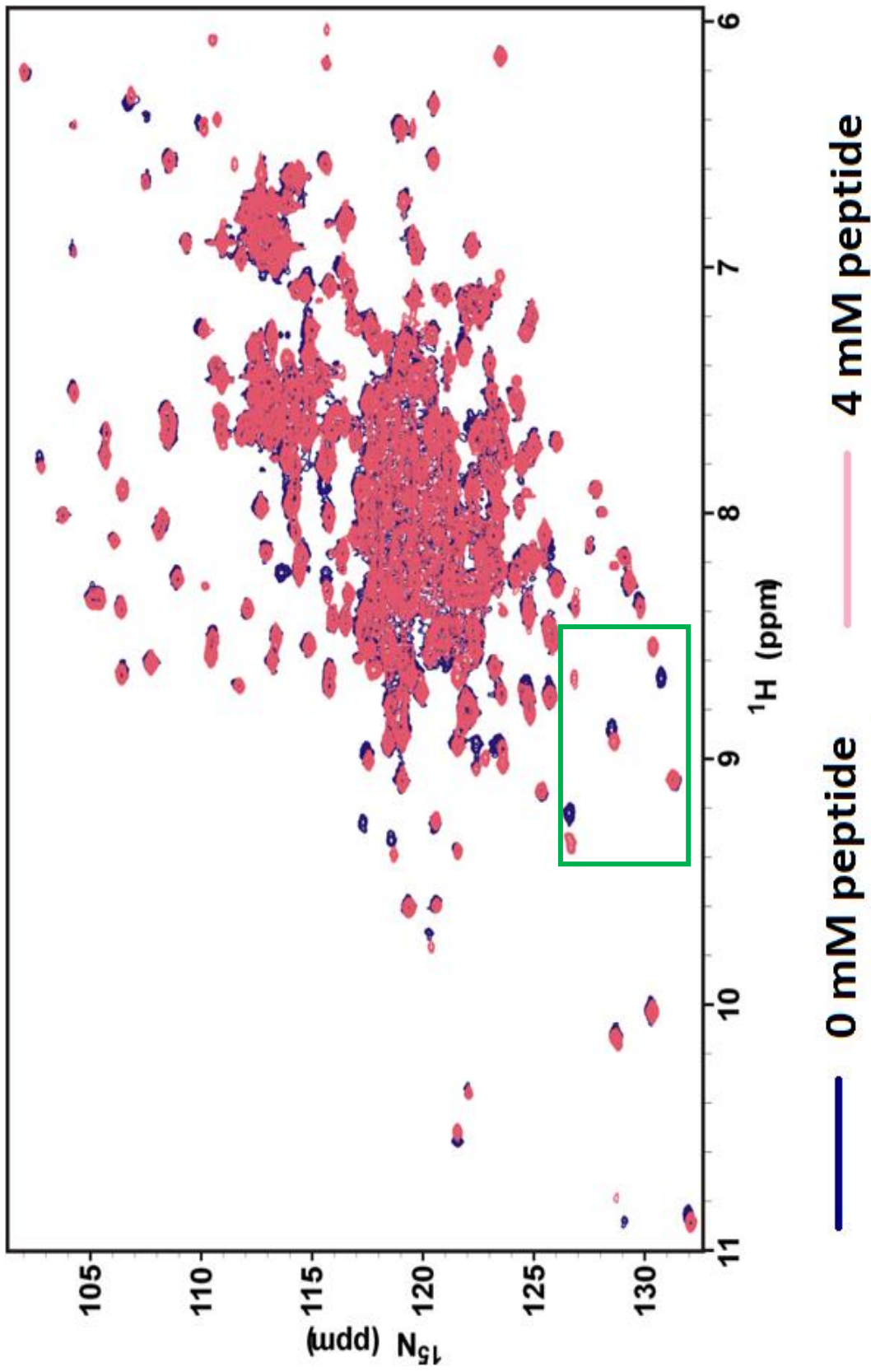


Figure 5.7 ^1H , ^{15}N -chemical shift perturbation experiments to map the interaction surfaces on C3d for factor H-derived peptide

The titration experiments were performed in 20 mM NaH_2PO_4 , 150 mM NaCl , pH 6.8. Overlapping spectra collected in the absence and the presence of 4 mM factor H-derived peptide show a number of residues are involved in the binding event. Fortunately, residues exhibiting changes in chemical shifts were mostly found in a non-crowded region of the spectra. It is thus possible to use the limited information from the chemical shift perturbation analysis to map the binding site for the factor H-derived peptide on C3d.

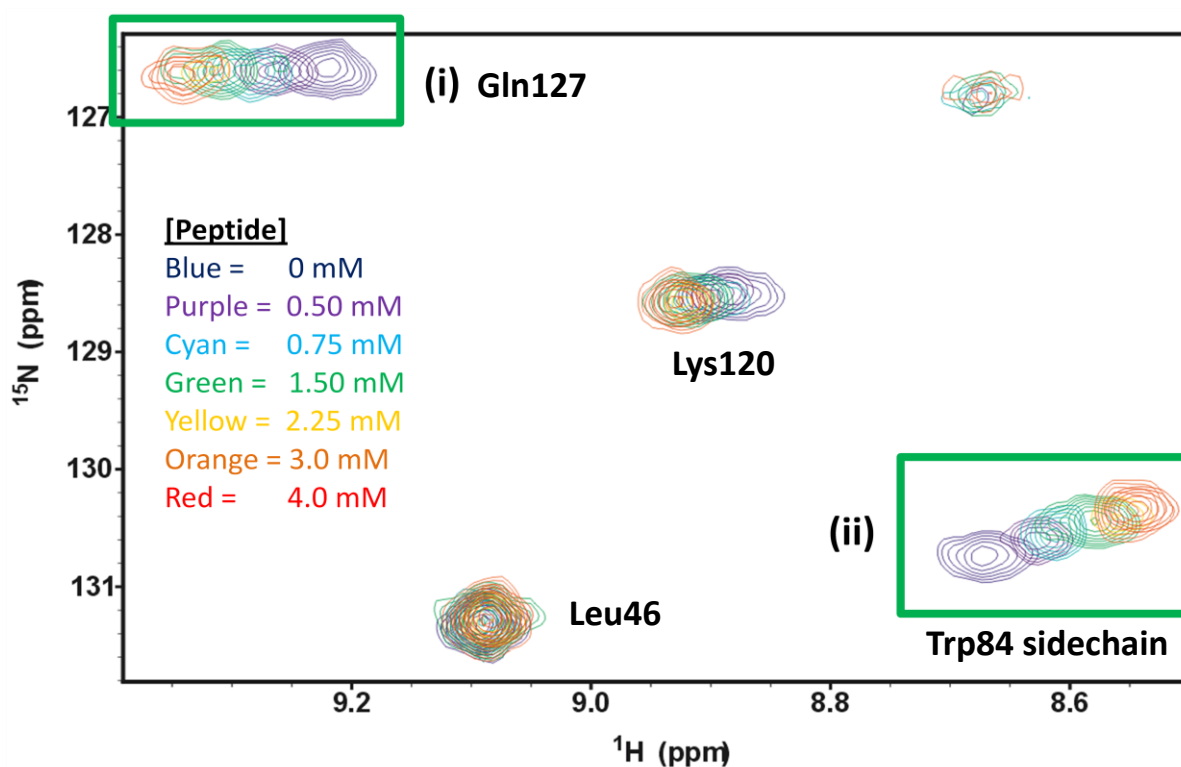
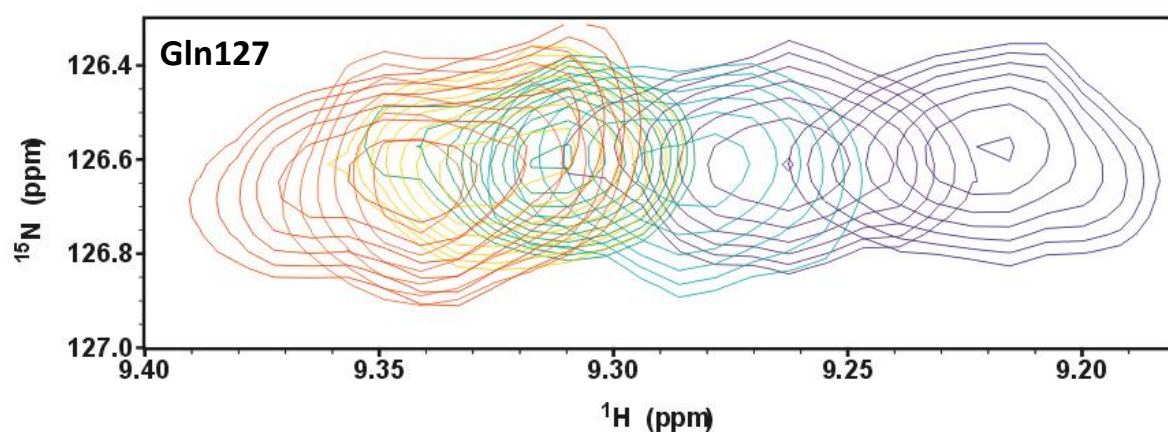


Figure 5.8 Inserts showing expanded views of the indicated areas in the ^1H , ^{15}N -chemical shift perturbation experiments

Different titration points were overlaid and chemical shifts that were able to be assigned to specific residues in C3d were annotated. The chemical shifts corresponding to Leu46 is unaffected over the course of the titration experiments, suggesting this residue is not involved in the binding to factor H-derived peptide. On the other hand, Gln127, Lys 120 and Trp84 sidechain show substantial chemical shift changes of fast exchange characteristics. Plotting the change in proton chemical shifts of Gln127 and Trp84 sidechain against ligand concentrations allow determination of the K_D between C3d and factor H-derived peptide (see insert figures on the following pages).

(i)



Concentration of factor H-derived peptide

Blue = 0 mM

Purple = 0.50 mM

Cyan = 0.75 mM

Green = 1.50 mM

Yellow = 2.25 mM

Orange = 3.0 mM

Red = 4.0 mM

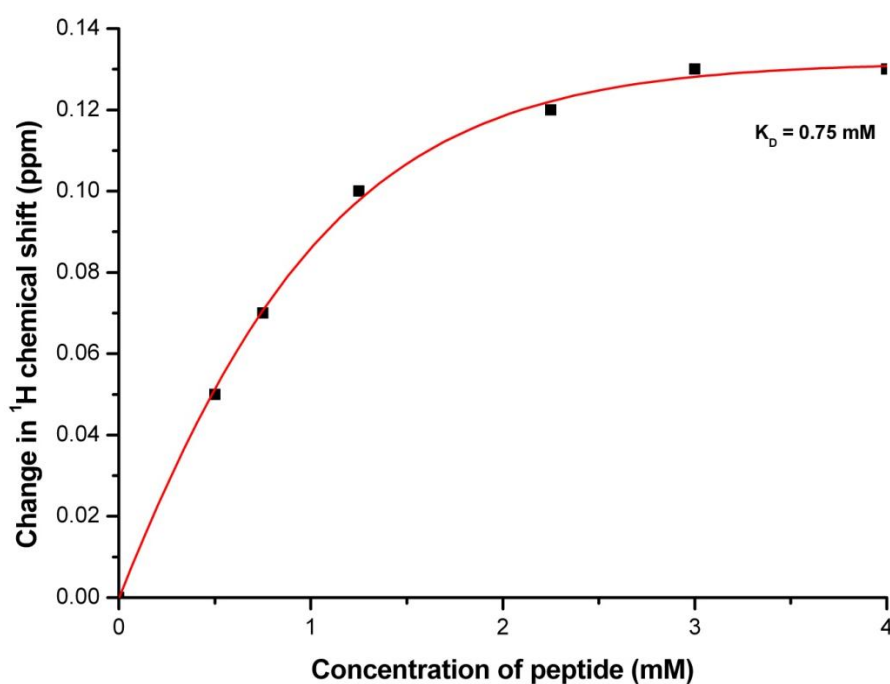
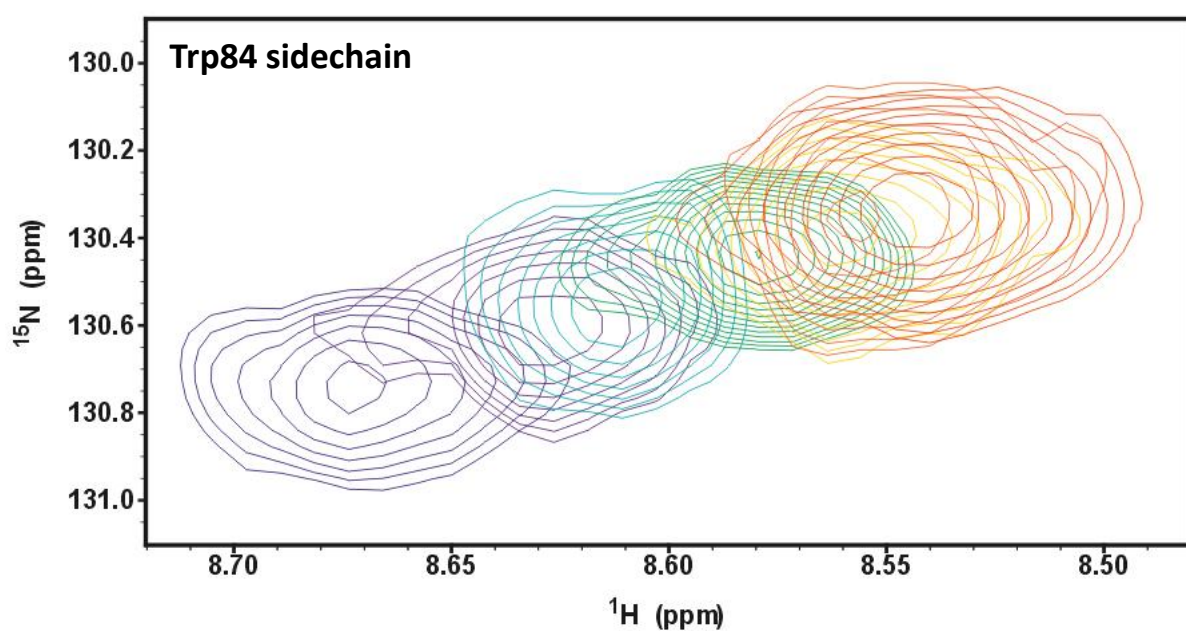


Figure 5.9 | Saturation binding curves based on change in proton chemical shift position of Gln127

- (i) The proton chemical shift of Gln127 was perturbed significantly upon incremental addition of factor H-derived peptide. Plotting change in proton chemical shift against ligand concentrations showed that the binding of C3d to factor H-derived peptide is saturable with a K_D of 0.75 mM.

(ii)



Concentration of factor H-derived peptide

Blue = 0 mM Purple = 0.50 mM Cyan = 0.75 mM Green = 1.50 mM
 Yellow = 2.25 mM Orange = 3.0 mM Red = 4.0 mM

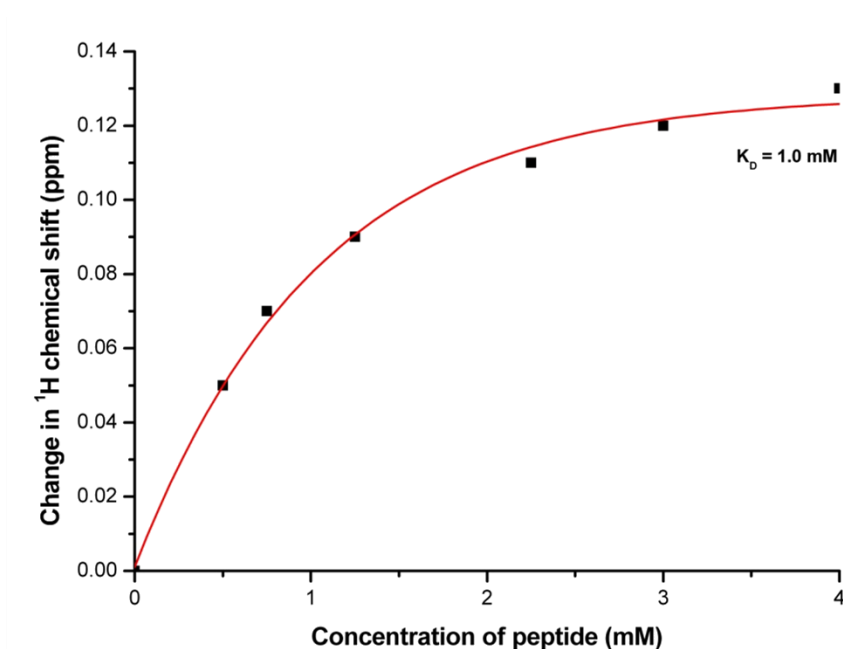


Figure 5.8 | Saturation binding curves based on change in proton chemical shift position of Trp84 sidechain

(ii) The proton chemical shift of Trp84 sidechain was perturbed significantly upon incremental addition of factor H-derived peptide. Plotting change in proton chemical shift against ligand concentrations showed that the binding of C3d to factor H-derived peptide is saturable with a K_D of 1.0 mM.

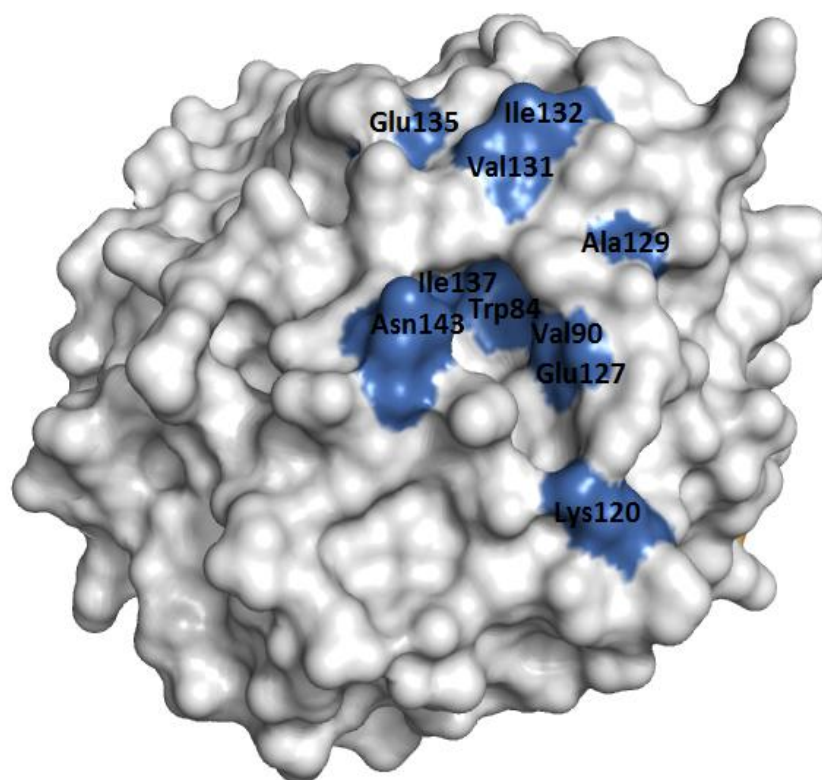


Figure 5.10 | The binding surfaces for factor H-derived peptide on C3d defined by ^1H , ^{15}N -chemical shift perturbation analysis

Residues that were perturbed upon addition of factor H-derived peptide were mapped on the crystal structure of C3d in sky blue. The peptide appeared to interact with a small hydrophobic pocket on C3d and a number of residues surrounding this pocket. The binding surface for factor H-derived peptide defined by NMR titration experiments is located spatially away from the prominent CD21 SCR1-2 binding site on C3d.

5.3 DISCUSSION

NMR chemical shift perturbation analysis presented in this chapter confirms the extent of the physiologically relevant CD21 SCR1-2 binding site on C3d, albeit the assignment of ^1H , ^{15}N -HSQC spectra of C3d has only achieved partial completion owing to technical difficulties. The prominent binding site for CD21 SCR1-2 on C3d defined by NMR studies is in good agreement with mutagenesis studies and biophysical characterisation detailed in Chapter III. Both mutagenesis data (Chapter III; Clemenza L and Isenman DE, 2000) and NMR interaction analysis have pointed out that the CD21 CR1-2 binding surfaces on C3d span a surface area of approximately $30 \times 30 \text{ \AA}$ with two negatively charge patches surrounding a concave hydrophobic groove. The

proteins appear to bind strongly to a single contiguous binding site under physiologically relevant ionic strength in solution. The binding site for C3d on CD21 SCR1-2 defined by mutagenesis and NMR studies in solution (Bomb M DPhil thesis, 2008; Kovacs JM *et al.*, 2009) show a high degree of charge and shape complementarity to the CD21 SCR1-2 binding site on C3d as described in Chapter III. Furthermore, the CD21 SCR1-2 binding site is located on the opposite side to the antigen attachment site (Nagar B *et al.*, 1998) when modelled onto the crystal structure of C3d. These features certainly allow C3d to function as a bridge between innate and adaptive immune response (Chapter I; Ahearn JM *et al.*, 1996; Dempsey PW *et al.*, 1996) by simultaneously binding to antigen and CD21 SCR1-2 exploiting discrete binding surfaces. Mutagenesis studies, microcalorimetric and NMR interaction analysis presented in this thesis provide three separate lines of evidence to conclude that the binding interface observed in crystal structure of the C3d/CD21 SCR1-2 protein complex solved by Szakonyi G *et al.* in 2001 (PDB:1GHQ) is entirely a result of crystallization artefact. Collectively, our data support the binding mode described in Chapter III and are consistent with the idea of the interaction between C3d and CD21 SCR1-2 being dominated by either electrostatic or hydrophobic binding events. This binding mode, as suggested by docking programme ClusPro with user-defined binding restraints, is in excellent agreement with a new crystal structure of C3d/CD21 SCR1-2 protein complex solved by van den Elsen JMH and Isenman DE in 2011 (PDB:3OXU) and will be discussed further in Chapter VI.

The C3d/CD21 SCR1-2 interaction exhibits strong ionic strength dependency and the K_D is reduced at high salt concentrations (Chapter III; Diefenbach RJ and Isenman DE, 1995; Guthridge JM *et al.*, 2001). However, ionic strength sensitivity of the interaction between C3d and CD21 SCR1-2 is not reflected in NMR chemical shift perturbation analysis. The K_D of the interaction in 150 mM and 400 mM NaCl buffer are $0.92 \pm 0.07 \mu\text{M}$ and $4.89 \pm 0.07 \mu\text{M}$ as determined by ITC in Chapter III, respectively. Hence, raising the concentration of NaCl from 150 mM to 400 mM in NMR titration experiments does not alter the intermediate exchange characteristic of this interaction, which is typically observed when affinity of a binding interaction falls between $100 \text{ nM} < K_D < 10 \mu\text{M}$. While residues on C3d that are perturbed upon the addition of CD21 SCR1-2 in 150 mM NaCl are largely the same as those observed in titration experiments performed in 400 mM NaCl, experiments performed in high salt condition offer incremental benefits in spectral sensitivity and reduction in chemical

shift overlaps. These observations suggest C3d is prone to ionic strength dependent self-aggregation and it exploits the same residues to mediate the protein-protein interaction between CD21 SCR1-2 even beyond physiologically relevant ionic strengths. Nevertheless, the binding interface becomes more complex at high protein concentrations, most likely attributable to non-specific, charge driven interactions between the overall negatively charged C3d and positively charged CD21 SCR1-2. These charge driven non-specific interactions observed *in vitro* are unlikely to occur *in vivo* as C3d circulates in plasma at 1.3 mg/ml (Correia JJ and Detrich W, 2008), *i.e.* 37 μM , implying C3d is normally monomeric under physiologically relevant conditions. Given the fact that the concentration of neither protein is unlikely to reach 200 μM in body plasma, the prominent binding site (*i.e.* hydrophobic groove surrounded by two negatively charged patches on C3d and hydrophobic core surrounded by positively charged residues on CD21 SCR1) identified by mutagenesis, NMR and recent crystallographic data is expected to be the sole interaction interface between C3d and CD21 SCR1-2 *in vivo*. Although the nonamer factor H-derived peptide appears to interact with C3d in a 1:1 stoichiometry and the K_D of the interaction is 0.88 ± 0.18 mM, NMR chemical perturbation analysis clearly shows that the peptide occupies a small hydrophobic pocket located spatially away from the CD21 SCR1-2 binding site on C3d. Hence, the peptide does not serve as an appropriate starting point to develop potent inhibitors to disrupt the C3d/CD21 SCR1-2 interaction. Future work could include screening C3d against commercial phage display libraries to identify original C3d-binding peptide.

CHAPTER VI

CRYSTALLIZATION TRIALS OF THE C3d/CD21 SCR1-2 PROTEIN COMPLEX

6.1 STRUCTURE DETERMINATION OF RECOMBINANT C3d

Complement component C3 is composed of two disulfide linked polypeptide chains, α chain (115 kDa) and β chain (75 kDa), with a total mass of 190 kDa and circulates in plasma in a biologically inert form at 1 – 1.6 mg/ml (Müller-Eberhard HJ, 1988). Activation of the complement system leads to proteolytic cleavage of a single peptide bond in the α chain of C3 by C3 convertases to give complement fragment C3a and C3b. Complement fragment C3b is composed of the resulting α' chain and β chain. The cleavage of C3 is also accompanied with a conformational change that leads to the exposure of an intramolecular thioester bond formed between the side chain of Cys988 and Gln991 in C3b (*i.e.* Cys17 and Gln20 in C3d), which is responsible for the thioester linkage of C3b to nearby pathogenic antigens. This type of bond is only found in two other plasma proteins, namely complement C4 and α_2 -macroglobulin (Isenman DE, 1983). The thioester

bond allows C3b to attach covalently to target molecules on pathogenic surfaces through ester or amide linkages (Bhagavan NV, 2002). In native C3, the thioester bond is formed by a transacylation reaction between the thiol group of Cys988 and the γ -amide group of Gln991, and the bond is protected within a hydrophobic pocket (Figure 6.1). Cleavage of C3 allows transient exposure of the thioester bond with a half-life of approximately 100 μ s (Lambris JD and Holers VM, 2000) to participate in a transacylation reaction with nucleophilic groups present on cell surface antigens, such as hydroxyl groups on glycoconjugates and amines on protein surfaces. The attachment of C3b to antigens is nonspecific and activated C3b with an exposed thioester moiety in solution is rapidly hydrolyzed by water in a competing inactivation step to prevent covalent attachment to self antigens (Law SKA and Dodds AW, 1997). Antigen-bound C3b is subjected to proteolytic degradation and C3d is the final protease-resistant degradation product of C3b. An amino acid substitution of C17 to A17 was introduced in the recombinant C3d molecule of 35 kDa. The substitution removes the reactive thioester antigen attachment site originally at C17 and ensures correct disulfide bonding of two other native cysteines during the expression of recombinant C3d in *E. coli*. Although the crystal structure of human C3d has already been solved at 1.8 Å resolution (Nagar B *et al.*, 1998), crystallization trials in thesis work identified two different crystallization conditions of C3d using commercially available screens. Obtaining reproducible diffraction quality C3d crystals provided an efficient fragment-based drug discovery platform for subsequent small-molecule soaking experiments.

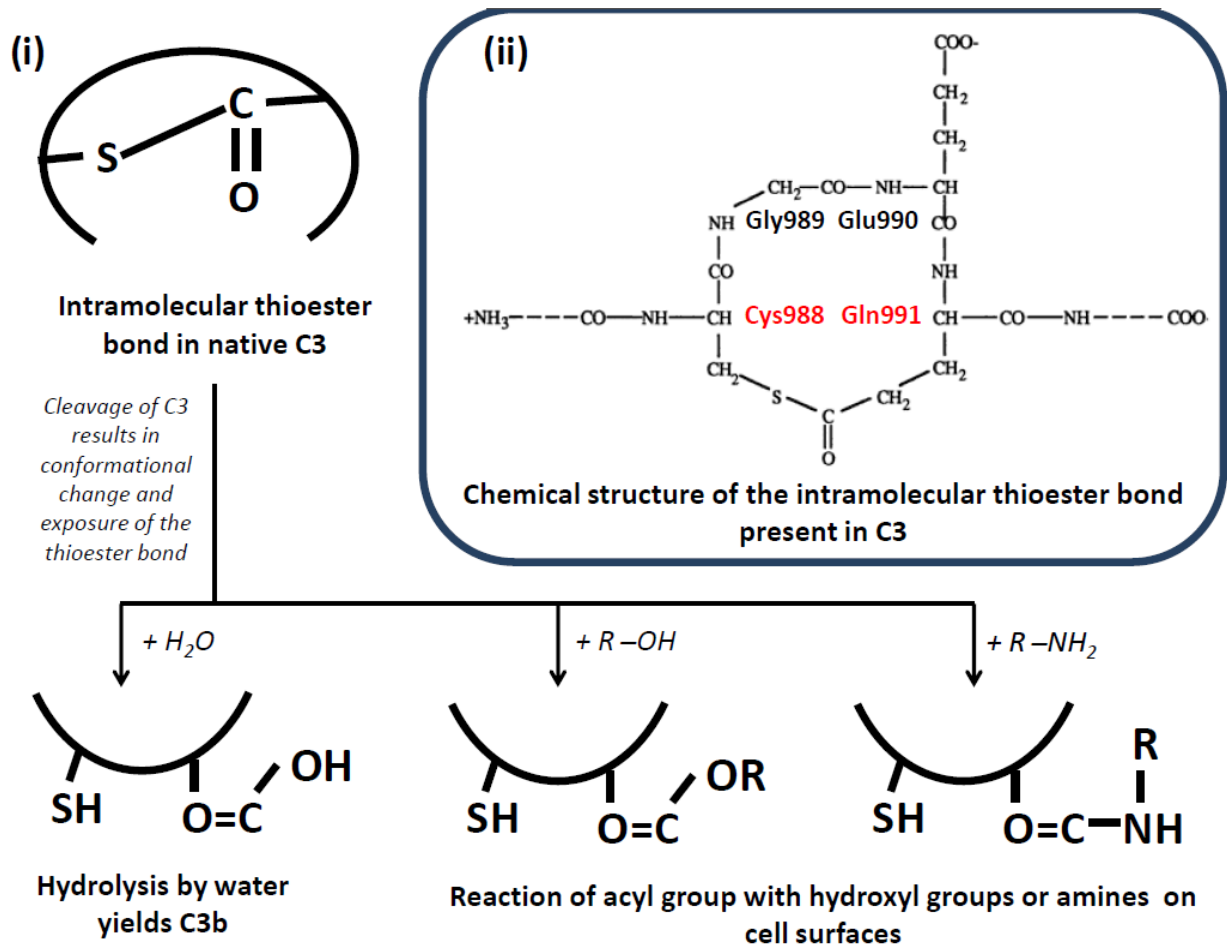


Figure 6.1 | Intramolecular thioester bond in complement C3

- (i) Complement activation triggers the proteolytic cleavage of the α -chain of C3 and induces a conformational change in C3b. The thioester bond is thus exposed and becomes extremely reactive with nearby water and other nucleophiles. The activated C3b with an exposed thioester moiety could be rapidly hydrolysed by water to form an inactivated form of C3b, or could react with hydroxyl groups and amines on cell surface antigens.
- (ii) Chemical structure of the intramolecular thioester bond present in native C3 (picture adapted from Rother K, Till GO and Hansch GM, 1998). The 15-membered thiolactone ring is comprised of four residues, Cys988-Gly989-Glu990-Gln991, with the thioester bond forming between Cys988 and Gln991. The chemical structure and composition are conserved in most C3, C4 and α_2 -macroglobulins over a diverse range of species, except occasionally the glycine may be substituted by alanine in some cases.

C3d crystals were grown in two different crystallization conditions— 0.2 M potassium nitrate, 20 w/v polyethylene glycol (PEG) 3350, as well as 10%w/v PEG 1000, 10% PEG 8000. Crystals that were grown in the former conditions diffracted to 2.0 Å resolution in $P6_1$ space group, whereas those that were grown in the later conditions diffracted to 1.6 Å in $P2_12_12_1$ space groups (Table 6.1). Regardless of the space groups, the overall molecular architecture of C3d solved in the thesis work is almost identical to the published C3d crystal structure in $P2_12_12_1$ space group (PDB:1C3D) when the two structures are superimposed in the asymmetric unit. Superposition of our C3d structure in $P6_1$ space group and 1C3D results in r.m.s.d. values of 0.29Å for 1176 backbone atoms and 0.35 Å all 2317 atoms. Likewise, superposition of our C3d structure in $P2_12_12_1$ space group and 1C3D results in r.m.s.d. values of 0.27Å for backbone atoms and 0.33 Å for all atoms. It appears C3d adopts a fairly rigid structure so that different crystallization conditions and crystal packings do not have any significant effect on the overall molecular architecture of C3d. The positively charged concave CD21 SCR1-2 binding surface defined by mutagenesis studies appeared not as solvent accessible in the $P6_1$ crystal form of C3d than the $P2_12_12_1$ crystal form (Figure 6.2). Hence, C3d crystals in the $P2_12_12_1$ crystal form were used in subsequent small-molecule soaking experiments.

Table 6.1 | Summary of data collection and refinement statistic

	C3d ($P6_1$)	C3d ($P2_12_12_1$)
Resolution	2.0 - 31.48	1.58 - 39.33
Space group	$P6_1$	$P2_12_12_1$
Unit cell dimension	a=54.64 b=54.64 c=168.64 $\alpha=90.00 \beta=90.00 \gamma=120.00$	a=59.93 b=65.02 c=87.25 $\alpha=90.00 \beta=90.00 \gamma=90.00$
Number of reflection (high resolution shell)	102412 (14939)	182690 (26553)
Rmerge	0.081 (0.524)	0.056 (0.402)
I/δI	14.4 (3.1)	14.6 (3.6)
Completeness	100 (100)	99.4 (99.9)
Multiplicity	5.3 (5.3)	3.9 (3.9)
R factor	0.24	0.20
Rfree	0.27	0.23
Rmsd angle	1.832	1.141
Rmsd bond	0.021	0.0084

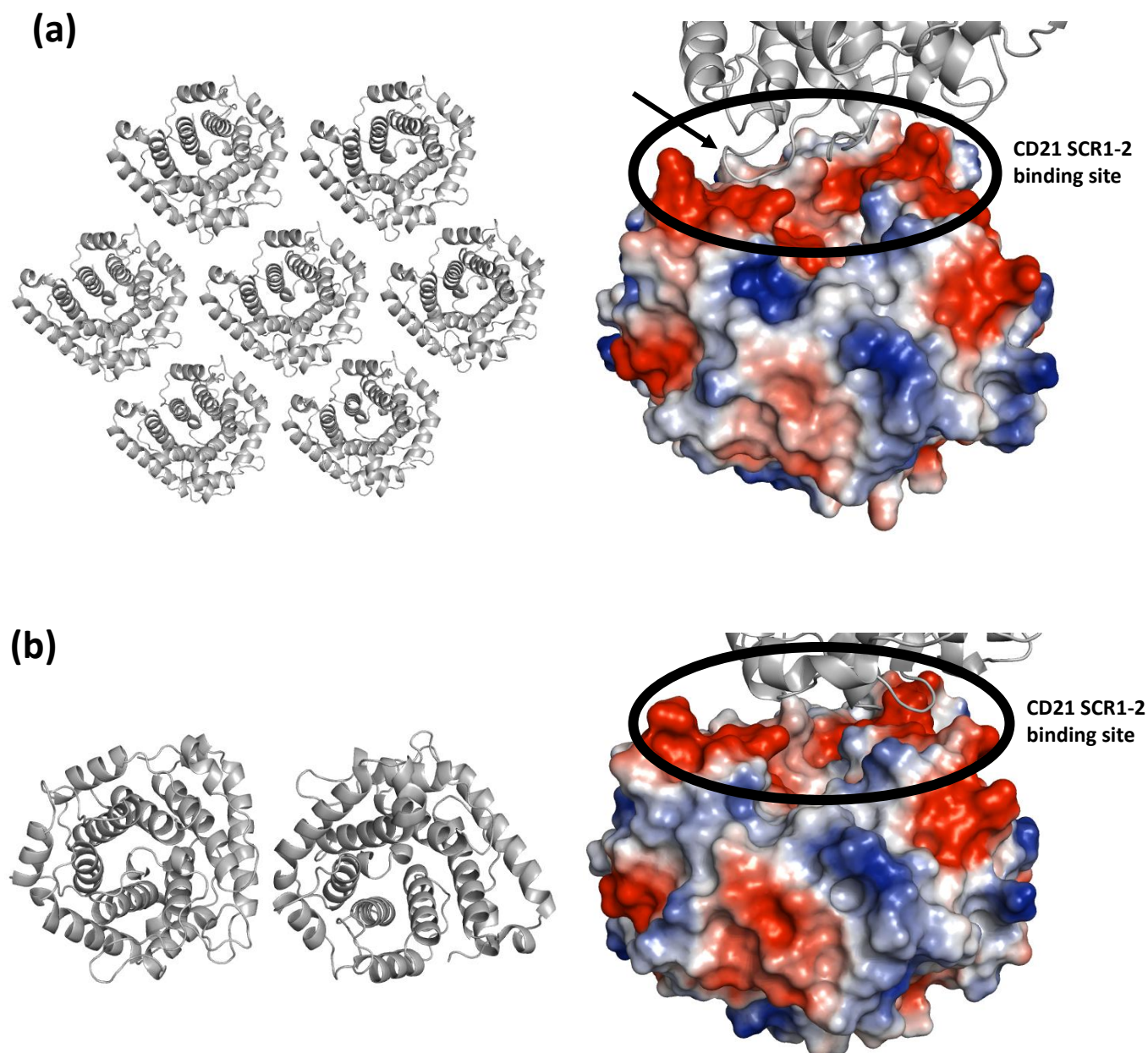


Figure 6.2 | Crystal structure of C3d in $P6_1$ and $P2_12_12_1$ spacegroups

- (a) C3d in $P6_1$ crystal form. C3d is shown as electrostatic surface representation on the right with another C3d molecule in a symmetry cell shown as cartoon to illustrate crystal packing. The CD21 SCR1-2 binding site is circled and is partly obstructed by a loop region of another C3d molecule (arrow) due to crystal packing.
- (b) C3d in $P2_12_12_1$ crystal form. C3d is shown as electrostatic surface representation on the right with another C3d molecule in a symmetry cell shown as cartoon to illustrate crystal packing. The CD21 SCR1-2 binding site is circled and is more solvent accessible than in the $P6_1$ crystal form.

6.2 CRYSTALLIZATION TRIALS OF THE PROTEIN COMPLEX

6.2.1 ANALYTICAL SIZE EXCLUSION CHROMATOGRAPHY

Analytical size exclusion chromatography was used to examine the formation of the C3d/CD21 SCR1-2 protein complex. The molecular weights of C3d and CD21 SCR 1-2 are 35 kDa and 15 kDa, respectively. Hence, a 1:1 stoichiometric complex of C3d/CD21 SCR1-2 is expected to have a molecular weight of 50 kDa. A mixture of the C3d/CD21 SCR1-2 complex at a molar ratio of 1:1.3 was prepared to ensure binding site saturation and passed on to an analytical Superdex 75 HR 10/30 column (GE Healthcare), nonetheless a single peak was eluted and visual inspection suggested it corresponded to a 40 kDa protein. Figure 6.3 shows an elution profile of molecular weight standards, C3d monomer and the C3d/CD21 SCR1-2 mixture. Molecular weight standards used were BSA (66 kDa), ovalbumin (45 kDa) and carbonic anhydrase (30 kDa). Plotting a graph of elution volume vs. Log molecular weight of the molecular weight standards (Figure 6.4) gives a calibration line to precisely determine the apparent molecular weight of protein sample. The peak elution volumes of the C3d monomer and the C3d/CD21 SCR1-2 complex are 10.6 ml and 10.4 ml, respectively. Using the calibration line to map the apparent molecular weight of the protein samples, C3d monomer has an apparent molecular weight of 35 kDa and is in agreement with the calculated as well as the mass spectrometry determined molecular weight of C3d. The single peak presumably corresponding to the C3d/CD21 SCR1-2 complex has an apparent molecular weight of 40 kDa as determined with the use of calibration line. Fractions from this single peak were analyzed on an SDS-polyacrylamide gel (Figure 6.5) and bands corresponding to C3d and CD21 SCR1-2 were observed across all fractions. Despite the fact that the proteins appeared to elute as a protein complex, it was not straightforward to assess the stoichiometry of the complex from an SDS-polyacrylamide gel. Although a 1:1 stoichiometry of the protein complex was determined by ITC (Chapter III), it was speculated that the proteins do not form stable complexes in solution owing to their fast dissociation kinetics as observed in SPR sensorgrams presented in Chapter III. The experiment was also repeated at room temperature and the same elution profile was obtained. Hence, temperature does not appear to be a major variable parameter in the formation of the protein complex.

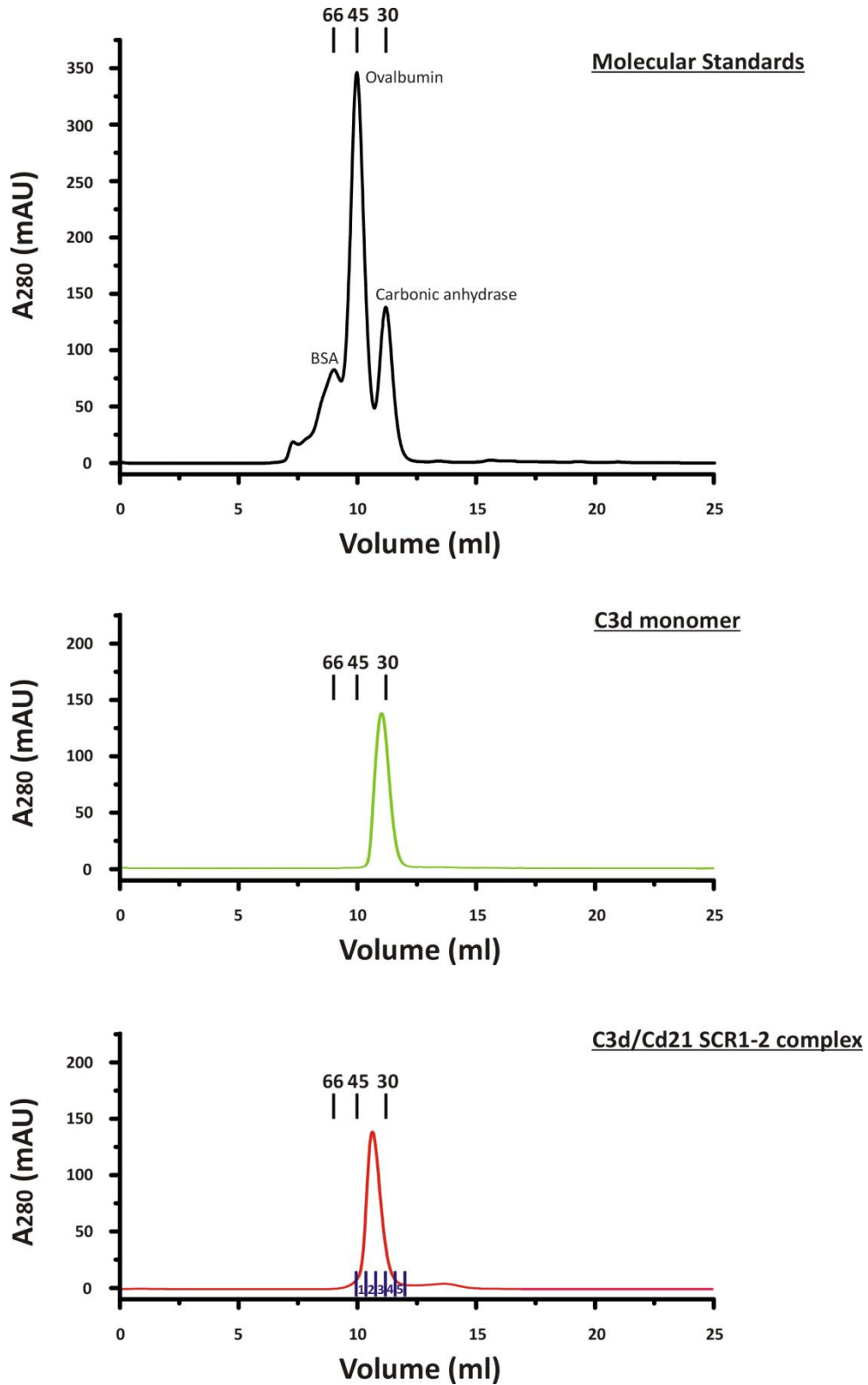


Figure 6.3 | Analytical size exclusion chromatography profile of the molecular weight standards, C3d and the C3d/CD21 SCR1-2 complex

Top: Elution profile of molecular standards – BSA (30 kDa), Ovalbumin (45 kDa) and Carbonic anhydrase (66 kDa). The elution volume of the molecular weight standards was extrapolated and plotted against their Log MW to generate a calibration line as shown in Figure 6.3.

Middle: Elution profile of C3d monomer.

Bottom: Elution profile of the C3d/CD21 SCR1-2 complex.

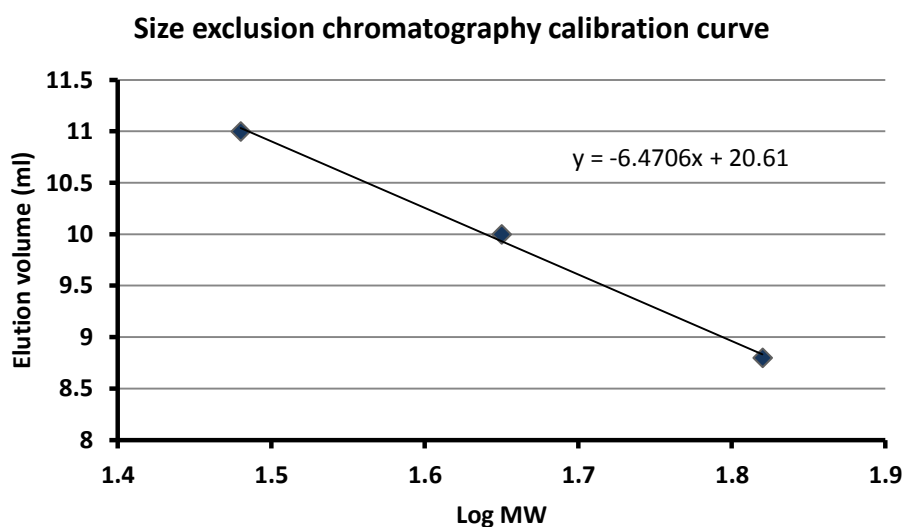


Figure 6.4 | Size exclusion chromatography calibration curve for proteins with molecular weight between 30 kDa to 66 kDa

The elution volume of the molecular weight standards as shown in Figure 6.2 was extrapolated and plotted against their Log MW to generate a calibration line. The elution volume and Log MW has a linear relationship defined by $y = -6.47x + 20.61$.

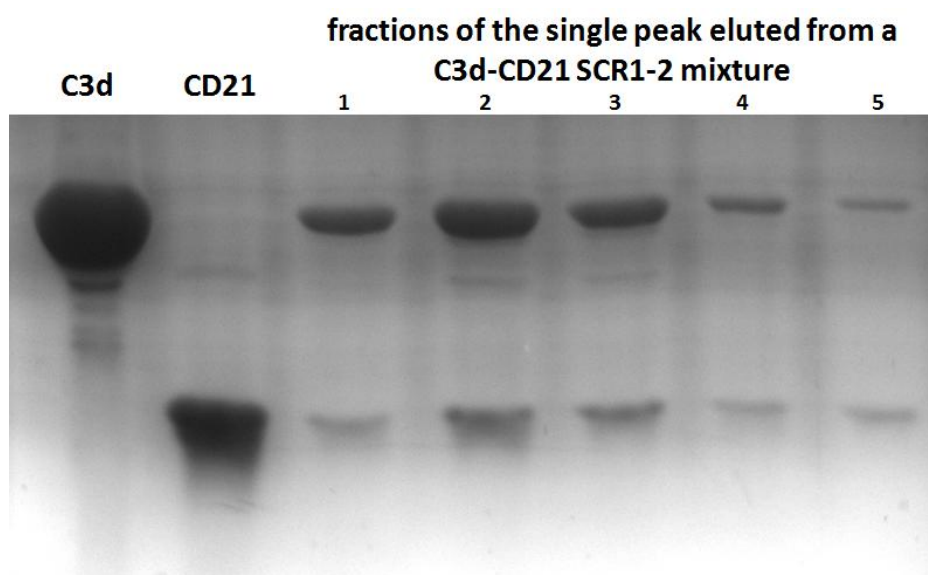


Figure 6.5 | SDS-polyacrylamide gel showing fractions of the single peak eluted from a C3d/CD21 SCR1-2 mixture

Fraction number corresponded to the fractions marked on the C3d/CD21 SCR1-2 gel filtration elution profile on Figure 6.2. It appeared both C3d and CD21 SCR1-2 were present across the fractions when the protein mixture was eluted as a single peak from an analytic SD75 HR10/30 column. The apparent molecular weight of the protein complex is 40 kDa, whereas the calculated molecular weight of the complex is 50 kDa. It was uncertain whether the proteins were in a 1:1 stoichiometric complex by visual inspection of the SDS-polyacrylamide gel.

The influence of ionic strength on protein complex formation was also examined. The C3d/CD21 SCR1-2 complex at a molar ratio of 1:1.3 was run on an analytical Superdex 75 HR 10/30 column (GE Healthcare) equilibrated at different ionic strengths. The formation of the C3d/CD21 SCR1-2 complex was very sensitive to the concentration of NaCl and the proteins appeared to dissociate in 250 mM NaCl (Figure 6.6). Two peaks corresponding to C3d and CD21 SCR1-2 were resolved at an ionic strength of 250 mM NaCl, suggesting that protein complex formation was disrupted; This was expected because the interaction between C3d and CD21 SCR1-2 was predominantly electrostatic driven. The front peak eluted at 10.6 ml certainly corresponded to C3d monomer with an apparent molecular weight of 35 kDa, whereas the second peak eluted at 13.2 ml corresponded to CD21 SCR1-2 monomer with an apparent molecular weight of 14 kDa. A small degree of protein complex dissociation was also observed in 150 mM NaCl, as indicated by a minor peak of CD21 SCR1-2 'leaking' from the preceding protein complex peak. The protein complex appeared to be most stable in 100 mM NaCl and the peak corresponding to the protein complex overlaid perfectly with the one in 150 mM NaCl with no apparent protein dissociation. Although the complex appeared stable in 50 mM NaCl as well, the interaction was further complicated by the fact that CD21 SCR1-2 was notoriously known to adhere to Superdex columns at low ionic strength, *i.e.* 50 mM NaCl. As a result, the single peak corresponded to the C3d/CD21 SCR1-2 complex broadened in 50 mM NaCl and protein adhesion to Superdex column was thought to play a role in this broadening effect.

Based on the observations from the analytical size exclusion experiments, it is concluded that the C3d/CD21 SCR1-2 protein complex does not form a stable protein complex in solution. The protein complex has an apparent molecular weight of 40 kDa in 100 mM NaCl as well as 150 mM NaCl, while the calculated molecular weight of the complex is 50 kDa. Moreover, the protein complex formation is sensitive to ionic strength and falls apart in 250 mM NaCl as separate monomeric proteins. CD21 SCR1-2 eluted at 12.8 ml in 150 mM NaCl and 13.2 ml in 250 mM NaCl, which corresponded to an apparent molecular weight of 16 kDa and 14 kDa, respectively. Ionic strength and the presence on C3d appear to influence the retention of CD21 SCR1-2 on a Superdex gel filtration column, and thus the integrity of C3d/CD21 SCR1-2 complex showing a fast-on/fast-off kinetics depends on the attraction of CD21 SCR1-2 to C3d and Superdex column.

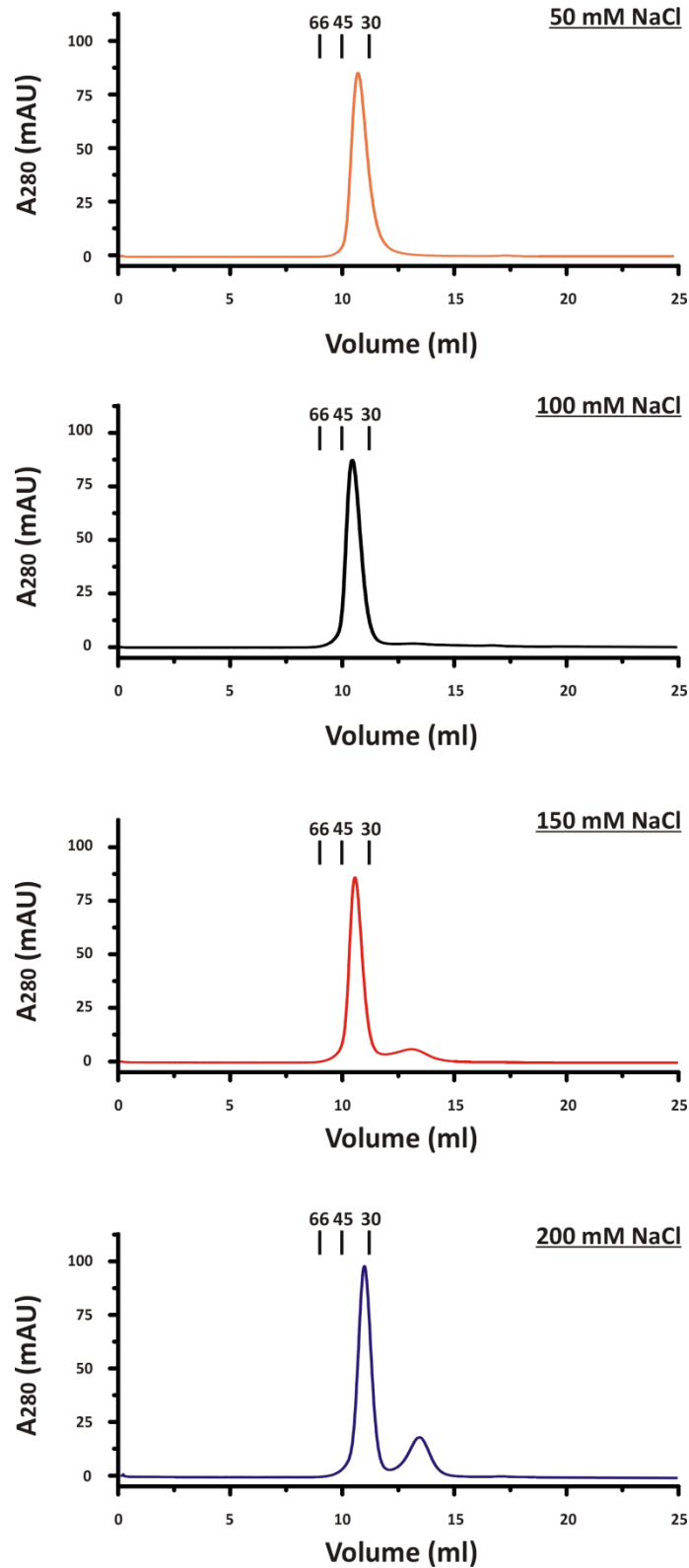


Figure 6.6 | Influence of ionic strength on the formation of C3d/CD21 SCR1-2 complex

In 50 mM NaCl, the protein complex eluted as a single peak but trailing suggested protein adhesion to the Superdex column.

In 100 mM NaCl, the protein complex appeared most stable with no apparent dissociation observed. However, the apparent molecular weight of the complex is still 40 kDa.

In 150 mM NaCl, the protein complex exhibited minor dissociation and the apparent molecular weight of the complex is 40 kDa.

In 250 mM NaCl, the protein complex dissociated completely with two separate peaks observed corresponding to C3d and CD21 SCR1-2, respectively.

6.2.2 CRYSTALLIZATION TRIALS OF THE C3d/CD21 SCR1-2 PROTEIN COMPLEX

Since the C3d/CD21 SCR1-2 protein complex appears to be most stable in 100 mM NaCl, crystallization trials were set up primarily with proteins in 10 mM HEPES, 100 mM NaCl, pH 6.8. Crystallization trays were set up in two ways – either by using a mixture of C3d and CD21 SCR1-2 at an equal molar ratio directly or by using the size exclusion peak corresponding to the C3d/CD21 SCR1-2 complex. Optimal protein concentrations for tray setup were determined to be no less than 25 mg/ml (or 450 μ M of each protein) according to preliminary trials. Trays were set up at 4 °C and 18 °C with eight commercially available screens, namely Natrix (Hampton Research), PEG/Ion (Hampton Research), Proplex (Molecular Dimensions), JCSG (Molecular Dimensions), PACT (Molecular Dimensions), Wizard (Emerald Biosystems), Molecular Dimension 1/2 (Molecular Dimensions) and Morpheus (Molecular Dimensions). The drop ratio of protein solution to mother liquor was initially tested at 1:1, as well as 2:1 in subsequent trials to compensate for high salt content of crystallization buffers (typically at 0.5 M for most screens). Unfortunately, screening a total number of 6144 different crystallization conditions in these trials only yield a small number of salt crystals.

Table 6.2 | Crystallization trial of the C3d/CD21 SCR1-2 protein complex

Crystallization screen	Protein sample preparation	Drop ratio	Temperature (°C)	Crystals (salt)
Proplex	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	2
		2:1	4 and 18	1
PEG/Ion	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	2
		2:1	4 and 18	2
Natrix	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	1
		2:1	4 and 18	2
JCSG	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	1
		2:1	4 and 18	1
PACT	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	0
		2:1	4 and 18	0
Wizard	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	0
		2:1	4 and 18	0
MD 1/2	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	2
		2:1	4 and 18	2
Morpheus	1. Manual mixing; 2. Gel filtration of the protein complex	1:1	4 and 18	0
		2:1	4 and 18	0

6.3 DISCUSSION

Until very recently, the precise binding surfaces of C3d and CD21 SCR1-2 remained controversial in the published crystal structure of the protein complex solved by Szakonyi G *et al.* in 2001, owing to a disagreement between the crystal structure (Szakonyi G *et al.*, 2001), biochemical (Clemenza L and Isenman DE, 2000; Hannan JP *et al.*, 2005; Bomb M, 2008) and recent NMR data (Bomb M, 2008; Kovacs JM *et al.*, 2009). The interaction of C3d and CD21 SCR1-2 exhibited strong ionic strength dependence and salt content in solution had significant influence on the formation of the C3d/CD21 SCR1-2 protein complex. Much effort was made to identify conditions that would stabilize the protein complex in attempt to crystallize a physiologically relevant C3d-CD21 SCR1-2 protein complex. However, crystallization trials were unsuccessful and yielded several salt crystals in various crystallization conditions. Fortunately, a new crystal structure of the C3d-CD21 SCR1-2 complex was published by a separate group in April 2011 (PDB:3OED; van den Elsen JMH and Isenman DE, 2011). The new crystal structure delineated the protein binding interfaces at atomic resolution and provided solid evidence to confirm the interactions observed in the old structure (PDB:1GHQ) were entirely artifactual. It was suggested that Szakonyi's crystal structure represented a non-physiologically relevant binding mode as a result of favorable crystal-packing energetics in the presence of zinc acetate in the crystallization conditions. According to the crystallization methods published by van den Elsen JMH and Isenman DE, purified C3d and CD21 SCR1-2 were mixed at a 1:1 stoichiometric ratio and concentrated to a combined protein concentration of 11.4 mg/ml (equivalent to 230 μ M each). The protein buffer used was 10 mM PIPES, 100 mM NaCl, 0.02% NaN₃, pH 6.25 and crystallization conditions were 0.1 M HEPES, 0.2 M CaCl₂, 20% PEG 6000, pH 7.0 at 18 °C. Several crystals grew in a few days and one of them diffracted to 3.2 Å resolution.

There are two C3d/CD21 SCR1-2 protein complexes in an asymmetric unit of the new crystal structure (Figure 6.7; PDB:3OED). Despite the fact that the orientation of CD21 SCR1-2 domains is the same in both complexes, there are some differences in symmetry-related crystal contacts in the two complexes. It suggests that the interactions between the proteins are not entirely stable, which could be due to the fast-on/fast-off interaction kinetics as characterized by SPR experiments. Structural alignments of the C3d and CD21 SCR1-2

molecules in the protein complex with their unliganded counterparts showed no major conformational changes upon binding. The buried surface area (BSA) of the C3d/CD21 SCR1 complex is 2250 Å² and shows striking shape and charge complementarity (van den Elsen JMH and Isenman DE, 2011). Contact surfaces between the negatively-charged concave surface on C3d and the SCR1 domain of CD21 account for about 75% of the total BSA of the complex. This is in contrast to the interface observed in the previous C3d/CD21 SCR1-2 complex (PDB:1GHQ); the BSA between C3d and the SCR2 domain of CD21 is less than 800 Å² and is comparable to the contact surfaces formed between the non-physiological homodimer of CD21 SCR1-2 (880 Å²). Existing biochemical and NMR data (Clemenza L and Isenman DE, 2000; Bomb M, 2008; Kovacs JM *et al.*, 2009; van den Elsen JMH and Isenman DE, 2011), as well as new microcalorimetric data presented in Chapter III of the thesis are in agreement with the contacts observed in the new crystal structure of the protein complex. Microcalorimetric data of the double mutant cycles presented in Chapter III precisely identified two salt bridges that were observed in the new crystal structure of the protein complex, namely CD21 SCR1-2 R13 with C3d E37, as well as CD21 SCR1-2 K41 with C3d D163. Only a few charged interactions are observed between the SCR2 domain of CD21 and C3d, such as CD21 K108 with C3d E166 and CD21 D92 with C3d K251. Interestingly, side chain of residue V26 on the SCR1 domain of CD21 formed an intricate network of van der Waals interactions with a small pocket formed by residues V97, I100, A101, I102, D103, S104 and I164 on C3d. Several residues from this pocket on C3d are also utilized by immune-evasion proteins from *S. aureus* (Figure 1.5, Chapter I). These proteins of *S. aureus* share significantly overlapping contact surfaces with CD21 SCR1-2 on C3d and thus explain their inhibitory activities towards the C3d/CD21 SCR1-2 interaction (Isenman DE *et al.*, 2010; Burman JD *et al.*, 2008; Ricklin D *et al.*, 2008; Clark EA *et al.*, 2011; Hammel M *et al.*, 2007).

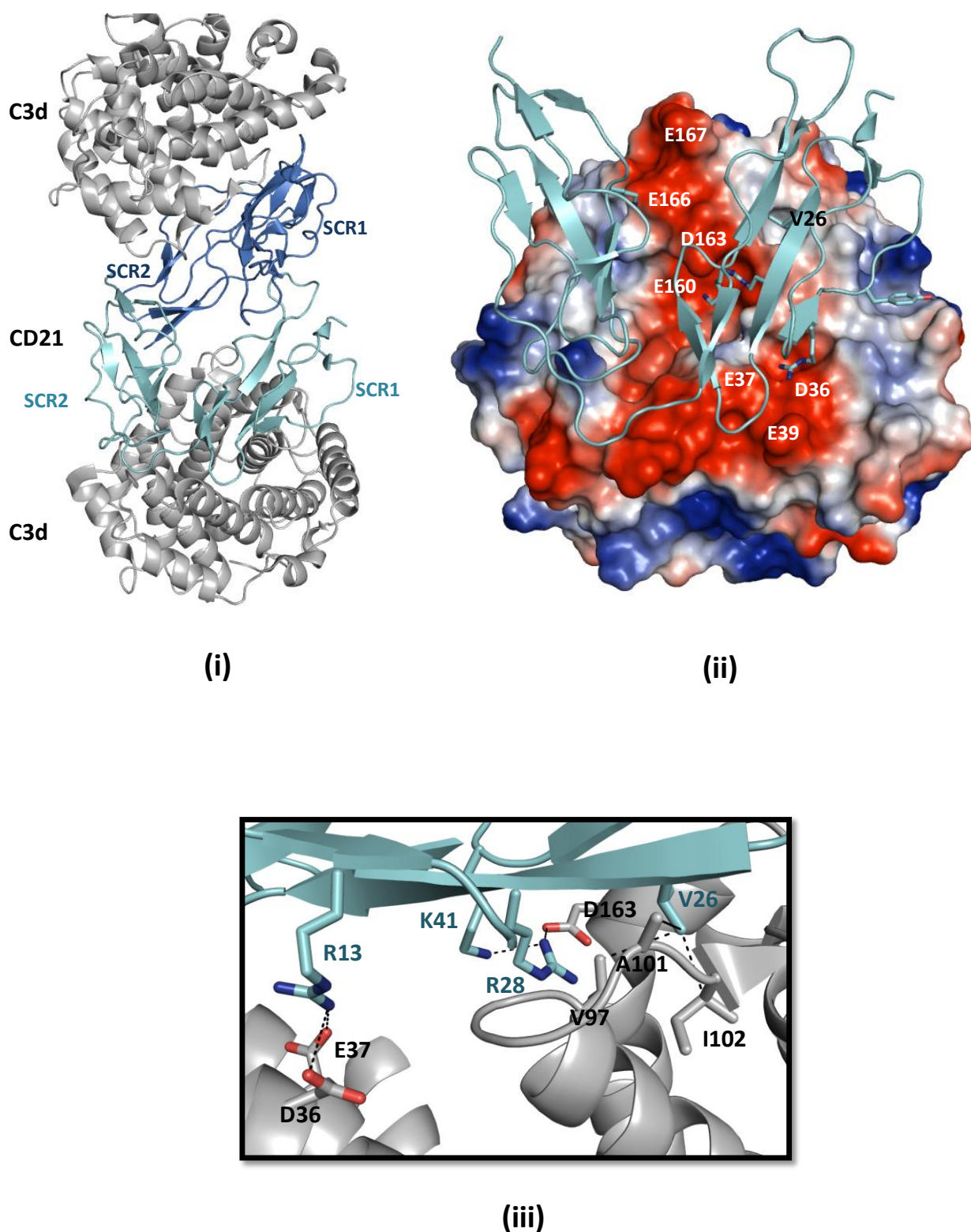


Figure 6.7 | Crystal structure of the C3d/CD21 SCR1-2 protein complex solved by Isenmen DE *et al.* (PDB:3OED)

- (i) Structure of the C3d-CD21 SCR1-2 protein complex showing overall molecular architecture as cartoon diagram. There are two C3d-CD21 SCR1-2 protein complexes in an asymmetric unit with CD21 SCR1-2 coloured in light blue and dark blue, respectively.
- (ii) Structure of the protein complex with C3d shown as electrostatic surface representation. The side chain of V26 on CD21 SCR1 occupies a small pocket formed by V97, I100, A101, I102, D103, S104 and I164 on C3d.
- (iii) Close-up diagram of the crystallographic C3d/CD21 SCR1 binding interface.

A model of mutagenesis data-constrained C3d/CD21 SCR1-2 protein complex was generated from molecular docking server ClusPro (Comeau SR *et al.*, 2004), as detailed in Chapter III. The model provides atomic-level information regarding specific interactions between C3d and CD21 SCR1-2 and the binding mode shows substantial similarity with the new crystal structure (Figure 6.8). The docking program relies on an input of the crystal structure of C3d (PDB: 1C3D) and CD21 SCR1-2 (PDB:1LY2), as well as user-restrained residue specific information such as residues that are known to mediate the protein-protein interaction from prior mutagenesis experiments. The data is sent to an automated docking server, which then generates several potential binding modes with downloadable PDB files of the protein complex based on low energy binding results calculated from the docking server. As illustrated in Figure 6.8, the new crystal structure of C3d/CD21 SCR1-2 complex provides a solid validation on our mutagenesis and docking studies in defining the native binding interfaces of the protein-protein interaction between C3d and CD21 SCR1-2. Superposition of our C3d/CD21 SCR1-2 structure predicted by ClusPro and the crystal structure of the C3d/CD21 SCR1-2 complex solved by Isenmen DE *et al.* (PDB:3OED) results in r.m.s.d. values of 0.53Å for backbone atoms and 0.93 Å for all atoms. The binding mode predicted by ClusPro is in surprisingly close agreement with the binding mode determined crystallographically by Isenmen DE *et al.*. Hence, it is possible to deduce the native binding mode of two proteins by using microcalorimetric analysis of mutagenesis studies coupled with a robust molecular docking server (*e.g.* ClusPro), even in the absence of a physiologically relevant crystal structure of the protein complex.

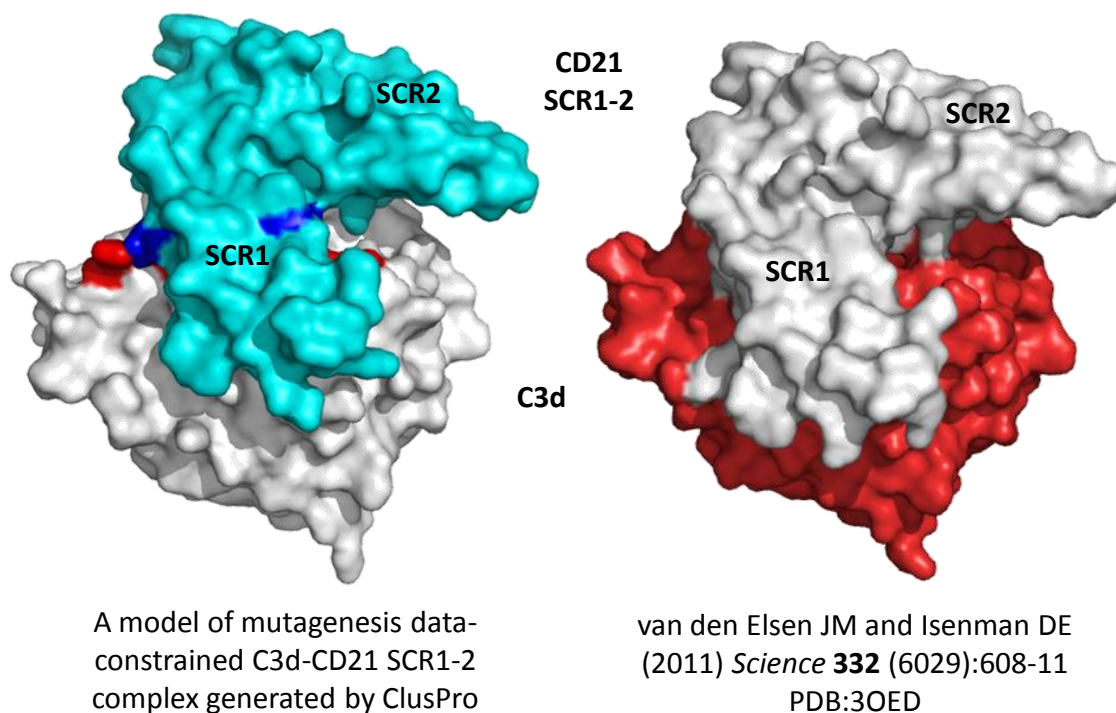


Figure 6.8 | Comparison of the C3d/CD21 SCR1-2 complex predicted by ClusPro and solved by X-ray crystallography

The binding mode predicted by ClusPro shows a high degree of similarity to the binding mode determined crystallographically by Isenman DE *et al.*. Salt bridges identified from our microcalorimetric analysis of double mutant cycles presented in Chapter III are highlighted in the ClusPro model. Superposition of the ClusPro model and 3OED results in r.m.s.d. values of 0.53Å for backbone atoms and 0.93 Å for all atoms.

CHAPTER VII

FRAGMENT-BASED DRUG SCREENING

7.1 INTRODUCTION TO FRAGMENT-BASED DRUG DISCOVERY

7.1.1 RATIONAL BEHIND FRAGMENT-BASED DRUG DESIGN

Fragment-based drug design is a drug discovery strategy pioneered by Astex Therapeutics Ltd. (now Astex Pharmaceuticals) in 2000 and is now firmly established as an effective approach to general novel drug leads against protein targets for which structural information is readily available. The fragmentation of drug-like molecule into smaller pieces (<200 Da), such as discrete functional groups, is a concept that has been used *in silico* to simplify the computational analysis of ligand binding and to map out different pharmacophoric elements required for high-affinity binding (Bohm HJ, 1995; Miranker A *et al.*, 1991). This approach allows proper optimization of each unique interaction in the binding site and subsequent incorporation into a single

molecular entity. The resulting compound in theory would possess a binding affinity that is the sum of the individual interactions. In order to effectively use fragments for drug discovery research in an experimental setting, rapid and reliable screening methodologies are required to assay thousands of low-molecular-weight fragments to test for weak binding (K_D values in the high μM to mM range) to the target protein. The strategy involves the design and screening of fragment libraries, evaluation of potential hits, structural validation of fragment binding modes and structure-guided evolution of novel tight binders in a step-by-step approach. Using fragments with favourable physiochemical, adsorption, distribution, metabolism, excretion (ADME) profiles and pharmacokinetics properties offers a feasible starting point to find hits that avoid undesirable pharmacophores in the drug discovery process. Fragments are considered efficient binders despite of their low potency owing to their low complexity and higher binding energy per unit mass than conventional high-throughput screening (HTS) hits.

A conventional high-throughput screening (HTS) campaign attempts to evaluate as many full-size compounds (~ 500 Da) as technologically possible, typically in a range of a million or more, in the anticipation of finding relatively potent drug leads with K_D values ideally lower than $1 \mu\text{M}$. However, it has been estimated the size of the chemical universe is in the range of 10^{60} compounds (Martin YC, 1997) and as a result screening 10^6 compounds merely covers a small fraction of available chemical space. The fragment universe is much smaller with an estimation of the size of the chemical universe below 160 Da is approximately 14 million compounds (Fink T *et al.*, 2005). Therefore, screening a fragment library of $10,000$ compounds captures substantially more chemical diversity space than conventional HTS campaign. Furthermore, less complex molecules exhibit higher hit rates against protein targets as theoretically calculated by Hann MM *et al.* at GlaxoSmithKline (Hann MM *et al.*, 2001). Although a typical fragment screen only explores less than 1% of the available low-molecular-mass universe, the probability to find leads is substantially higher than conventional HTS campaign and further increases the value of fragment screen. The idea has been validated by drug discovery efforts at Novartis (Schuffenhauer A *et al.*, 2005), in which the observed hit rates for fragment screens were $10 - 1,000$ times higher than a conventional HTS campaign. The combination of broader sampling of the potential chemical universe and increased hit rates for molecules of low complexity makes fragment-based

screening a powerful strategy for lead generation. A recent analysis indicated that a sufficiently large fragment screen provides excellent coverage of clinically useful chemical space (Hajduk PJ *et al.*, 2005), as the ability to find fragment leads against a protein target correlates well with the success in producing potent small molecule inhibitors of protein activity.

The fragment library used in this thesis was kindly provided by Dr. David Mountford at the Institute of Pharmaceutical Science, King's College London. Criteria to select fragments was based on the Congreve's rule of three for lead likeness (Congreve M *et al.*, 2005), in which the molecular weight of fragment ≤ 300 Da, the number of hydrogen bonds of both acceptors and donors ≤ 3 and $\log P \leq 3$. This is a more stringent criteria than the Lipinski's rule of five for drug likeness (Lipinski CA *et al.*, 1996). Our small fragment library comprised of one thousand chemical fragments was acquired from Maybridge (Fisher Scientific) and presents a wide range of chemical and structural diversity. Although structure guidance was available for the CD21 SCR1-2 binding site on C3d, the fragment library was constructed without the aid of *in silico* screening against C3d. Emphasis was placed on chemical novelty and the library was designed to be applicable for different types of targets. A selection filter was also applied for the removal of chemical scaffolds with toxicophores or inappropriate functionality (*e.g.* reactive groups that might covalently react with target proteins or groups that were not amenable to chemical evolution during hit-to-lead progression). The fragments were dissolved as 50 mM in 100% DMSO with assured purity of $\geq 95\%$.

7.1.2 PROSPECTIVE SMALL MOLECULES BINDING GROOVE ON THE SURFACE OF C3d

Mutational studies and microcalorimetric analysis presented in Chapter III delineated the physiologically relevant binding interface on C3d and CD21 SCR1-2. *In silico* docking suggested the contact surfaces on both proteins were broad and showed high degree of charge and shape complementarity. It was shown that only a small subset of residues at the binding interface contributed most of the free energy of binding (Table 3.2, Chapter III). The CD21 SCR 1-2 binding site on C3d comprises two predominant clusters of negatively charged residues surrounding a deep hydrophobic groove (Figure 7.1). The clusters constitute approximately two-thirds of the binding interface on C3d and it was postulated small drug-like molecules could bind deeper

in the hydrophobic groove with higher efficiencies than the native protein partner, *i.e.* CD21 SCR1-2. It was hoped to develop novel small molecule inhibitors that would make optimal use of this hydrophobic groove on C3d and block out the access of CD21 SCR1-2. Successful inhibitors could provide a new class of drugs to complement traditional therapies for rheumatoid arthritis, or be used as research tools for the analysis of the functional relevance of the interaction between C3d and other native ligands utilizing the same binding site in basic research studies.

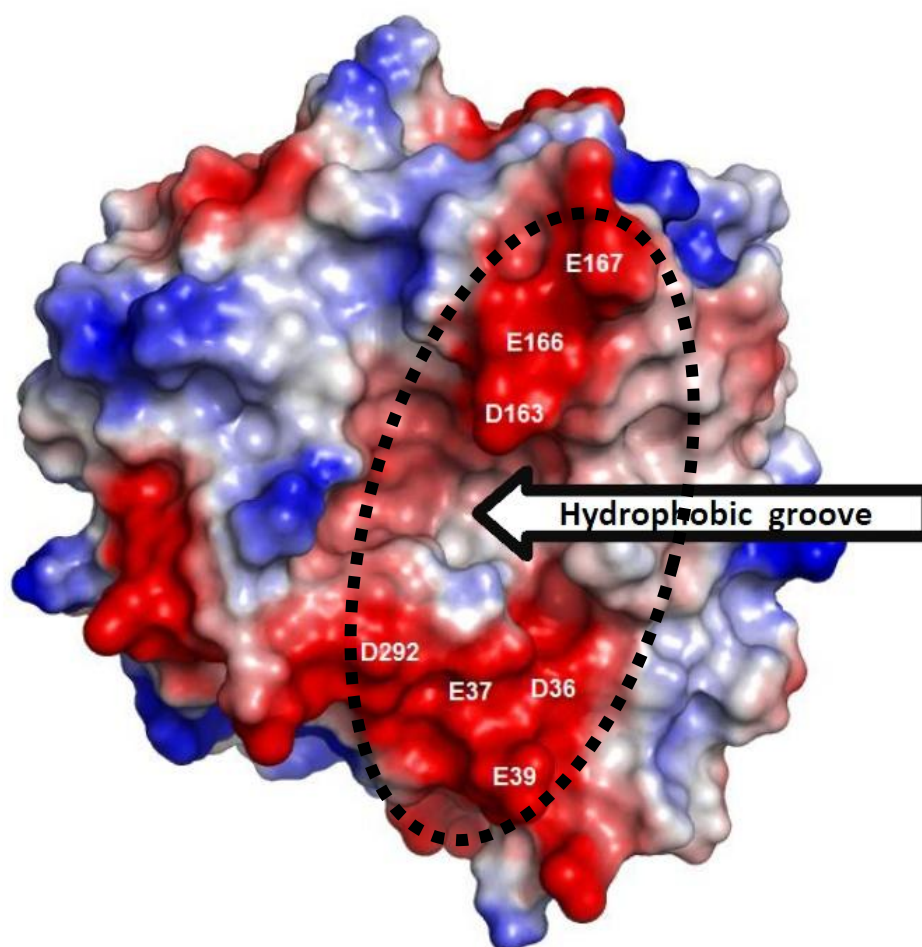


Figure 7.1 | Prospective small molecule binding groove on C3d

The CD21 SCR1-2 binding site on C3d is encircled with dotted line. Negatively charged residues highlighted on C3d were shown to be the key residues mediating the interaction. The deep hydrophobic groove at the binding interface presents a promising site for small molecule inhibitors to disrupt the association of C3d and CD21 SCR1-2.

7.2 HIGH-THROUGHPUT FRAGMENT SCREENING TECHNIQUES

7.2.1 DIFFERENTIAL SCANNING FLUORIMETRY

The technique was pioneered by 3D Pharmaceuticals which became part of Johnson & Johnson (Cummings MD *et al.*, 2006) and the interest in this technology was also raised by Structural Genomics Consortium to screen for ligands that promote protein stability (Senisterra GA *et al.*, 2008; Niesen FH *et al.*, 2008). It is now readily used as a cost-effective, reagent friendly, high-throughput screening of small molecules with low micromolar target affinity (Abad MC *et al.*, 2008; Lo MC *et al.*, 2004).

Differential scanning fluorimetry (DSF) assay measures protein unfolding via a change in the fluorescent properties of bound dye molecules upon protein denaturation. The ability of ligand binding to enhance protein stability is a well recognized phenomenon. The degree of stabilization can be observed through the increased resistance of a protein to physically denaturing condition, *i.e.* temperature. A fluorescence dye, *e.g.* Sypro Orange, is used in the assay to probe protein denaturation over a temperature range spanning the typical protein unfolding temperatures of 25–95 °C. Intrinsic fluorescence of the dye is quenched in aqueous environment but is enhanced when bound to hydrophobic patches on proteins. Plotting sample fluorescence as a function of increasing temperature gives a protein denaturation curve and the melting temperature of the protein (T_m) is defined by the midpoint of the curve. Ligand binding induces a shift in protein thermal stability and ΔT_m represents the difference of the protein melting temperature in the absence and the presence of ligand. It has been demonstrated that the increase in T_m of a protein by ligand interaction is proportional to the concentration as well as affinity of the ligand (Matulis D *et al.*, 2005). Nevertheless, the affinity of a ligand cannot be accurately determined from the saturation value of ΔT_m . In order to qualify DSF as a thermodynamic readout technique, the folding transition monitored has to be reversible and the interaction of ligand with the unfolded state of the target protein has to be excluded. Since these conditions are not fulfilled in the majority of cases, DSF is generally regarded as a high-throughput assay to prioritize the rank order of ligand-binding strength.

The protein concentration used in DSF assay depends on the protein of interest and typically falls in the range

of 1 – 5 μM . For high-throughput screening applications, it is necessary to establish the lowest limit of protein concentration that show large enough signal for fluorescence change upon protein unfolding to give a precise determination of T_m . Small proteins below 15 kDa (*e.g.* ubiquitin) often do not have a substantially different portion of hydrophobic surface area between folded and unfolded states and exhibit similar affinities of fluorescent dye to both states (Kranz JK and Schalk-Hihi C, 2011). As a result, the adaptability of certain proteins to DSF assay is restricted due to insufficient signal amplitude unless much higher protein concentration is used.

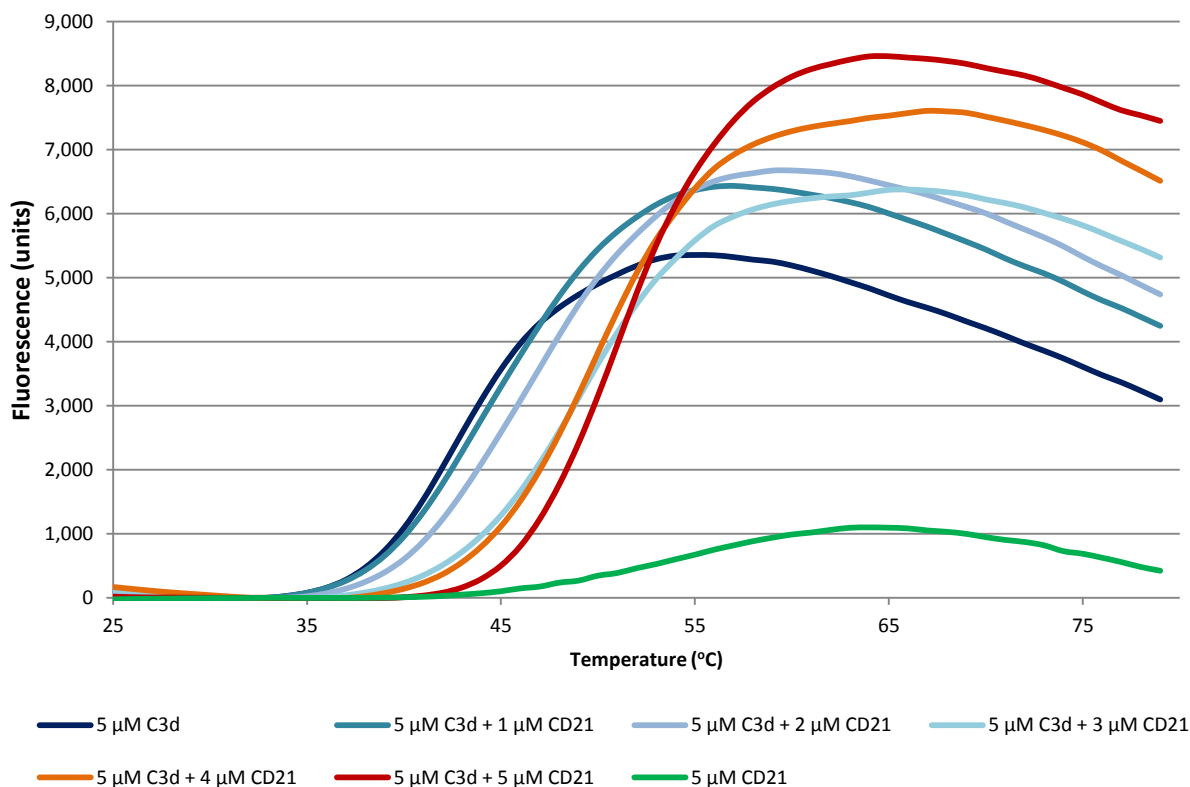
7.2.2 PILOT DSF ASSAY WITH C3d AND CD21 SCR1-2

The wild type C3d-CD21 SCR1-2 interaction was investigated to establish whether C3d was applicable to DSF assay. The results are presented in Figure 7.2 and data analysis was carried out with a well established 'DSF analysis' Microsoft Excel template available for download from <ftp://ftp.sgc.ox.ac.uk/pub/biophysics> (Niesen FH *et al.*, 2008). At a protein concentration of 5 μM , the T_m values of C3d and CD21 SCR1-2 alone were 44.8 $^{\circ}\text{C}$ and 56.2 $^{\circ}\text{C}$, respectively. The pilot screen was done in triplicate and the denaturation profiles were highly reproducible at a given scenario. The standard deviations of T_m were in the range of $\pm 0.15 - 0.20$ $^{\circ}\text{C}$ across all scenarios. As expected with small proteins, CD21 SCR 1-2 (15 kDa) produced little fluorescent intensity and the maximal level observed was 5-fold lower than C3d (35 kDa) at the same protein concentration. As the proteins unfolded in a temperature ramp, greater portions of the hydrophobic core of the proteins were exposed and allowed interaction with Sypro Orange. It thus enhanced fluorescent readout until the melting proteins were intolerable at a high enough temperature. At this point the melting proteins were denatured and precipitated out of solution, thereby reducing the amount of protein that could interact with Sypro Orange and led to a gradual decline in fluorescent intensity.

A dose-dependent experiment using CD21 SCR1-2 as a native protein ligand for C3d indicated that ΔT_m was saturable (Figure 7.2a). A stepwise increase in the concentration of CD21 SCR1-2 positively shifted the thermal stability of C3d. The T_m value of C3d was 51.8 $^{\circ}\text{C}$ when both proteins were at an equal molar ratio, giving a ΔT_m of +7.1 $^{\circ}\text{C}$. Figure 7.2b showed a graph of ΔT_m plotted against concentration of CD21 SCR1-2. The combined

fluorescent intensity of C3d and CD21 SCR1-2 was also higher in a dose-dependent manner. Such behavior was likely due to additional interaction of Sypro Orange with the melting C3d-CD21 SCR1-2 protein complex.

(a)



(b)

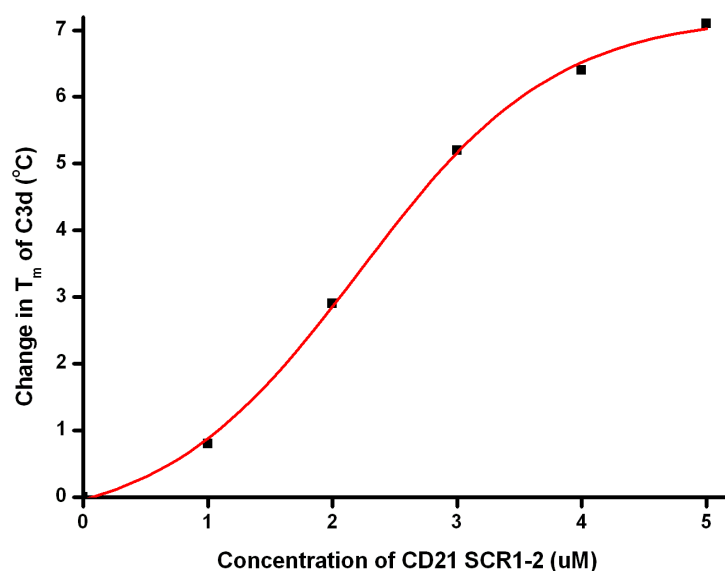


Figure 7.2 | Pilot DSF assay with C3d and CD21 SCR1-2

- Denaturation curves showing CD21 SCR1-2 increased the thermal stability of C3d in a concentration-dependent manner.
- A plot of ΔT_m of C3d versus concentration of CD21 SCR1-2 derived from DSF assay showing the interaction was saturable.

7.2.3 DSF-BASED FRAGMENT LIBRARY SCREENING AGAINST C3d

The denaturing effect of DMSO on the thermal stability of C3d was examined in order to initiate the screening campaign. Briefly, C3d was tested across a range of DMSO content ranging from 0 – 5 % and no changes to T_m were observed up to 2% DMSO. C3d was well tolerated and the T_m of C3d was decreased by merely 0.4 °C in 4% DMSO, *i.e.* giving T_m of 44.4 °C. Subsequent fragment screens were carried out at a C3d concentration of 5 μ M in 20 mM NaH_2PO_4 , 150 mM NaCl, pH 6.8 buffer supplemented with 4% DMSO to ensure fragment solubility. Fragment concentration used in the screen was 1 mM and ΔT_m of >0.6 °C (*i.e.* three-fold of the standard deviation for T_m of C3d in the absence of ligand in 4% DMSO) was taken as a cutoff. Among the one thousand fragments screened, 17.3% did not affect the thermal stability of C3d, 64.8% induced a positive shift in the T_m of C3d (Figure 7.3a and 7.3b) and 16.6% exhibited abnormal thermal profiles possibly due to potential fragment interference with fluorescent readouts (Figure 7.3c). Only thirteen fragments (1.3%) induced a significant negative shift in the thermal stability of C3d (Figure 7.3d). The destabilizing effect of these fragments was in the range of $-0.7 \leq \Delta T_m \leq -2.5$ °C.

For fragments that increased the thermal stability of C3d, forty-two fragments produced standard, clean looking denaturation curves (Figure 7.3a), whereas a majority of the fragments induced substantial amount of thermal stabilization accompanied with multiphasic denaturation profiles (Figure 7.3b). Such unusual stabilization behavior could arise from promiscuous binding of fragments to the melting hydrophobic core of C3d. The multiphasic nature of the denaturing profiles suggested that these fragments might initially interact with a single domain on C3d and subsequently bind to multiple sites as more hydrophobic core of C3d was revealed in the thermal denaturing process. In a typical DSF assay, the denaturing curve of a protein would gradually decline after reaching maximal fluorescent level as the protein precipitated out of solution. Promiscuous binding of fragments to the hydrophobic core of the protein appears to greatly extend protein stability and prevent it from precipitating. As a result, the denaturation curves of C3d plateau after reaching fluorescent maxima and the binding of these fragments to C3d appeared mostly insaturable. For the forty-two fragments that produced significant positively-shifted denaturation profiles, the ΔT_m values were in a range of 1 – 5 °C. Using DSF as a primary fragment library screen produced a hit rate of 4.2%, nonetheless most

denaturation profiles seemed peculiar and a great portion of fragments were not characterized owing to fluorescent interference. Therefore, surface plasmon resonance was employed as an alternative technique to screen the same fragment library.

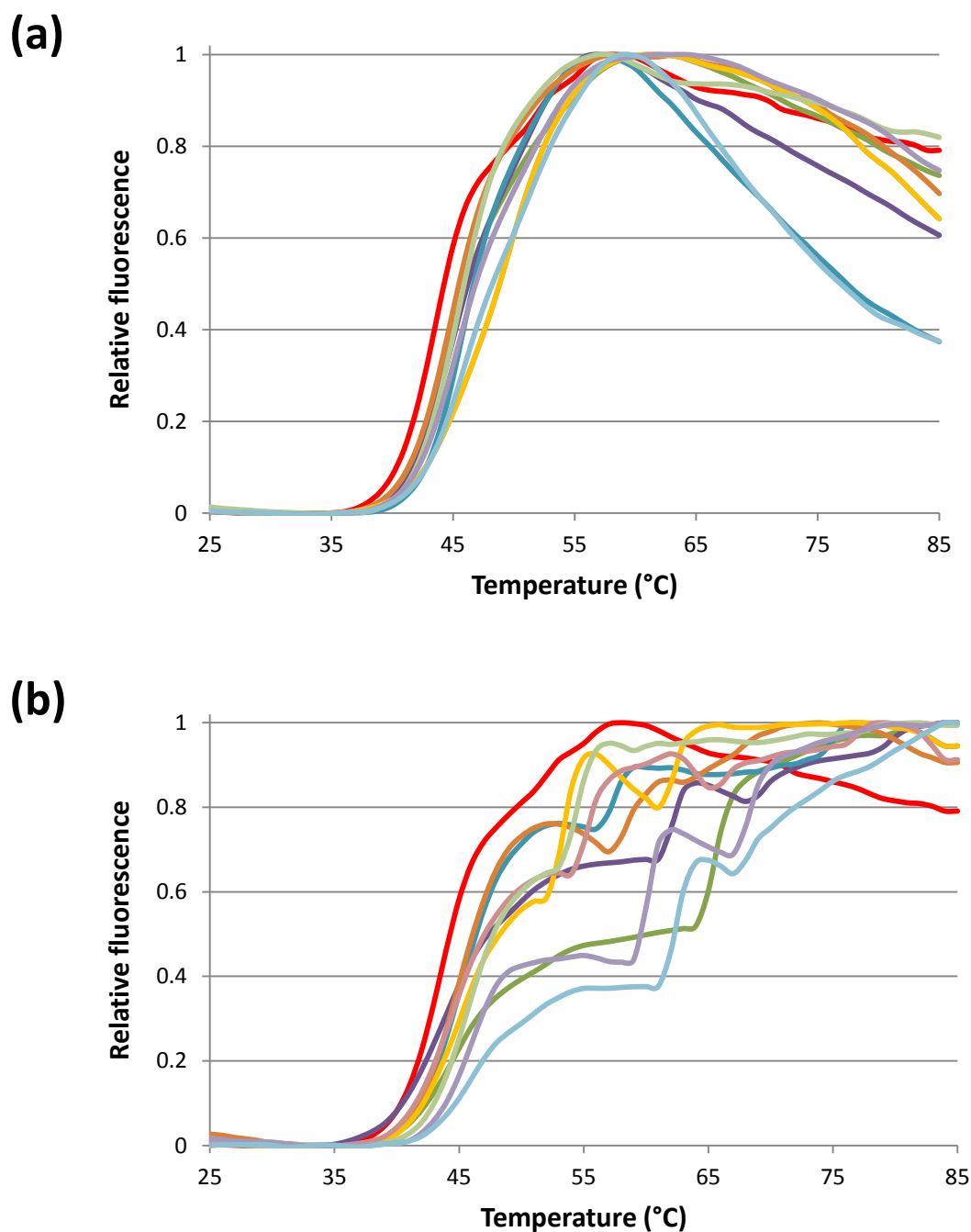


Figure 7.3 | Representative denaturation curves showing the influence of fragments on the thermal stability of C3d

Curve in red = C3d only in 4% DMSO buffer; Curves in other colours = C3d and fragments in 4% DMSO buffer.

- Examples of fragments that produced clean denaturation curves and increased the thermal stability of C3d. Such profiles represented 4.2% of the fragment library.
- Examples of fragments that induced multiphasic denaturation curves. Such profiles represented 60.6% of the fragment library.

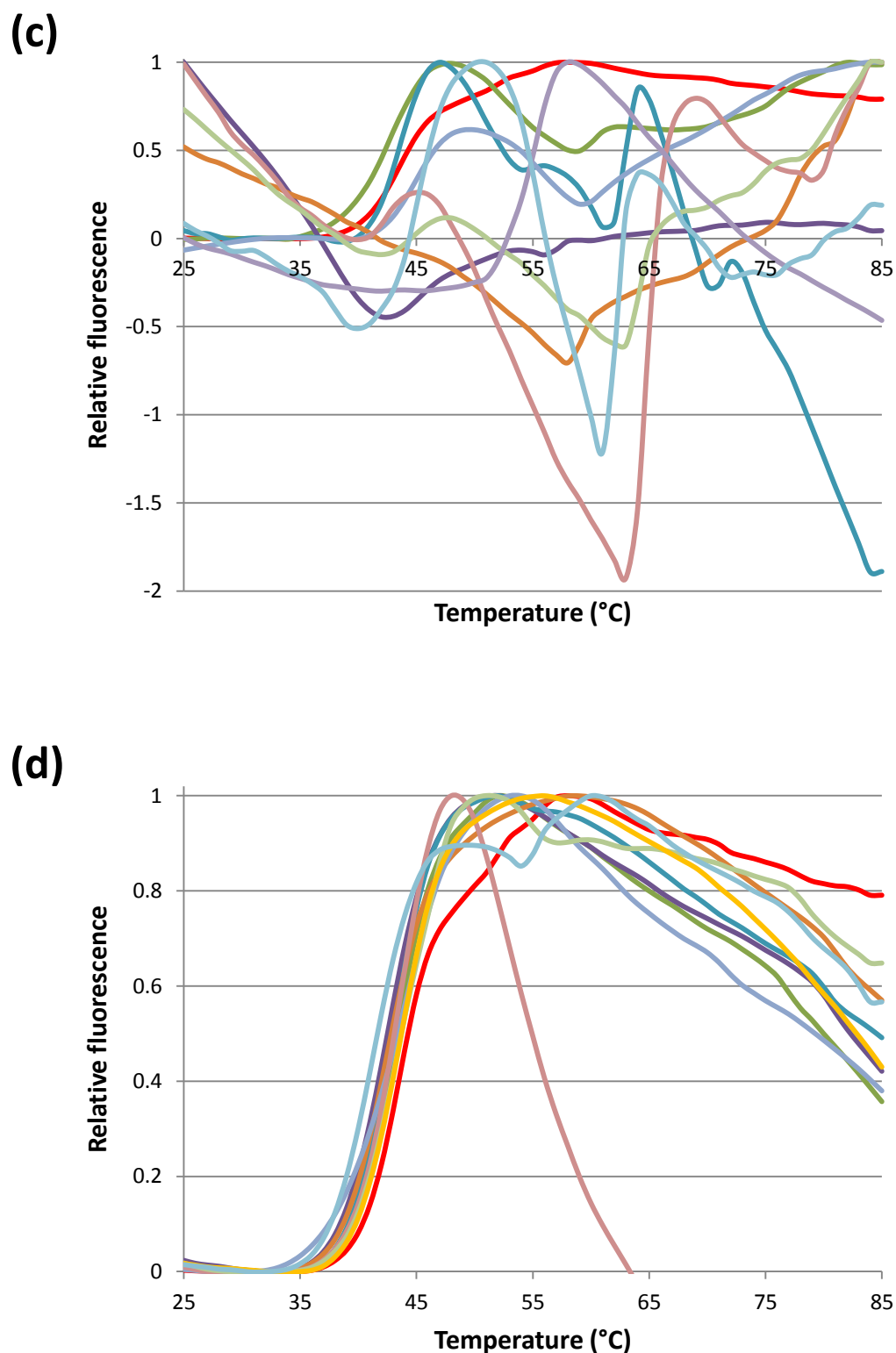


Figure 7.3 | Representative denaturation curves showing the influence of fragments on the thermal stability of C3d

Curve in red = C3d only in 4% DMSO buffer; Curves in other colours = C3d and fragments in 4% DMSO buffer.

- c) Examples of fragments that suffered chemical interference in DSF assay. Such profiles represented 16.6% of the fragment library.
- d) Examples of fragments that decreased the thermal stability of C3d. Such profiles represented 1.3% of the fragment library.

7.2.4 APPLICATION OF SURFACE PLASMON RESONANCE IN DRUG DISCOVERY

SPR-based fragment screening has gained wide recognition and demonstrated success in drug lead identification in the pharmaceutical industry (Hamalainen MD *et al.*, 2008; Nordstrom H *et al.*, 2008). It offers sufficient sensitivity and high-throughput screening on libraries of a few thousand fragments in several weeks per target. SPR provides quantitative binding information for ranking fragments by affinity and a detailed view of how the fragments interact with target proteins. Furthermore, it can support ongoing quantitative structure-activity efforts during fragment evolution in hit-to-lead development (Myszka DG and Rich RL, 2000). Current Biacore™ instruments require sample volumes of 100 µl and can analyze samples in a 96-well plate format. Nevertheless, screening low molecular weight compounds is associated with a number of technical challenges, including working with high refractive index solvents (*e.g.* DMSO), compound solubility limitations in aqueous solvents, non-specific binding and extremely low-affinity interactions. In most cases, the affinity of fragment to target protein is in the range of <10 mM and hence there will be little detectable binding interaction even at high fragment concentration. Since SPR signal is proportionate to mass (Malmqvist M, 1993), it is absolutely necessary to work at high protein immobilization density (typically in the range of 5000 – 10,000 RU) to reliably detect and characterize fragment binding.

Another advantage of the SPR-based drug screening platform lies in its ability to identify compounds that exhibit promiscuous binding to target proteins early in the drug discovery process. The mechanism of promiscuous binding is well documented (Shoichet BK, 2006) and elimination of promiscuous compounds is a crucial objective so as to provide high-quality fragment candidates for drug lead generation. Investigation of these compounds indicates that they often form soluble colloidal aggregates in solution with diameters of 30 – 400 nm (McGovern SL *et al.*, 2002). They are capable of binding protein targets with high stoichiometry and affinity resulting in steep dose-response curves (Walters W and Namchuk M, 2003) owing to nonspecific interactions. Promiscuous compounds achieve inhibition by masking protein surfaces and thereby preventing substrates or native ligands access to the binding site (McGovern SL *et al.*, 2003). Several studies have indicated that promiscuous binding is sensitive to ionic strength (McGovern SL *et al.*, 2002) and the presence of detergent (McGovern SL *et al.*, 2003; Ryan AJ *et al.*, 2003; Feng BY *et al.*, 2006) or BSA (Coan KE *et al.*,

2007). Therefore, the occurrence of non-specific binding can be reduced by using buffers at moderate ionic strength supplemented with 0.01% Triton X-100 in screening assay.

Based on observations in SPR-based fragment screening across multiple libraries, the probability of compound forming colloidal aggregate in solution increases substantially at and above 500 μM (Hamalainen MD *et al.*, 2008; Navratilova I *et al.*, 2010). The preference of fragment concentration for the primary screen has been reported in the range of millimolar (Hamalainen MD *et al.*, 2008) or around 100 μM (Navratilova I and Myszka D, 2006). Although higher fragment screening concentration benefits from larger binding signals for hit detection, it also increases the probability of compounds binding promiscuously to the target proteins. Conversely, lower screening concentration is viable for the primary screen as the goal is to identify potential hits and not observing their binding at saturation. Modern Biacore™ instruments have a short term noise of ± 0.2 RU, making the threefold signal-to-noise level equivalent to approximately 5% compound occupancy on a typically dense protein surface. In twenty SPR-based fragment screens across four target classes performed by Genentech at fragment concentration of 100 μM and 5% binding taken as a cutoff (Giannetti AM, 2011), the least potent hits detected were subsequently characterized and showed K_D values of 2 – 5 mM. In general, fragments weaker than this possess ligand efficiency too low to be of interest and represent challenges to obtaining ligand-bound protein co-crystal structures for structure-based drug design programme.

7.2.5 SPR-BASED FRAGMENT LIBRARY SCREENING AGAINST C3d

An initial SPR binding assay conducted at a fragment concentration of 100 μM did not give much observable binding to C3d in the primary screen. It appeared the interaction between fragments and C3d might be fairly weak and it was decided to raise the fragment concentration to 500 μM for hit detection. The expected maximal response (R_{max}) for a single-site interaction can be calculated with the following relation:

$$R_{\text{max}} (\text{RU}) = \text{Immobilization density (RU)} \times \frac{\text{Mass of ligand (Da)}}{\text{Mass of immobilized protein (Da)}}$$

Given the mass of fragments was <300 Da and the mass of C3d was 35 kDa, the $R_{\max}$ for a single-site interaction would be <60 RU at an immobilization density of 7000 RU. Likewise, the $R_{\max}$ for fragments of 150 Da would be 30 RU and a 20% binding site occupancy (*i.e.* 6 RU) was taken as a lower limit cutoff. A secondary dose-response experiment was used to confirm fragments with an observed binding response (R_{obs}) in the range of $6 < R_{\text{obs}} < 60$ RU form the primary screen.

In the primary screen, 92.4% of the fragments in the library showed significant interaction with C3d even in the presence of 0.01% Triton X-100 which was added to reduce non-specific binding. A majority of the cases exceeded the $R_{\max}$ which was 60 RU. The inability to saturate a single binding site suggested the fragments were binding non-specifically to the surface of C3d. These fragments appeared to bind C3d in a non-stoichiometric manner and represented 74.1% of the fragment library (Figure 7.4). Some of these fragments produced unusually high R_{obs} in the range of 200 – 600 RU, and up to 2000 RU in a few odd cases. Only 183 fragments generated R_{obs} between 6 and 60 RU, giving a hit rate of 18.3% in the SPR-based primary screen. These fragments were examined in a secondary dose-response experiment to test for saturability.

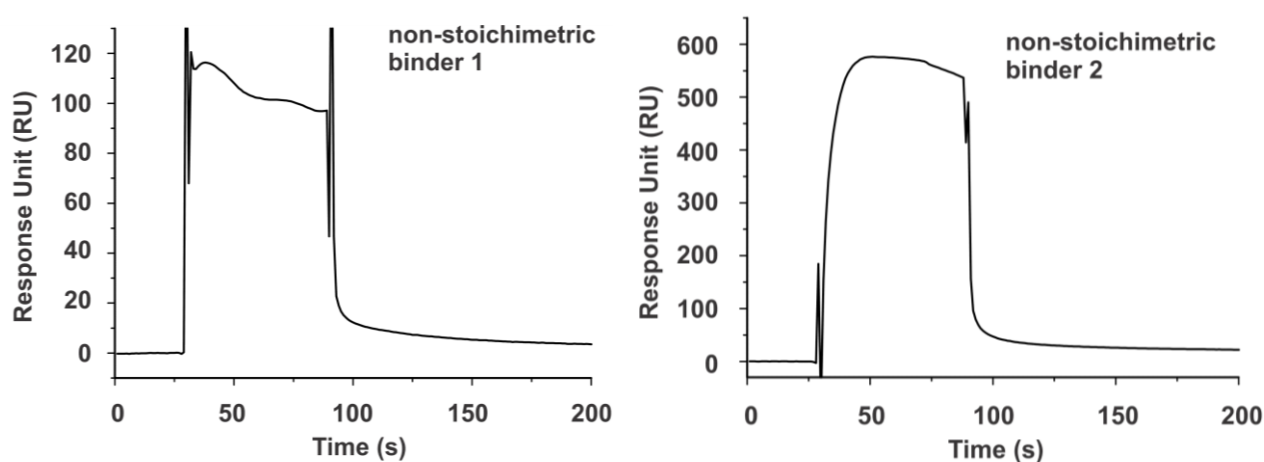


Figure 7.4 | Representative SPR sensorgrams showing fragments that appeared to bind promiscuously to C3d

The $R_{\max}$ of a typical fragment for a single-site interaction with C3d is 60 RU. Most fragments in the library (74.1%) produced binding responses which exceeded the $R_{\max}$, suggesting these fragments were occupying more than one binding site on C3d. Two examples of non-stoichiometric binders tested at 500 μM are shown with binding responses of approximately 120 μM and 600 μM , respectively.

7.2.6 SPR-BASED DOSE-RESPONSE EXPERIMENT

Binding response at six fragment concentrations was assessed in the dose-response experiment. A concentration series was set up with two-fold dilution for each of the 183 fragments tested, with the highest fragment concentration at 2 mM and the lowest at 63 μ M. Only nineteen fragments exhibited promising binding characteristic (Figure 7.5), nevertheless plotting R_{obs} versus concentration of fragment suggested the binding in all cases did not show any sign of saturability at the concentrations tested. The K_D could not be determined and binding stoichiometry clearly exceeded 1:1. It appeared that the binding affinities of these fragments were very weak and at least beyond 2 mM if tested with higher range of fragment concentrations. Such weak-binding fragments represent a less than ideal starting point for a structure-based drug design campaign. The remainder of the fragments tested in the secondary screen all displayed a common binding characteristic (Figure 7.6). Doubling of fragment concentration below 500 μ M did not have considerable effect on the binding responses produced by these fragments. However, binding responses noticeably increased above 500 μ M. Such behavior seemed to agree well with observations of increased probability of compound forming colloidal aggregate in solution ≥ 500 μ M (Hamalainen MD *et al.*, 2008; Navratilova I *et al.*, 2010). Therefore, these fragments were most likely binding promiscuously to the surface of C3d.

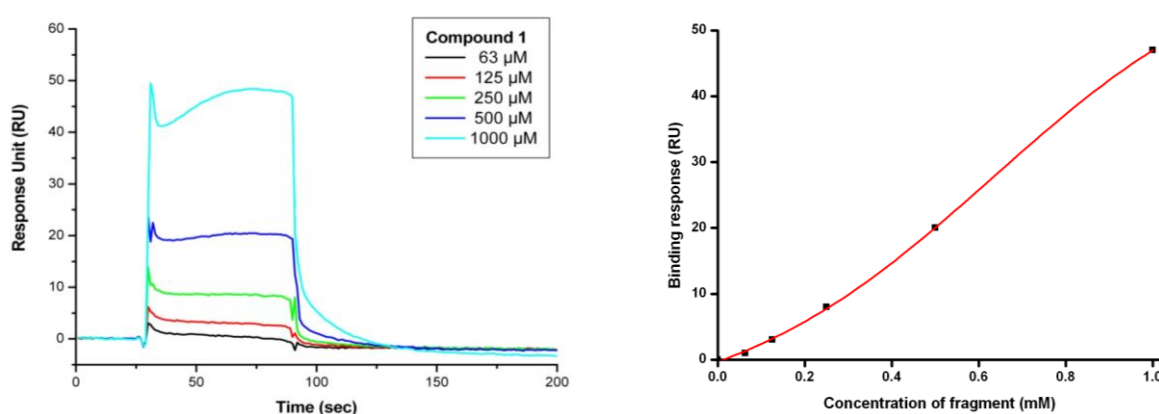


Figure 7.5 | Representative SPR sensorgrams showing fragments with decent binding characteristics to C3d

Examples of fragments that exhibited decent binding characteristics to C3d in SPR-based dose-response experiment. However, the binding lacked saturability at the concentrations tested.

Left = SPR sensorgram showing fragment binding to C3d tested in a series of fragment concentrations; *Right* = binding curve derived from the sensorgram on the left.

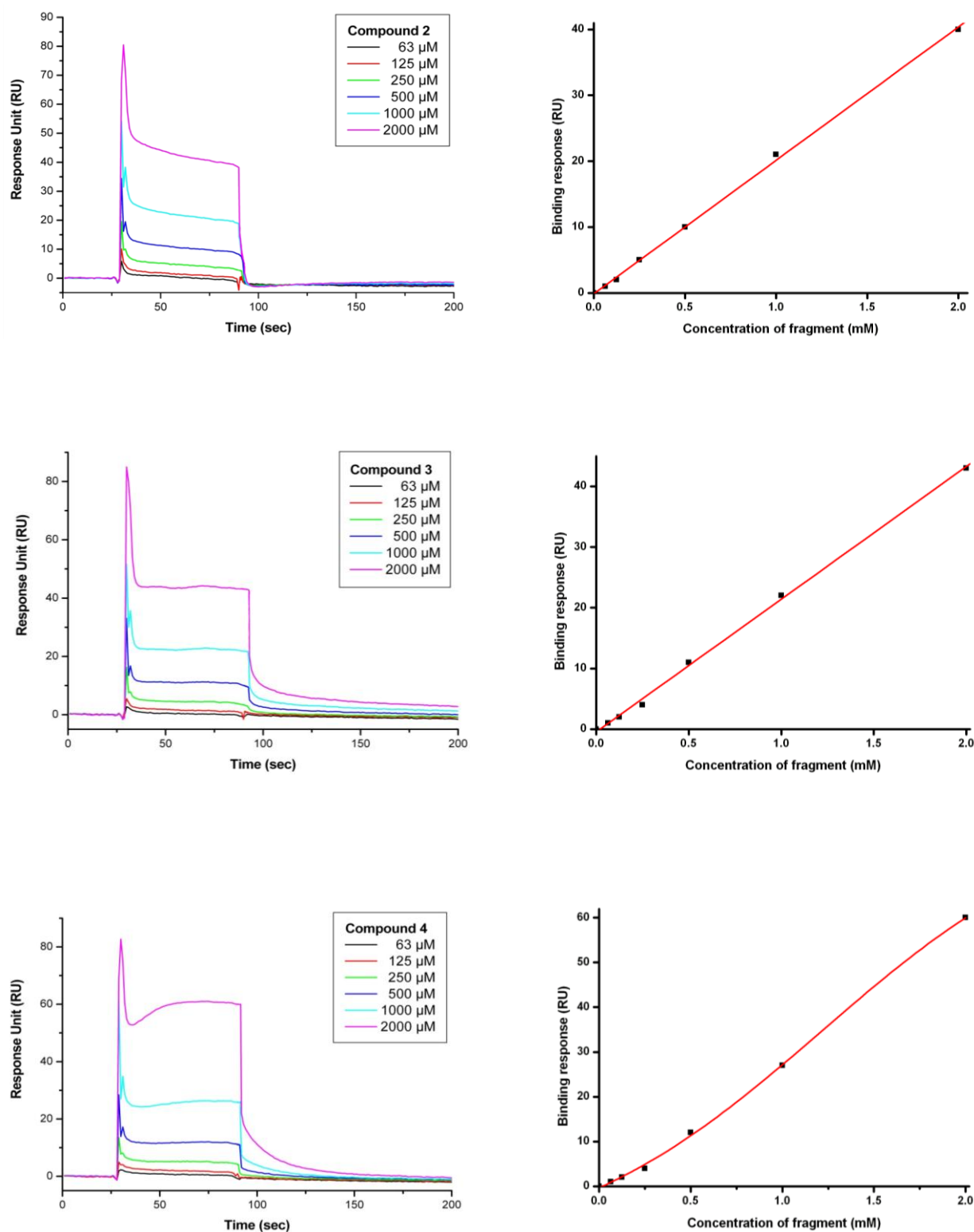


Figure 7.5 | Representative SPR sensorgrams showing fragments with decent binding characteristics to C3d

Examples of fragments that exhibited decent binding characteristics to C3d in SPR-based dose-response experiment. However, the binding lacked saturability at the concentrations tested.

Left = SPR sensorgram showing fragment binding to C3d tested in a series of fragment concentrations; *Right* = binding curve derived from the sensorgram on the left.

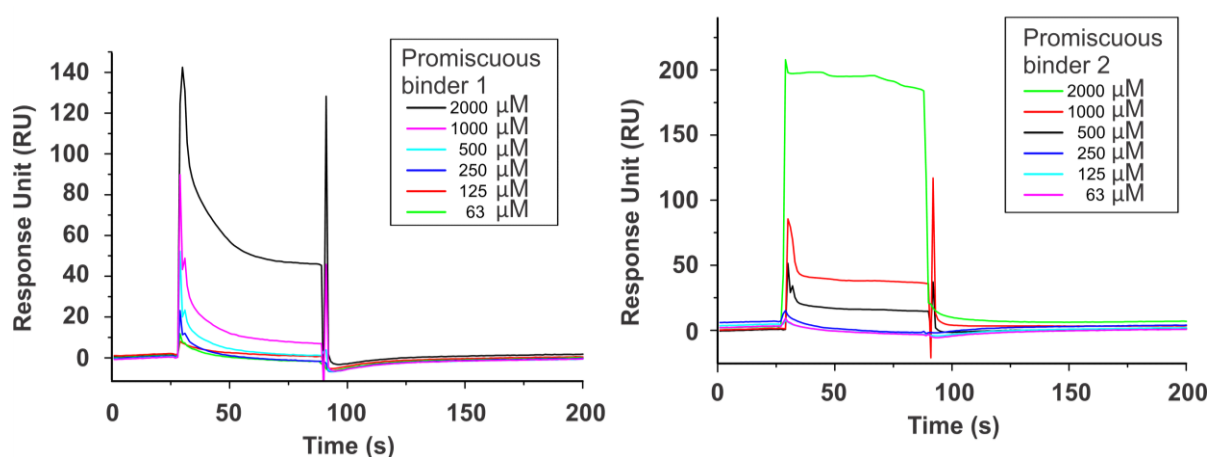


Figure 7.6 | Representative SPR sensorgram showing fragment binding promiscuously to C3d

Two examples of promiscuous binders tested for binding to C3d in a concentration series. Doubling of fragment concentration below 500 μM did not have considerable effect on the binding response, but binding stepped up considerably when the fragment concentration was raised beyond 500 μM .

7.2.7 DSF VERSUS SPR AS PRIMARY FRAGMENT LIBRARY SCREENING METHODS

Results of using DSF and SPR as primary fragment library screening methods against C3d are summarized in Figure 7.7. On the whole, SPR offered clean and quantitative single-step hit detection without the drawback of potential chemical interference in fluorescence-based technique such as DSF. The BIAcore™ T100 instrument provided good sensitivity to precisely measure binding response of fragments to C3d and gave clear confirmation to eliminate promiscuous binders in the primary screen. In contrast, most denaturation curves generated in DSF assay in screening this particular fragment library appeared peculiar. The multiphasic nature of the denaturation curves hindered curve-fitting and T_m estimation, making determination of ΔT_m induced by ligand binding less reliable. There was ambiguity in deducing the exact binding characteristics of most fragments, let alone fragments that were suspected to interfere with fluorescent readouts in DSF assay. It was necessary to employ an alternative screening technique (*e.g.* SPR) to clarify the binding characteristics of fragments.

One interesting observation from using DSF and SPR as primary fragment screen was that the two techniques did not give an identical set of hits. There was no strong correlation between fragments that were classified

as potential hits in DSF-based and SPR-based screening as illustrated with the following examples. One fragment did not produced any observable binding in the SPR screen but was shown to increased the thermal stability of C3d presumably by binding to the protein (Figure 7.8a). Likewise, another fragment that was classified as a potential hit in the SPR screen did not induced significant changes in the thermal stability of C3d in DSF assay (Figure 7.8b). Occasionally, hit detection from DSF-based and SPR-based primary screen was in agreement with each other (Figure 7.8c). SPR appeared to be a more robust technique in fragment screening and the secondary dose-response experiment offered affinity estimation, as well as discrimination of fragments that possessed good binding characteristics towards C3d. Therefore, only the nineteen potential hits selected from the SPR secondary dose-response experiment were brought forward to hit validation by NMR and X-ray crystallography.

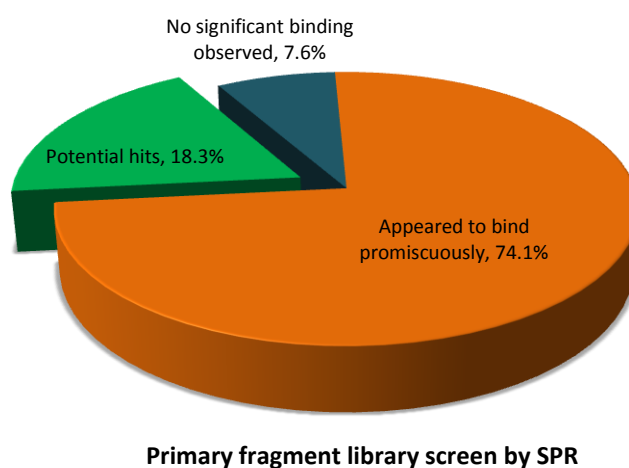
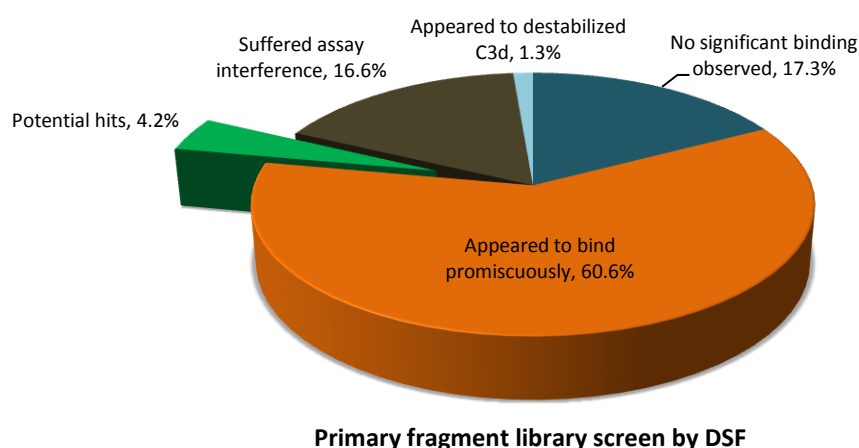


Figure 7.7 | DSF and SPR as primary fragment library screening methods for hit detection

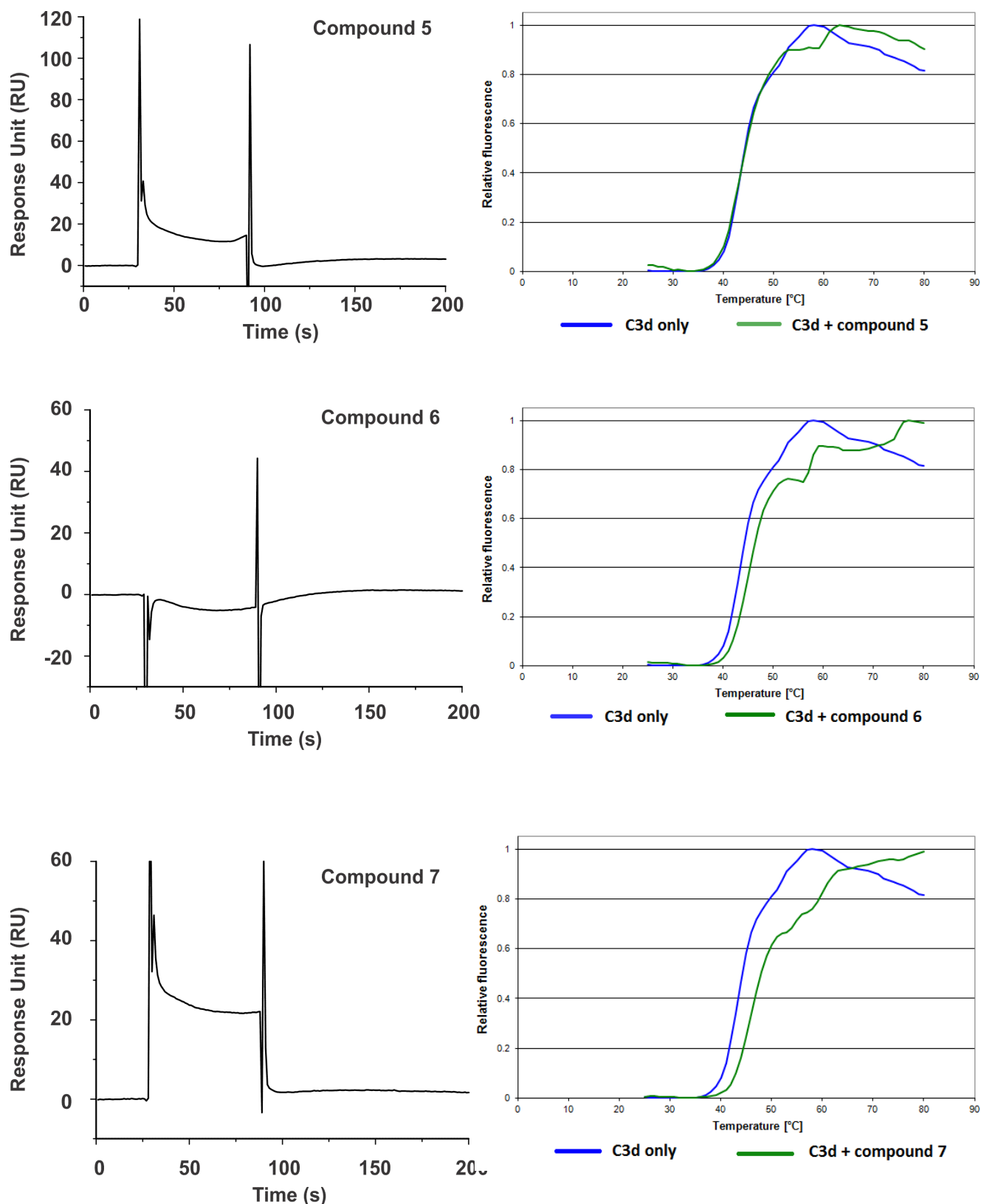


Figure 7.8 | Results of DSF-based and SPR-based fragment library screening were sometimes inconsistent

Compound 5 did not increase the thermal stability of C3d, but appeared to bind C3d as indicated by SPR sensorgram.

Compound 6 increased the thermal stability of C3d but did not appear to bind C3d as observed in SPR sensorgram.

Compound 7 showed consistent results in both DSF and SPR experiments.

7.3 FRAGMENT HIT VALIDATION

7.3.1 PROTEIN-BASED NMR SPECTROSCOPY

NMR is a versatile technique for monitoring ligand interaction with C3d in solution as discussed in Chapter V. There are two distinct categories of NMR experiments in the detection of low molecular-weight ligand binding, namely ligand-based and protein-based methods. The protein-based method was used here as a fragment hit validation step and changes in the HSQC spectra of C3d before and after the addition of fragment molecules were monitored. It provided residue-specific information and enabled the identification of ligand binding sites even in the absence of successful ligand-bound crystal structures. An HSQC spectrum of C3d acquired in the usual NMR phosphate buffer supplemented with 4% DMSO did not show any significant changes when compared to the native spectrum in the absence of DMSO (data not shown). Binding of the nineteen potential hits selected from the secondary SPR dose-response experiment were checked at a single concentration at 1 mM in 20 mM NaH₂PO₄, 150 mM NaCl, pH6.8 supplemented with 4% DMSO.

Figure 7.9 shows representative HSQC spectra of C3d before and after the addition of a potential fragment hit. Concentration of ¹⁵N-labeled C3d was held constant at 150 μM which was at the minimum end of protein concentration needed for typical HSQC experiments. Although low protein concentration diminished spectral resolution, it provided an efficient ¹⁵N-labeled protein reservoir for medium-throughput hit validation process. While six fragments did not produce significant chemical shift perturbation, other fragments induced changes in a small subset of peaks in the HSQC spectra of C3d. Unfortunately, peaks perturbed by these fragments did not appear to be identical to the subset of peaks that was perturbed by CD21 SCR1-2 (Chapter V). The findings suggested these fragments were interacting with C3d at sites that were different from CD21 SCR1-2. The amide resonance of solvent-exposed histidine 133 was notably perturbed by six fragments when they were titrated to ¹⁵N-labeled C3d sample (Figure 7.9). It was possible the change in the chemical shift of histidine 133 was associated with a change in sample pH. Although pH of fragment samples were checked with pH paper strips and adjusted as close to pH 7.0 as possible prior to NMR titration experiments, it was difficult to make precise pH adjustment to tiny microliter aliquot of fragment samples. The imidazole

sidechain of histidine has a pKa of 6.0 (Berg JM *et al.*, 2002). Since NMR titration experiments were done at physiologically relevant pH 6.8, relatively small alterations in sample pH could change the average charge of histidine and hence the chemical shifts of histidine 133. While some fragments only affected the chemical shifts of histidine 133, a few of them also induced chemical shift perturbation of glutamic acid 135. It appears that these fragments occupy a small binding patch comprised of histidine 133 and glutamic acid 135 on the surface of C3d, which nonetheless do not form the CD21 SCR1-2 binding site on C3d. Due to poor signal-to-noise ratio and high degree of overlaps in the HSQC spectra of C3d, it was not possible to obtain further residue specific information regarding the binding site of these fragments on C3d.

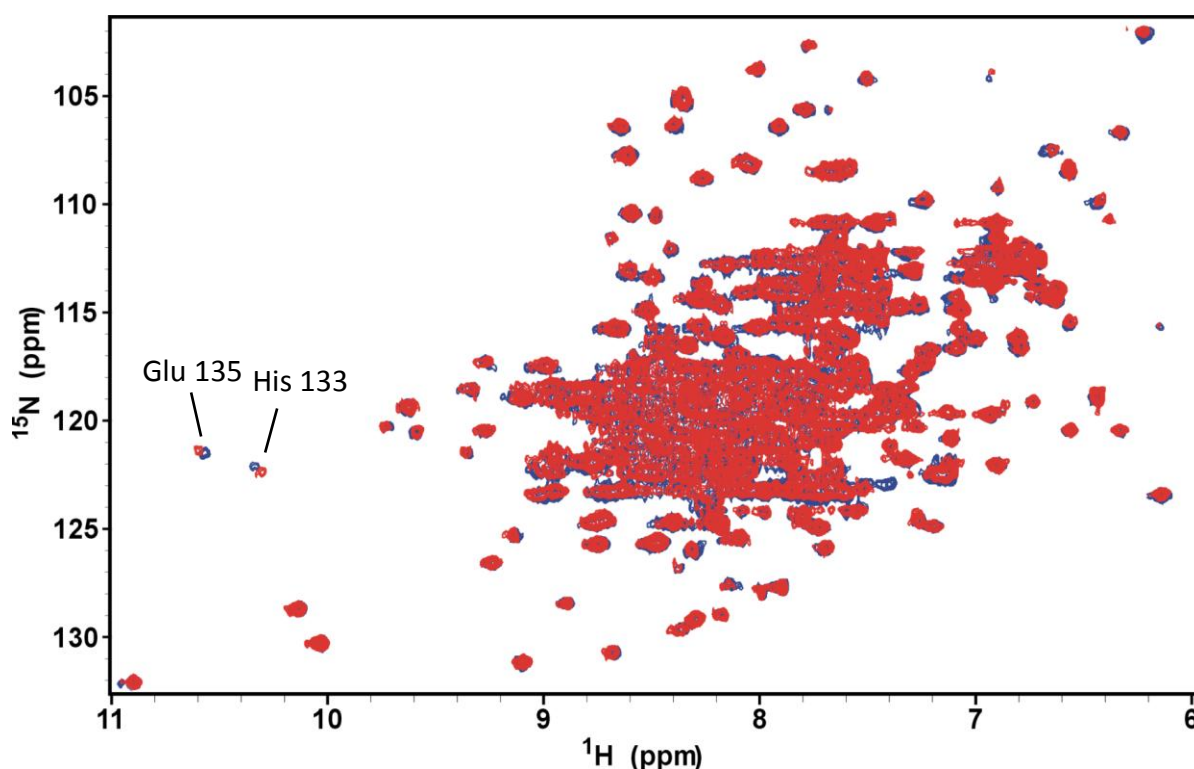


Figure 7.9 | Representative HSQC spectra of C3d collected on a Bruker 700 MHz spectrometer before and after the addition of potential fragment hits

Some potential fragment hits that were shown to bind C3d in dose-dependent SPR experiments induced changes in small subset of peaks in the HSQC spectra of C3d. Blue = starting spectra; Red = after addition of 1 mM fragment. However, peaks that were perturbed by these fragments were not the same as those that were perturbed by CD21 SCR1-2. Notably, histidine 133 and glutamic acid 135 (annotated) on C3d appeared to interact with some of the potential fragment hits. The overall spectral resolution is too poor to allow any further residue specific information to be obtained regarding the exact binding site of these fragments on C3d.

7.3.2 X-RAY CRYSTALLOGRAPHY

Since it was relatively straight forward to grow diffraction-quality C3d crystals (Chapter VI), soaking experiments were performed in attempt to obtain protein co-crystal structure with a bound fragment. Nineteen fragments which exhibited good binding characteristics in the secondary SPR dose-response experiment were selected for soaking trials. Since these fragments appeared to interact very weakly with C3d, it was necessary to use a fragment concentration as high as possible to realistically obtain successful soaks. General rule was to use a fragment concentration that was five times the K_D of a fragment to ensure full binding site occupancy. DMSO tolerability of C3d crystals was checked at 4% DMSO for eighteen hours and the crystals appeared intact without much sign of disintegration. The pH of fragment samples was checked with pH paper strips and adjusted to pH 7.0 when needed.

Initial fragment soaks were performed at fragment concentrations of 50 mM in a buffer containing mother liquor supplemented with 4% DMSO for six hours. The conditions were too harsh and most crystals fully dissolved or cracked in the end. Subsequent soaks were performed at fragment concentration of 25 mM for eighteen hours in the same buffer conditions. Seventeen fragments appeared to produce a darkening effect on C3d crystals. These crystals were no longer transparent and appeared as grey or brown in colour, suggesting precipitation might have occurred during the soak. The remaining seven fragments did not have nuisance effect on the crystals and only four of them induced minor cracks after the soak. All crystals were cryo-protected with buffer containing mother liquor with 25 mM fragment and supplemented with 4% DMSO and 15% ethyleneglycol. Data were collected at a synchrotron and crystal structures were solved as described previously. Unfortunately, all soaking experiments only yielded C3d apo-structures at 1.8 Å and no extra electron density for bound fragment was observed in the Fo-Fc map.

7.4 DISCUSSION

Several biophysical techniques were employed to screen and validate potential hits from a small fragment library against C3d. Unfortunately, the work did not yield novel inhibitors that would disrupt the interaction between C3d and CD21 SCR1-2 owing to inherent difficulties in targeting protein-protein interfaces (PPI). Developing PPI inhibitors remains a huge challenge even with recent advances in drug discovery approach and rich chemical resources in the pharmaceutical industry. Typically flat protein-protein interfaces often lack small, deep cavities at the surfaces of proteins that could serve as small molecule binding sites (Hopkins AL and Groom CR, 2002). The contact surfaces involved in PPIs are approximately 1500 – 3000 Å² (Jones S and Thornton JM, 1996; Conte LL *et al.*, 1999), which are substantially larger than the typical range of 300 – 1000 Å² observed in protein – small molecule interactions (Smith RD *et al.*, 2006; Cheng AC *et al.*, 2007). Such broad interaction areas mean that contacts in PPIs involve a large number of non-contiguous amino acid residues on the protein surface. For this reason, short peptides derived from contiguous sequences at the binding interface generally fail to generate suitable chemical templates to initiate structure-based drug design campaigns. Moreover, often times there are no mechanism-based or natural product lead structures that could serve as templates for designing small molecule inhibitors for PPIs. On rare occasions random empirical compound screens have identified novel inhibitors that disrupted protein-protein complexes, and a few of them have progressed through clinical trials (Wells JA and McClendon CL, 2007; Berg T, 2008). Recent successful examples from the literature include PPI inhibitors that targeted cytokine IL-2, cytokine TNF, anti-apoptotic Bcl-2 family proteins, oncogenic regulator HDM2, human papilloma virus transcription factor E2, bacterial replication proteins ZipA, polo-like kinase 1 PBD domain and leukemic protein CBFβ. In these examples, hits were identified from screening conventional HTS libraries of more than 250,000 compounds or fragment libraries of 7000 – 15,000 fragment molecules. The few leads identified were relatively hydrophobic, rigid and often found to contain multiple aromatic groups. Most final PPI drug candidates had a molecular weight <1000 Da and did not comply with Lipinski's rule of five (Sperandio O *et al.*, 2010). Computer algorithms that extensively analyse known PPI inhibitors from the literature and utilize in silico property filtering have been developed recently to enable design of PPI focused libraries (Reynes C *et al.*, 2010). Future

work in identifying potential inhibitors for the C3d-CD21 SCR1-2 interaction could benefit from screening larger PPI focused libraries against C3d.

With regards to screening techniques, it is unclear what might have caused the difference in hit detection using DSF and SPR in the current work. In general, coupling two or more different screening techniques in hit detection allows filtering of false positive hits or finding false negative hits that were misinterpreted with one single screening approach. There are experimental limitations and effective affinity ranges for detection of small molecule binding associated with each biophysical technique. Figure 7.10 was adapted from a review paper in *Methods in Enzymology* (Hubbard RE and Murray JB, 2011) and summarizes current hit detection methods and their sensitivity limit for detecting small molecule binding to protein targets.

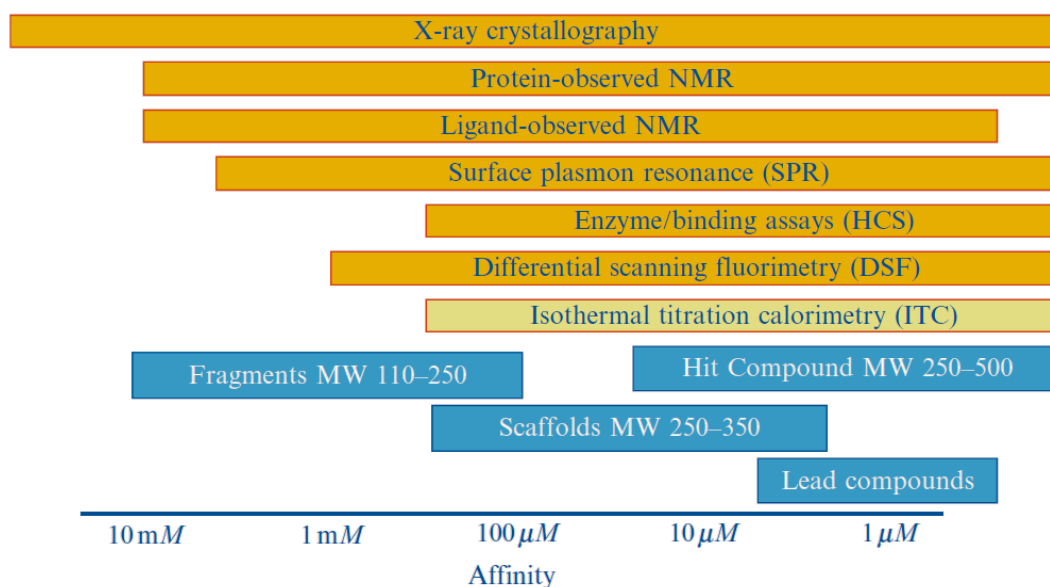


Figure 7.10 | Current hit detection methods and their sensitivity limit for detecting small molecule binding to proteins. Several biophysical methods currently used for hit detection (yellow) are shown against some indicative ranges for binding affinities of different categories of drug-like compounds (blue).

It is likely that the fragment library used in this study presents molecules with overall poor affinities to C3d. DSF is not sensitive enough to detect fragments with binding affinities greater than 1 mM to target proteins. As a result, only a handful of fragments were detected as potential hits using DSF. Although DSF provided a fast and economical screen, it only offers an apparent affinity as the readout and potential chemical interference in fluorescence-based screening technique seems inevitable. Furthermore, observations from

fragment-based screening practice at Vernalis indicates that the DSF assay, as well as the SPR-based assay to a lesser extent, appear to be more susceptible to aggregation of organic molecules while no compound aggregation was detected in ligand-based NMR experiments (Hubbard RE and Murray JB, 2011). Such pernicious behavior could be the cause of the overwhelming population of promiscuous binders detected in this particular fragment library against C3d. DSF is a less than ideal primary screening method against libraries with relatively low affinity to target proteins. However, it could be used routinely for rapid evaluation of more potent lead molecules generated from fragment hits in fragment evolution process. Findings in this chapter demonstrate that SPR and protein-based NMR spectroscopy offer complementary biophysical approaches in fragment hit detection. SPR provides information on binding behavior of fragments as well as an estimation of binding affinities of fragments to C3d. Protein-based NMR methods delivers information about binding modes and confirms specificity of fragments. Although no real hits were obtained against this challenging PPI target in the end, NMR was able to confirm most of the SPR hits exhibited some sort of binding to C3d.

Promiscuous binders appeared to dominate the fragment library used in this study when screened against C3d. Observations from SPR-based fragment screening practice at Roche indicate that a given compound may aggregate and bind promiscuously to one protein target but not to another (Giannetti AM *et al.*, 2008). Consequently, promiscuous binders are an inevitable feature of all screens as the behavior is target-dependent. Unless the same fragment library is screened against different protein targets, it would be debatable whether the library presented fragments that were mostly super-stoichiometric binders to protein targets, or C3d possessed inherent stickiness to a variety of molecule classes.

CHAPTER VIII

CONCLUSIONS AND FUTURE DIRECTIONS

Biophysical characterisation of protein – ligand interactions can provide vital information to dissect complex biochemical binding mechanisms. C3d has been shown to interact with a number of different protein ligands, namely CD21 SCR1-2 (Molina H *et al.*, 1995; Clemenza L and Isenman DE *et al.*, 2000; Zhang L *et al.*, 2007; Bomb M, 2008; Guthridge JM *et al.*, 2001; Kovacs JM *et al.*, 2009; Szakonyi G *et al.*, 2001; van den Elsen JMH and Isenman DE, 2011), *S.aureus* Efb-C (Hammel M *et al.*, 2007), *S.aureus* Ehp (Hammel M *et al.*, 2007), *S.aureus* Sbi (Clark EA *et al.*, 2011) and complement regulatory protein factor H (Morgan HP *et al.*, 2011). The aim of this thesis was to gain a detailed understanding of C3d/CD21 SCR1-2 interactions using different biophysical methods and to identify potential low molecular weight inhibitors of the interaction.

The prominent CD21 SCR1-2 binding site has been mapped on the surface of C3d which comprised a deep hydrophobic pocket surrounded by two negatively charged patches. Interaction analysis by mutagenesis, SPR,

ITC and NMR methods described in this thesis has confirmed this binding site being the only significant binding interface of the C3d/CD21 SCR1-2 interaction under physiologically relevant conditions. Collectively with the CD21 SCR1-2/C3d interaction analysis from the CD21 SCR1-2 perspective (Molina H *et al.*, 1995; Bomb M, 2008; Kovacs JM *et al.*, 2009), the solution studies from the C3d perspective presented in this thesis provide clarification to resolve controversy in two crystal structures of the C3d/CD21 SCR1-2 protein complex showing different binding modes. It is clear that crystal structure of the protein complex solved by van den Elsen JMH and Isenman DE in 2011 (PDB:3OED) represents a physiologically relevant binding mode of the C3d/CD21 SCR1-2 interaction. On contrary, the crystal structure of the protein complex solved by Szakonyi G *et al.* in 2001 (PDB:1GHQ) is widely recognized to be artifactual. This crystal structure illustrated a non-physiological binding mode that is not supported by residue specific information derived from mutagenesis (Chapter III) and NMR studies (Chapter V). Moreover, the C3d/CD21 SCR1-2 interaction has an unusual thermodynamic signature and is predominately entropically-driven as evident in microcalorimetric analysis (Chapter III). Entropically-driven interactions are typically dominated by either electrostatic or hydrophobic binding events, which is consistent with the binding interface observed in the new crystal structure of the protein complex (van den Elsen JMH and Isenman DE, 2011; PDB:3OED) and the binding site defined by mutagenesis (Chapter III; Bomb M, 2008; Molina H *et al.*, 1995; Clemenza L and Isenman DE, 2000) and NMR studies (Chapter V; Bomb M, 2008; Kovacs JM *et al.*, 2009).

Through the interactions with CD21 SCR1-2, C3d plays an integral part in bridging the innate and adaptive immune response (Fearon DT and Locksley RM, 1996). Co-engagement of antigen-bound C3d with CD21 SCR1-2 and the B cell receptor lowers the threshold for B cell activation by a magnitude of 10,000-fold (Dempsey PW *et al.*, 1996). Nonetheless, undue activation of humoral response leads to auto-antibody production and the pathological involvement of the C3d/CD21 SCR1-2 interaction is hold responsible for the development of rheumatoid arthritis (Del Nagro CJ *et al.*, 2005; Hietala MA *et al.*, 2002). Hence, identification of low molecular weight inhibitors to disrupt the C3d/CD21 SCR1-2 is of particular interest given the importance of this interaction in the pathological development of arthritis. Two biophysical techniques, SPR

and DSF, were employed to screen a small fragment library comprised one thousand chemical fragment against C3d. NMR chemical shifts perturbation analysis was used as a secondary hit validation methods to map the binding site of potential C3d-g fragments identified from SPR-based screening. Results from DSF assays suggested that most of the fragments appear to bind to C3d in a non-specific manner and such interactions are usually associated with compound aggregation (Shoichet BK, 2006). Many of these compounds are thought to destabilize protein tertiary structure and show non-ideal binding behaviours in DSF or other biophysical assays. SPR-based binding assays provide a high-throughput screening platform to monitor direct ligand binding to immobilized target protein. In addition, SPR-based dose-response experiments provide information about changes in binding behaviour over a range of compound concentrations. It is particularly useful in identifying promiscuous compounds that exhibit aggregation-dependent binding to C3d. Unfortunately, all potential C3d-binding compounds display curve shapes and binding behaviour that do not represent the desirable site-specific binding properties of druglike leads. The work did not yield novel inhibitors of the C3d/CD21 SCR1-2 interaction owing to inherent difficulties in targeting protein-protein interfaces (PPI). If a collection of C3d-binding fragments with desirable binding characteristics are indentified in the future, SPR-based competition assays will be performed to distinguish between site-selective and non-site-selective binders. In the quest for low molecular weight inhibitors to disrupt the C3d/CD21 SCR1-2 interaction, a novel nonamer peptide derived from factor H SCR20 was designed based on crystal structure of the C3d/factor H SCR19-20 (Morgan HP *et al.*, 2011; PDB:3OXU). SPR-based binding assay and NMR chemical shift perturbation analysis were used to evaluate the protein-ligand interaction between C3d and factor H-derived peptide. Although the peptide does bind to C3d weakly with a 1:1 stoichiometry, it appears to occupy a small hydrophobic pocket spatially away from the CD21 SCR1-2 binding surface on C3d. For this reason, the peptide was not exploited further as an potential starting point to develop inhibitors of the C3d/CD21 SCR1-2 interaction. Future works could include screening a PPI-focused chemical library with a larger collection of chemical scaffold against C3d, such as PPI chemical library (7000 compound) available from Asinex. PPI is engineered to bind proteins by nature and therefore peptide or antibodies are likely to have a higher successful rate to target the C3d/CD21 SCR1-2 binding

interface. Thus, it is sensible to screen C3d against a phage display library comprised of variants of a peptide or protein expressed on a phage virion, such as phage display library kit from New England Biolabs. A multitude screening approach with carefully selected libraries is likely to increase the probability of identifying potential hits that show desirable binding characteristics to C3d.

The NMR backbone resonance assignment of C3d makes it possible to define the extent of the CD21 SCR1-2 binding site on C3d under physiologically relevant ionic strength. Despite the resonance assignment only achieving 48% completion, the assigned resonances in the ^1H , ^{15}N -HSQC spectra of C3d contain sufficient information to monitor the binding between C3d and CD21 SCR1-2. Based on NMR chemical perturbation analysis from the C3d prospective, the delineated binding site for CD21 SCR1-2 is different to the one observed in crystal structure of the C3d/CD21 SCR1-2 protein complex solved by Szakonyi G *et al.* in 2001 (PDB:1GHQ), but is in full agreement with the one solved by van den Elsen JMH and Isenman DE in 2011 (PDB:3OXU). Full completion of the NMR backbone resonance assignment of C3d would yield a full picture of the C3d/CD21 SCR1-2 interaction in solution, nevertheless assigning the reminder of the backbone resonances is challenging and requires a much more sophisticated approach specially designed for high molecular weight systems. Future work could include selective methyl group labelling and methyl-TROSY NMR experiments to reduce spectral complexity. It enables the assignment of Ile, Leu and Val which are usually concentrated in the core of a protein. These residues are more difficult to assign in big proteins due to fairly undifferentiated chemical environment in the core. Details of the experiments were discussed at the end of Chapter IV. Completing NMR backbone assignment of C3d will continue to be a foremost priority in future work and it will be enable a full mapping of the CD21 SCR1-2 binding site on C3d. Furthermore, the assigned ^1H , ^{15}N -HSQC spectra of C3d can be used to validate other published crystal structures of C3d in complex with other protein ligands, such as C3d/Efb-C (PDB:2GOX), C3d/Ehp (PDB:2NOJ), C3d/Sbi-IV (PDB:2WY8) and C3d/factor H SCR19-20 (PDB:3OXU).

Bibliography

Abad MC, Askari H, O'Neill J, Klinger AL, Milligan C, Lewandowski F, Springer B, Spurlino J and Rentzeperis D. (2008). Structural determination of estrogen-related receptor γ in the presence of phenol derivative compounds. *J. Steroid Biochem. Mol. Biol* , **108**, 44–54.

Agostoni A, Cicardi M, Bergamaschini L, Boccassini G and Tucci A. (1980). C1-inhibitor concentrate for treatment of hereditary angioedema. *N Engl J Med* , **303**, 527.

Ahearn JM, Fischer MB, Croix D, Goerg S, Ma M, Xia J, Zhou X, Howard RG, Rothstein TL and Carroll MC. (1996). Disruption of the Cr2 locus results in a reduction in B-1a cells and in an impaired B cell response to T-dependent antigen. *Immunity* , **4** (3), 251-62.

Alexion Pharmaceuticals Inc. (2007). FDA approves Alexion's Soliris(TM) for all patients with PNH—first therapy approved for this rare and life-threatening blood disease. (*Alexion, Cheshire, CT*) , <http://ir.alexionpharm.com/releasedetail.cfm?ReleaseID=234156>.

Amero C, Schanda P, Dura A, Ayala I, Marion D, Franzetti B, Brutscher B and Boisbouvier J. (2009). Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *JACS* , **131** (10), 3448-3449.

Atkins KL, Burman JD, Chamberlain ES, Cooper JE, Poutrel B, Bagby S, Jenkins AT, Feil EJ and van den Elsen JM. (2008). S. aureus IgG-binding proteins SpA and Sbi: host specificity and mechanisms of immune complex formation. *Mol Immunol* , **45**, 1600–1611.

Avant Immunotherapeutics Inc. (2007). Press Release: AVANT Restructures Organization to Focus Resources on Core Programs and Operations. (*Avant Immunotherapeutics, Needham, MA*) , <http://phx.corporate-ir.net/phoenix.zhtml?c=93243&p=irol-newsArticle&t=Regular&id=985098&>.

Barrington RA, Schneider TJ, Pitcher LA, Mempel TR, Ma M, Barteneva NS and Carroll MC. (2009). Uncoupling CD21 and CD19 of the B-cell coreceptor. *Proc Natl Acad Sci USA* , **106** (34), 14490-5.

Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, Keystone EC, Genovese MC, Wasko MC, Moreland LW, Weaver AL, Markenson J and Finck BK. (2000). A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. *N Engl J Med* , **343** (22), 1586-93.

Bax A and Grzesiek S. (1993). Methodological advances in protein NMR. *Accounts Chem Res* , **26**, 131-138.

Bax A and Ikura I. (1991). An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α -carbon of the preceding residue in uniformly $^{15}\text{N}/^{13}\text{C}$ enriched proteins. *J. Biomol. NMR*, **1**, 99-104.

Berg JM, Tymoczko JL and Stryer L. (2002). Biochemistry, 5th ed. *W.H. Freeman*.

Berg T. (2008). Small-molecule inhibitors of protein-protein interactions. *Curr Opin Drug Discov Devel*, **11** (5), 666-74.

Beutler B. (2004). Inferences, questions and possibilities in Toll-like receptor signalling. *Nature*, **430**, 257–263.

Beutler B. (2003). Not “molecular patterns” but molecules. *Immunity*, **19**, 155–158.

Bhagavan NV. (2002). Chapter 35 Molecular immunology. In *Medical Biochemistry, 4th Edition* (pp. 853-854). Harcourt/Academic Press.

Bodenhausen D and Ruben DJ. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett*, **69**, 185-189.

Bohm HJ. (1995). Site-directed structure generation by fragment-joining. *Persp. Drug. Disc. Design*, **3**, 21–33.

Bomb M. (2008). Biophysical characterisation of the CD21 – C3d interaction . In *DPhil thesis - Characterising the interactions of CD21 with multiple ligands*.

Bonifati DM and Kishore U. (2007). Role of complement in neurodegeneration and neuroinflammation. *Mol Immunol*, **44**, 999–1010.

Bredius RG, Driedijk PC, Schouten MF, Weening RS and Out TA. (1992). Complement activation by polyclonal immunoglobulin G1 and G2 antibodies against *Staphylococcus aureus*, *Haemophilus influenzae* type B, and tetanus toxoid. *Infect Immun*, **60**, 4838–4847.

Brooks E, Herbert AP, Jenkins HT, Soares DC and Barlow PN. (2005). Opportunities for new therapies based on the natural regulators of complement activation. *Ann NY Acad Sci*, **1056**, 176–18.

Buerke M, Schwertz H, Seitz W, Meyer J and Darius H. (2001). Novel small molecule inhibitor of C1s exerts cardioprotective effects in ischemia-reperfusion injury in rabbits. *J Immunol*, **167**, 5375–5380.

Bureeva S, Andia-Pravdivy J and Kaplun A. (2005). Drug design using the example of the complement system inhibitors' development. *Drug Discov Today*, **10**, 1535–1542.

- Burman JD, Leung E, Atkins KL, O'Seaghdha MN, Lango L, Bernado P, Bagby S, Svergun DI, Foster TJ, Isenman DE and van den Elsen JM. (2008). Interaction of human complement with Sbi, a Staphylococcal immunoglobulin-binding protein: indications of a novel mechanism of complement evasion by *Staphylococcus aureus*. *J Biol Chem* , **283**, 17579–17593.
- Carroll MC. (2004). The complement system in regulation of adaptive immunity. *Nat Immunol* , **5** (10), 981-6.
- Carroll MC. (1998). The role of complement and complement receptors in induction and regulation of immunity. *Annu Rev Immunol* , **16**, 545-568.
- Carter RH and Fearon DT. (1992). CD19: Lowering the threshold for antigen receptor stimulation of B lymphocytes. *Science* , **256**, 105–107.
- Cavanagh J, Fairbrother W, Palmer AG and Sklepton NJ. (2007). Protein NMR Spectroscopy - Principles and Practice 2nd ed. Elsevier Academic Press.
- Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC and Huang ES. (2007). Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnol* , **25**, 71–75.
- Clark EA, Crennell S, Upadhyay A, Zozulya AV, Mackay JD, Svergun DI, Bagby S and van den Elsen JMH. (2011). A structural basis for Staphylococcal complement subversion: X-ray structure of the complement-binding domain of *Staphylococcus aureus* protein Sbi in complex with ligand C3d. *Mol Immunol* , **48** (4), 452–462.
- Clemenza L and Isenman DE. (2000). Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). *J Immunol* , **165**, 3839-3848.
- Clubb RT, Thanabal VT and Wagner G. (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**, 213-217.
- Coan KE and Shoichet BK. (2007). Stability and equilibria of promiscuous aggregates in high protein milieus. *Mol Biosyst* , **3**, 208-213.
- Coggins BE and Zhou P. (2003). PACES: Protein sequential assignment by computer-assisted exhaustive search. *J. Biomol. NMR* , **26**, 93-111.

- Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, Keystone EC, Loveless JE, Burmester GR, Cravets MW, Hessey EW, Shaw T, Totoritis MC; REFLEX Trial Group. (2006). Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum* , **54** (9), 2793-806.
- Comeau SR, Gatchell DW, Vajda S and Camacho CJ. (2004). ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Research* , **32**, W96-W99.
- Congreve M, Carr R, Murray C and Jhoti H. (2005). A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* , **8** (19), 876-877.
- Conte LL, Chothia C and Janin J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol* , **285**, 2177-2198.
- Correia JJ and Detrich W. (2008). *Biophysical tools for biologistsL Vol. 1 in vitro techniques*. USA: Academic Press, Elsevier.
- Cronstein BN. (2005). Low-dose methotrexate: a mainstay in the treatment of rheumatoid arthritis. *Pharmacol Rev* , **57** (2), 163-72.
- Cummings MD, Farnum MA and Nelen MI. (2006). Universal Screening Methods and Applications of ThermoFluor®. *J Biomol Screen* , **11** (7), 854-863.
- Daha MR, Fearon DT and Austen KF. (1976). C3 requirements for formation of alternative pathway C5 convertase. *J Immunol* , **117**, 630-634.
- Dam TK and Brewer CF. (2002). Thermodynamic Studies of Lectin-Carbohydrate Interactions by Isothermal Titration Calorimetry. *Chem Rev* , **102**, 387-429.
- De Serres J, Groner A and Lindner J. (2003). Safety and efficacy of pasteurized C1 inhibitor concentrate (Berinert P) in hereditary angioedema: a review. *Transfus Apher Sci* , 247-254.
- Del Nagro CJ, Kolla RV, Rickert RC. (2005). A Critical Role for Complement C3d and the B Cell Coreceptor (CD19/CD21) Complex in the Initiation of Inflammatory Arthritis. *J Immunol* , **175** (8), 5379-89.
- Dempsey PW, Allison ME, Akkaraju S, Goodnow CC and Fearon DT. (1996). C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science* , **271**, 348-350.

- Dodd I, Mossakowska DE, Camilleri P, Haran M, Hensley P, Lawlor EJ, McBay DL, Pindar W, and Smith RA. (1995). Overexpression in *Escherichia coli*, folding, purification, and characterization of the first three short consensus repeat modules of human complement receptor type 1. *Protein Expr Purif*, **6**, 727-736.
- Eftink M and Biltonen R. (1980). Thermodynamics of interacting biological systems. In *Biological Microcalorimetry* (pp. 343-412). Academic Press.
- Elliott MJ, Maini RN, Feldmann M, Kalden JR, Antoni C, Smolen JS, Leeb B, Breedveld FC, Macfarlane JD and Bijl H. (1994). Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor alpha (cA2) versus placebo in rheumatoid arthritis. *Lancet*, **344** (8930), 1105-10.
- Epstein J, Eichbaum QE, Sheriff S and Ezekowitz RAB. (1996). The collectins in innate immunity. *Curr Opin Immunol*, **8**, 29-35.
- Erlanson DA. (2006). Fragment-based lead discovery: a chemical update. *Curr Opin Biotechnol*, **17**, 643-652.
- Evans JNS. (1995). Biomolecular NMR Spectroscopy. *Oxford, Oxford University Press*.
- Farmer BT, Venters RA, Spicer LD, Wittekind MG and Müller L. (1992). A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules. *J. Biomol. NMR*, **2** (2), 195-202.
- Fearon DT and Austen KF. (1975). Initiation of C3 cleavage in the alternative complement pathway. *J Immunol*, **115**, 1357-1361.
- Fearon DT and Austen KF. (1975). Properdin: Binding to C3b and stabilization of the C3b-dependent C3 convertase. *J Exp Med*, **142**, 856-863.
- Fearon DT and Carter RH. (1995). The CD19/CR2/TAPA-1 complex of B lymphocytes: linking natural to acquired immunity. *Annu Rev Immunol*, **13**, 127-149.
- Fearon DT and Locksley RM. (1996). The instructive role of innate immunity in the acquired immune response. *Science*, **272**, 50-53.
- Feng BY and Shoichet BK. (2006). A detergent-based assay for the detection of promiscuous inhibitors. *Nat Protoc*, **1**, 550-553.
- Fingerroth JD, Clabby ML and Strominger JD. (1988). Characterization of a T-lymphocyte Epstein-Barr virus/C3d receptor (CD21). *J Virol*, **62**, 1442-1447.

Fink T, Bruggesser H and Reymond J. (2005). Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew. Chem. Int. Ed. Engl.* , **44**, 1504–1508.

Fischer MB, Ma M, Goerg S, Zhou X, Xia J, Finco O, Han S, Kelsoe G, Howard RG, Rothstein TL, Kremmer E, Rosen FS and Carroll MC. (1996). Regulation of the B cell response to T-dependent antigens by classical pathway complement. *J Immunol* , **157** (2), 549-56.

Flynn PF, Mattiello DL, Hill HDW and Wand AJ. (2000). Optimal use of cryogenic probe technology in NMR studies of proteins. *J. Am. Chem. Soc.* , **122**, 4823–4824.

Freeman R and Kupce E. (1997). Decoupling: theory and practice I. Current methods and recent concepts. *NMR Biomed* , **10**, 372-380.

Frei Y, Lambris JD and Stockinger B. (1987). Generation of a monoclonal antibody to mouse C5 application in an ELISA assay for detection of anti-C5 antibodies. *Mol Cell Probes* , **1**, 141–149.

Frisch C, Schreiber G, Johnson CM and Fersht AR. (1997). Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation. *J Mol Biol* , **267** (3), 696-706.

Fritsch R, Eselböck D, Skrinier K, Jahn-Schmid B, Scheinecker C, Bohle B, Tohidast-Akrad M, Hayer S, Neumüller J, Pinol-Roma S, Smolen JS and Steiner G. (2002). Characterization of autoreactive T cells to the autoantigens heterogeneous nuclear ribonucleoprotein A2 (RA33) and filaggrin in patients with rheumatoid arthritis. *J Immunol* , **169** (2), 1068-76.

Fujita T, Matsushita M and Endo Y. (2004). The lectin-complement pathway--its role in innate immunity and evolution. *Immunol Rev* , **198**, 185-202.

Furst D, Breedveld F, Kalden J, Smolen J, Burmester G, Dougados m, Emery P, Gibofsky A, Kavanaugh A, Keystone E, Klareskog L, Russell A, van de Putte LBA and Weisman M. (2003). Updated consensus statement on biological agents for the treatment of rheumatoid arthritis and other immune mediated inflammatory diseases (May 2003). *Ann Rheum Dis* , **62** S2, ii2–ii9.

Gabriel SE. (2001). The epidemiology of rheumatoid arthritis. *Rheum Dis Clin North Am* , **27**, 269–282.

Gardner KH and Kay LE. (1997). Production and Incorporation of ¹⁵N, ¹³C, ²H (¹H-δ1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J Am Chem Soc* , **119** (32), 7599-7600.

Gardner KH and Kay LE. (1998). The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struc* , **27**, 357-406.

Gardner KH and Kay LE. (1998). The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* , **27**, 357–406.

Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A and Kalodimos CG. (2007). Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* , **131** (4), 756-69.

Giannetti AM. (2011). From experimental design to validated hits a comprehensive walk-through of fragment lead identification using surface plasmon resonance. *Methods Enzymol* , **493**, 169-218.

Giannetti AM, Koch BD and Browner MF. (2008). Surface Plasmon Resonance Based Assay for the Detection and Characterization of Promiscuous Inhibitors. *J Med Chem* , **51** (3), 574–580.

Gilbert HE, Eaton JT, Hannan JP, Holers VM, Perkins SJ. (2005). Solution structure of the complex between CR2 SCR 1-2 and C3d of human complement: an X-ray scattering and sedimentation modelling study. *J Mol Biol* , **346** (3), 859-73.

Gill SC and von Hippel PH. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* , **182** (2), 319-325.

Goel N and Stephens S. (2010). Certolizumab Pegol. *mAbs* , **2** (2), 137 - 147.
<http://dx.doi.org/10.4161/mabs.2.2.11271>.

Goto NK, Gardner KH, Mueller GA, Willis RC and Kay LE. (1999). A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J Biomol NMR* , **13** (4), 369-74.

Grzesiek S and Bax A. (1992). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* , **99**, 201-207.

Grzesiek S and Bax A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* , **114**, 6291-6293.

Gulati P, Guc D, Lemercier C, Lappin D and Whaley K. (1994). Expression of the components and regulatory proteins of the classical pathway of complement in normal and diseased synovium. *Rheumatol Int* , **14** (1), 13-9.

Guntert P. (1998). Structure calculation of biological macromolecules from NMR. *Q Rev Biophys* , **31**, 145-237.

Guo RF and Ward PA. (2005). Role of C5a in inflammatory responses. *Annu Rev Immunol* , **23**, 821–852.

Guthridge JM, Rakstang JK, Young KA, Hinshelwood J, Aslam M, Robertson A, Gipson MG, Sarrias MR, Moore WT, Meagher M, Karp D, Lambris JD, Perkins SJ and Holers VM. (2001). Structural studies in solution of the recombinant N-terminal pair of short consensus/complement repeat domains of complement receptor type 2 (CR2/CD21) and interactions with its ligand C3dg. *Biochemistry* , **40** (20), 5931-41.

Hajduk PJ, Huth JR and Fesik SW. (2005). Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* , **48**, 2518–2525.

Hamalainen MD, Zhukov A, Ivarsson M, Fex T, Gottfries J and Karlsson R. (2008). Label-free primary screening and affinity ranking of fragment libraries using parallel analysis of protein panels. *J Biomol Screen* , **13** (3), 202-209.

Hämäläinen MD, Zhukov A, Ivarsson M, Fex T, Gottfries J, Karlsson R and Björnsne M. (2008). Label-free primary screening and affinity ranking of fragment libraries using parallel analysis of protein panels. *J Biomol Screen* , **13** (3), 202–209.

Hamel DJ and Dahlquist FW. (2005). The contact interface of a 120 kD CheA-CheW complex by methyl TROSY interaction spectroscopy. *JACS* , **127**, 9676-7.

Hammel M, Sfyroera G, Pyrpasopoulos S, Ricklin D, Ramyar KX, Pop M, Jin Z, Lambris JD and Geisbrecht BV. (2007). Characterization of Ehp, a secreted complement inhibitory protein from *Staphylococcus aureus*. *J Biol Chem* , **282** (41), 30051-61.

Hammel M, Sfyroera G, Ricklin D, Magotti P, Lambris JD and Geisbrecht BV. (2007). A structural basis for complement inhibition by *Staphylococcus aureus*. *Nat Immunol* , **8** (4), 430-7.

Hann M M, Leach AR and Harper G. (2001). Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* , **41**, 856–864.

Hannan JP, Young KA, Guthridge JM, Asokan R, Szakonyi G, Chen XS and Holers VM. (2005). Mutational analysis of the complement receptor type 2 (CR2/CD21)-C3d interaction reveals a putative charged SCR1 binding site for C3d. *J Mol Biol* , **346**, 845-858.

Hietala MA, Jonsson IM, Tarkowski A, Kleinau S and Pekna M. (2002). Complement deficiency ameliorates collagen-induced arthritis in mice. *J Immunol* , **169** (1), 454-9.

- Hoffmann JA, Kafatos FC, Janeway CA and Ezekowitz RA. (1999). Phylogenetic perspectives in innate immunity. *Science* , **284**, 1313–1318.
- Holland MC, Morikis D and Lambris JD. (2004). Synthetic small-molecule complement inhibitors. *Curr Opin Investig Drugs* , **5**, 1164–1173.
- Hopkins AL and Groom CR. (2002). The druggable genome. *Nature Rev Drug Discov* , **1**, 727–730.
- Horiuchi T, Takahashi M, Kikuchi J, Yokoyama S and Maeda H. (2005). Effect of dielectric properties of solvents on the quality factor for a beyond 900 MHz cryogenic probe model. *Magn. Reson.* , **174**, 34–42.
- Hosea SW, Brown EJ and Frank MM. (1980). The critical role of complement in experimental pneumococcal sepsis. *J Infect Dis* , **142**, 903–909.
- Hubbard RE and Murray JB. (2011). Experiences in fragment-based lead discovery. *Methods Enzymol* , **493**, 509–31.
- Ikura M, Kay LE and Bax A. (1990). A novel approach for sequential assignment of ¹H, ¹³C, and ¹⁵N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* , **29**, 4659–4667.
- Isaacson RL, Simpson PJ, Liu M, Cota E, Zhang X, Freemont P and Matthews S. (2007). A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J Am Chem Soc* , **129** (50), 15428–9.
- Isenman DE, Leung E, Mackay JD, Bagby S and van den Elsen JM. (2010). Mutational analyses reveal that the staphylococcal immune evasion molecule Sbi and complement receptor 2 (CR2) share overlapping contact residues on C3d: implications for the controversy regarding the CR2/C3d cocrystal structure. *J Immunol* , **184**, 1946–1955.
- Janeway CA and Medzhitov R. (2002). Innate immune recognition. *Annu Rev Immunol* , **20**, 197–216.
- Janssen BJ, Halff EF, Lambris JD and Gros P. (2007). Structure of compstatin in complex with complement component C3c reveals a new mechanism of complement inhibition. *J Biol Chem* , **282**, 29241–29247.
- Jelesarov I and Bosshard HR. (1999). Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J Mol Recognit* , **12**, 3–18.

- Johnston A, Gudjonsson JE, Sigmundsdottir H, Ludviksson BR and Valdimarsson H. (2005). The anti-inflammatory action of methotrexate is not mediated by lymphocyte apoptosis, but by the suppression of activation and adhesion molecules. *Clin Immunol* , **114** (2), 154-63.
- Jones S and Thornton JM. (1996). Principles of protein–protein interactions. *Proc Natl Acad Sci USA* , **93**, 13–20.
- Kalli KR, Ahearn JM, Fearon DT. (1991). Interaction of iC3b with recombinant isotypic and chimeric forms of CR2. *J Immunol* , **147** (2), 590-4.
- Katragadda M, Magotti P, Sfyroera G and Lambris JD. (2006). Hydrophobic effect and hydrogen bonds account for the improved activity of a complement inhibitor, compstatin. *J Med Chem* , **49**, 4616–4622.
- Kawasaki A, Takada H, Kotani S, Inai S, Nagaki K, Matsumoto M, Yokogawa K, Kawata S, Kusumoto S and Shiba T. (1987). Activation of the human complement cascade by bacterial cell walls, peptidoglycans, water-soluble peptidoglycan components, and synthetic muramylpeptides--studies on active components and structural requirements. *Microbiol Immunol* , **31** (6), 551-69.
- Kay LE and Gardner KH. (1997). Solution NMR spectroscopy beyond 25 kDa. *Curr Opin Struct Biol* , **7**, 722-31.
- Kay LE, Ikura M, Tschudin R and Bax A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* , **89** (3), 496-514.
- Kay LE, Keifer P and Saarinen T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* , **114**, 10663-10665.
- Keane J, Gershon S, Wise RP, Mirabile-Levens E, Kasznica J, Schwiertman WD, Siegel JN and Braun MM. (2001). Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med* , **345** (15), 1098-104.
- Keffer J, Probert L, Cazlaris H, Georgopoulos S, Kaslaris E, Kioussis D and Kollias G. (1991). Transgenic mice expressing human tumour necrosis factor: a predictive genetic model of arthritis. *EMBO J* , **10** (13), 4025–4031.
- Keller TH, Pichota A and Yin Z. (2006). practical view of 'druggability'. *Curr Opin Chem Biol* , **10**, 357–361.
- Kelly AE, Ou HD, Withers R and Dotsch V. (2002). Low-conductivity buffers for high-sensitivity NMR measurements. *J. Am. Chem. Soc* , **124**, 12013–12019.

- Kempeni J. (2002). Update on D2E7: a fully human anti-tumour necrosis factor alpha monoclonal antibody. *Ann Rheum Dis* , **59** S1, i44–i45.
- Kohl J. (2006). Drug evaluation: the C5a receptor antagonist PMX-53. *Curr Opin Mol Ther* , **8**, 529–538.
- Korb LC and Ahearn LM. (1997). C1q binds directly and specifically to surface blebs of apoptotic keratinocytes. *J Immunol* , **158**, 4525–4528.
- Kovacs H, Moskau D and Spraul M. (2005). Cryogenically cooled probes — a leap in NMR technology. *Prog. NMR Spectr* , **46**, 131–155.
- Kovacs JM, Hannan JP, Eisenmesser EZ and Holers VM. (2009). Mapping of the C3d Ligand Binding Site on Complement Receptor 2 (CR2/CD21) Using Nuclear Magnetic Resonance and Chemical Shift Analysis. *J Biol Chem* , **284** (14), 9513–9520.
- Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li K, Zheng J, Vakili P, Paschalidis IC and Vajda S. (2010). Achieving reliability and high accuracy in automated protein docking: Cluspro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. *Proteins: Structure, Function, and Bioinformatics* , **78**, 3124–3130.
- Kranz JK and Schalk-Hihi C. (2011). Protein thermal shifts to identify low molecular weight fragments. *Methods Enzymol* , **493**, 277–98.
- Kuo LC. (2011). How to avoid rediscovering the known. *Methods Enzymol* , **493**, 159–68.
- Lambris JD and Holers VM. (2000). *Therapeutic Interventions in the Complement System*. Totowa, NJ, USA: Humana Press.
- Law SKA and Dodds AW. (1997). The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Science* , **6**, 263–274.
- Leavitt S and Freire E. (2001). Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* , **11**, 560–566.
- Lee LY, Höök M, Haviland D, Wetsel RA, Yonter EO, Syribeys P, Vernachio J and Brown EL. (2004). Inhibition of complement activation by a secreted *Staphylococcus aureus* protein. *J Infect Dis* , **190** (3), 571–9.
- Lee LY, Liang X, Hook M and Brown EL. (2004). Identification and characterization of the C3 binding domain of the *Staphylococcus aureus* extracellular fibrinogen-binding protein (Efb). *J Biol Chem* , **279**, 50710–50716.

Lehrer RI and Ganz T. (1999). Antimicrobial peptides in mammalian and insect host defence. *Curr Opin Immunol* , **11**, 23–27.

Lehtinen MJ, Rops AL, Isenman DE, van der Vlag J and Jokiranta TS. (2009). Mutations of factor H impair regulation of surface-bound C3b by three mechanisms in atypical hemolytic uremic syndrome. *J Biol Chem* , **284** (23), 15650-8.

Li JS, Jaggars J and Anderson PA. (2006). The use of TP10, soluble complement receptor1 in cardiopulmonary bypass. *Expert Rev Cardiovasc Ther* , **4**, 649–654.

Li K, Okemefuna AI, Gor J, Hannan JP, Asokan R, Holers VM, Perkins SJ. (2008). Solution structure of the complex formed between human complement C3d and full-length complement receptor type 2. *J Mol Biol* , **384** (1), 137-50.

Lin Z, Schwarz FP and Eisenstein E. (1994). The hydrophobic nature of GroEL-substrate binding. *J Biol Chem* , **270**, 1011-1014.

Lipinski CA, Lombardo F, Dominy BW and Feeney PJ. (1996). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* , **23**, 3–25.

Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, Smolen JS, Weisman M, Emery P, Feldmann M, Harriman GR, Maini RN; Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. (2000). Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. *N Engl J Med* , **343** (22), 1594-602.

Liszewski MK and Atkinson JP. (1996). Membrane cofactor protein (MCP; CD46):Isoforms differ in protection against the classical pathway of complement. *J Immunol* , **156**, 4415–4421.

Liszewski MK, Farries TC, Lublin DM, Rooney IA and Atkinson JP. (1996). Control of the complement system. *Adv Immunol* , **61**, 201–283.

Lo MC, Aulabaugh A, Jina G, Cowling R, Bardb J, Malamasc M and Ellestad G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem* , **332**, 153–159.

Longhi MP, Harris CL, Morgan BP and Gallimore A. (2006). Holding T cells in check—a new role for complement regulators? *Trends Immunol* , **27**, 102-108.

Lundkvist J, Kastäng F and Kobelt G. (2008). The burden of rheumatoid arthritis and access to treatment: health burden and costs. *Eur J Health Econ* , **8** (Suppl 2) (1618-7598), S49-S60.

- Lyubchenko T, Dal Porto JM, Holers VM and Cambier JC. (2007). Cutting edge: Complement (C3d)-linked antigens break B cell anergy. *J Immunol* , **179** (5), 2695-9.
- Makrides SC. (1998). Therapeutic inhibition of the complement system. *Pharmacol Rev* , **50**, 59–87.
- Malmqvist M. (1993). Biospecific interaction analysis using biosensor technology. *Nature* , **361**, 186-187.
- Manderson AP, Botto M and Walport MJ. (2004). The role of complement in the development of systemic lupus erythematosus. *Annu Rev Immunol* , **22**, 431–456.
- Manna SK and Aggarwal BB. (1999). Immunosuppressive leflunomide metabolite (A77 1726) blocks TNF-dependent nuclear factor- κ B activation and gene expression. *J Immunol* , **162**, 2095–2102.
- Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AG and Clore GM. (1989). Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy. *Biochemistry* , **28** (15), 6150–6156.
- Marion D, Kay LE, Sparks SW, Torchia DA and Bax A. (1989). Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc.* , **111** (4), 1515–1517.
- Martin YC. (1997). Challenges and prospects for computational aids to molecular diversity. *Perspect. Drug Discov. Des.* , **7-8**, 159–172.
- Matsumoto AK, Kopicky-Burd J, Carter RH, Tuveson DA, Tedder TF and Fearon DT. (1991). Intersection of the complement and immune systems: a signal transduction complex of the B lymphocyte-containing complement receptor type 2 and CD19. *J Exp Med* , **173** (1), 55-64.
- Matsushita M and Fujita T. (2002). The role of ficolins in innate immunity. *Immunobiol* , **205**, 490–497.
- Matulis D, Kranz JK, Salemm FR and Todd MJ. (2005). Thermodynamic stability of carbonic anhydrase: Measurements of binding affinity and stoichiometry using thermofluor. *Biochemistry* , **44** (13), 5258-5266.
- Matulis D, Rouzina I and Bloomfield VA. (2000). Thermodynamics of DNA binding and condensation: isothermal titration calorimetry and electrostatic mechanism. *J Mol Biol* , **296**, 1053-1063.
- Mazumdar S and Greenwald D. (2009). Golimumab. *mAbs* , **1** (5), 422 - 431.
<http://dx.doi.org/10.4161/mabs.1.5.9286>.

McGovern SL, Caselli E, Grigorieff N and Shoichet BK. (2002). A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J Med Chem* , **45**, 1712-1722.

McGovern SL, Helfand BT, Feng B and Shoichet BK. (2003). A specific mechanism for non-specific inhibition. *J Med Chem* , **46**, 4265-4272.

McIntosh LP and Dahlquist FW. (1990). Biosynthetic incorporation of ¹⁵N and ¹³C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys* , **23**, 1–38.

Medzhitov R and Janeway CA Jr. (1997). Innate immunity: the virtues of a nonclonal system of recognition. *Cell* , **91**, 295–298.

Medzhitov R, Preston-Hurlburt P and Janeway CA. (1997). A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* , **388**, 394–397.

Meri S and Pangburn MK. (1994). Regulation of alternative pathway complement activation by glycosaminoglycans: specificity of the polyanion binding site on factor H. *Biochem Biophys Res Commun* , **198** (1), 52-9.

Miranker A. and Karplus M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* , **11**, 29–34.

Mohan N, Edwards ET, Cupps TR, Oliverio PJ, Sandberg G, Crayton H, Richert JR and Siegel JN. (2001). Demyelination occurring during anti-tumor necrosis factor alpha therapy for inflammatory arthritides. *Arthritis Rheum* , **44** (12), 2862-9.

Molina H, Holers VM, Li B, Fung Y, Mariathasan S, Goellner J, Strauss-Schoenberger J, Karr RW and Chaplin DD. (1996). Markedly impaired humoral immune response in mice deficient in complement receptors 1 and 2. *Proc Natl Acad Sci USA* , **93**, 3357–3361.

Molina H, Perkins SJ, Guthridge J, Gorka J, Kinoshita T and Holers VM. (1995). Characterization of a complement receptor 2 (CR2, CD21) ligand binding site for C3. An initial model of ligand interaction with two linked short consensus repeat modules. *J Immunol* , **154** (10), 5426-35.

Moreland L, Bate G and Kirkpatrick P. (2005). Abatacept. *Nat Rev Drug Disc* , **5** (3), 185–186.

Moreland LW, Baumgartner SW, Schiff MH, Tindall EA, Fleischmann RM, Weaver AL, Ettlinger RE, Cohen S, Koopman WJ, Mohler K, Widmer MB and Bloch CM. (1997). Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75)-Fc fusion protein. *N Engl J Med* , **337** (3), 141-7.

- Morgan BP, Daniels RH and Williams BD. (1988). Measurement of terminal complement complexes in rheumatoid arthritis. *Clin Exp Immunol* , **73** (3), 473-8.
- Morgan HP, Schmidt CQ, Guariento M, Blaum BS, Gillespie D, Herbert AP, Kavanagh D, Mertens HDT, Svergun DI, Johansson CM, Uhrin D, Barlow PN and Hannan JP. (2011). Structural basis for engagement by complement factor H of C3b on a self surface. *Nat Struct Mol Biol* , **18**, 463–470.
- Morikis D and Lambris JD. (2004). The electrostatic nature of C3d-complement receptor 2 association. *J Immunol* , **172**, 7537-7547.
- Müller-Eberhard HJ. (1988). Molecular organization and function of the complement system. *Annu Rev Biochem* , **57**, 321-47.
- Müller-Eberhard HJ. (1985). The killer molecule of complement. *J Invest Dermatol* , **85**, 47s–52s.
- Murphy KP, Freire E and Paterson Y. (1995). Configurational effects in antibody-antigen interactions studied by microcalorimetry. *Proteins* , **21**, 83-90.
- Murphy KP, Xie D, Garcia KC, Amzel LM, Freire E. (1993). Structural energetics of peptide recognition: angiotensin II/antibody binding. *Proteins* , **15**, 113-120.
- Murphy KP, Xie D, Thompson KS, Amzel LM, Freire E. (1994). Entropy in biological binding process: estimation of translational entropy loss. *Proteins* , **18**, 63-67.
- Myszka DG and Rich RL. (2000). Implementing surface plasmon resonance biosensors in drug discovery. *Pharm Sci Technolo Today* , **3** (9), 310-317.
- Nagar B, Jones RG, Diefenbach RJ, Isenman DE and Rini JM. (1998). X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* , **280**, 1277-1281.
- Narayana SV, Babu YS and Volanakis JE. (2000). Inhibition of complement serine proteases as a therapeutic strategy. In *Therapeutic Interventions in the Complement System* (eds Lambris JD & Holers VM) (pp. 57–74). Totowa, New Jersey, USA: Humana Press.
- Navratilova I and Hopkins AL. (2010). Fragment screening by surface plasmon resonance. *Acs Med. Chem. Lett* , **1**, 44-48.
- Navratilova I and Myszka D. (2006). Investigating biomolecular interactions and binding properties using SPR biosensors. In *Surface Plasmon Resonance Based Sensors*. Homola J, ed: Springer, New York.

- Neth O, Jack DL, Johnson M, Klein NJ and Turner MW. (2002). Enhancement of complement activation and opsonophagocytosis by complexes of mannose-binding lectin with mannose-binding lectin-associated serine protease after binding to *Staphylococcus aureus*. *J Immunol* , **169**, 4430–4436.
- Niesen FH, Berglund H and Vedadi M. (2008). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* , **2** (9), 2212-2221.
- Nordstrom H, Gossas T, Hamalainen M, Kallblad P, Nystrom S and Wallberg H. (2008). Identification of MMP-12 inhibitors by using biosensor-based screening of a fragment library. *J Med Chem* , **51**, 3449-3459.
- O'Brien R, Chowdhry BZ and Ladbury JE. (2001). Protein-ligand interactions: Hydrodynamics and Calorimetry. Oxford University Press, 263-286.
- Olsson TSG, Williams MA, Pitt WR and Ladbury JE. (2008). The Thermodynamics of Protein–Ligand Interaction and Solvation: Insights for Ligand Design. *J Mol Biol* , **384** (4), 1002-1017.
- Palma M, Nozohoor S, Schennings T, Heimdahl A and Flock JI. (1996). Lack of the extracellular 19-kilodalton fibrinogen-binding protein from *Staphylococcus aureus* decreases virulence in experimental wound infection. *Infect Immun* , **64** (12), 5284–5289.
- Pancer Z, Amemiya CT, Ehrhardt GRA, Ceitlin J, Gartland GL and Cooper MD. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* , **430**, 174-180.
- Pangburn MK. (2000). Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology* , **49** (1-2), 149-57.
- Pangburn MK, Schreiber RD and Müller-Eberhard HJ. (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** (1), 257-70.
- Parker CJ. (1992). Regulation of complement by membrane proteins: An overview. *Curr Top Microbiol Immunol* , **178**, 1-7.
- Pepys MB. (1976). Role of complement in the induction of immunological responses. *Transpl Rev* , **32**, 93–120.
- Perozzo R, Folkers G and Scapozza L. (2004). Thermodynamics of Protein-Ligand Interactions: History, Presence and Future Aspects. *Journal of receptors and signal transduction* , **24** (1), 1-52.

Pervushin K, Riek R, Wider G and Wüthrich K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* , **94** (23), 12366-71.

Pettigrew DM, Williams DT, Kerrigan D, Evans DJ, Lea SM, and Bhella D. (2006). Structural and functional insights into the interaction of echoviruses and decay-accelerating factor. *J Biol Chem* , **281**, 5169-5177.

Pierce MM, Raman CS and Nall BT. (1999). Isothermal titration calorimetry of protein-protein interactions. *Methods* , **19**, 213-221.

Pincus TE, Brooks RH and Callahan LF. (1994). Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. *Ann Intern Med* , **120**, 26–34.

Pugner KM, Scott DI, Holmes JW and Hieke K. (2000). The costs of rheumatoid arthritis: an international long-term view. *Semin Arthritis Rheum* , **29**, 305–320.

Raman B, Ramakrishna T and Rao CM (1996) Refolding of denatured/reduced lysozyme at high concentrations. *J Biol Chem*, **271**, 17067-72

Rattle H. (1995). An NMR primer for life scientists. *Partnership Press* .

Raulet DH. (2004). Interplay of natural killer cells and their receptors with the adaptive immune response. *Nat Immunol* , **5**, 996–1002.

Rehm T, Huber R and Holak TA. (2002). Application of NMR in Structural Proteomics: Screening for Proteins Amenable to Structural Analysis. *Structure* , **10**, 1613–1618.

Reid K and Turner M. (1994). Mammalian lectins in activation and clearance mechanisms involving the complement system. *Semin Immunopathol* , **15**, 307-26.

Reid KBM and Porter RR. (1981). The proteolytic activation systems of complement. *Annu Rev Biochem* , **50**, 433-64.

Rekharsky MV and Inoue Y. (2002). Solvent and Guest Isotope Effects on Complexation Thermodynamics of α -, β -, and 6-Amino-6-deoxy- β -cyclodextrins. *J Am Chem Soc* , **124**, 12361–12371.

Religa TL, Sprangers R and Kay LE. (2010). Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* , **328** (5974), 98-102.

- Reynès C, Host H, Camproux AC, Laconde G, Leroux F, Mazars A, Deprez B, Fahraeus R, Villoutreix BO and Sperandio O. (2010). Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. *PLoS Comput Biol*, **6** (3), e1000695.
- Ricklin D and Lambris JD. (2007). Complement-targeted therapeutics. *Nat Biotechnol*, **25** (11), 1265-75.
- Ricklin D, Ricklin-Lichtsteiner SK, Markiewski MM, Geisbrecht BV and Lambris JD. (2008). Cutting edge: members of the *Staphylococcus aureus* extracellular fibrinogen-binding protein family inhibit the interaction of C3d with complement receptor 2. *J Immunol*, **181**, 7463–7467.
- Rudolph MG and Wilson IA. (2002). The specificity of TCR/pMHC interaction. *Curr Opin Immunol*, **14**, 52–65.
- Ruschak AM and Kay LE. (2010). Methyl groups as probes of supra-molecular structure, dynamics and function. *J Biomol NMR*, **46** (1), 75-87.
- Ryan AJ, Gray NM, Lowe PN, and Chung CW. (2003). Effect of Detergent on “Promiscuous” Inhibitors. *J. Med. Chem*, **46** (16), 3448–3451.
- Sahu A and Lambris JD. (2000). Complement inhibitors: a resurgent concept in anti-inflammatory therapeutics. *Immunopharmacology*, **49**, 133–148.
- Sahu A, Kay BK and Lambris JD. (1996). Inhibition of human complement by a C3-binding peptide isolated from a phage-displayed random peptide library. *J Immunol*, **157**, 884–891.
- Salzmann M, Pervushin K, Wider G, Senn H and Wüthrich K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA*, **95** (23), 13585-90
- Sambrook J and Russell D. (2001). In *Molecular Cloning: A Laboratory Manual, 3rd Edition*. New York: Cold Spring Harbor Laboratory Press.
- Samuels J, Ng YS, Coupillaud C, Paget D and Meffre E. (2005). Impaired early B cell tolerance in patients with rheumatoid arthritis. *J Exp Med*, **201** (10), 1659-67.
- Schellekens GA, de Jong BA W and van den Hoogen F. (1998). Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. *J Clin Invest*, **101**, 273–281.
- Schnare M, Barton GM, Holt AC, Takeda K, Akira S and Medzhitov R. (2001). Toll-like receptors control activation of adaptive immune responses. *Nat Immunol*, **2** (10), 947-50.

- Schreiber G and Fersht AR. (1995). Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J Mol Biol* , **248**, 478-486.
- Schuck P. (1997). Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. *Annu Rev Biophys Biomol Struct* , **26**, 541-566.
- Schuffenhauer A et al. (2005). Library design for fragment based screening. *Curr. Top. Med. Chem.* , **5**, 751-762.
- Senisterra GA, Hong BS, Park HW and Vedadi M. (2008). Application of high-throughput isothermal denaturation to assess protein stability and screen for ligands. *J Biomol Screen* , **13** (5), 337-342.
- Seya T and Nagasawa S. (1985). Limited proteolysis of complement protein C3b by regulatory enzyme C3b inactivator: isolation and characterization of a biologically active fragment, C3dg. *J Biochem* , **97** (1), 373-82.
- Shai Y. (1999). Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α -helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim Biophys Acta* , **1462**, 55-70.
- Shoichet BK. (2006). Screening in a spirit haunted world. *Drug Discovery Today* , **11**, 607-615.
- Siemasko K, Chong AS, Jäck HM, Gong H, Williams JW and Finnegan A. (1998). Inhibition of JAK3 and STAT6 tyrosine phosphorylation by the immunosuppressive drug leflunomide leads to a block in IgG1 production. *J Immunol* , **160** (4), 1581-8.
- Sim RB and Tsiftoglou SA. (2004). Proteases of the complement system. *Biochem Soc Trans* , **32**, 21-27.
- Sittampalam GS, Ellis RM, Miner DJ, Rickard EC, Clodfelter DK. (1988). Evaluation of amino acid analysis as reference method to quantitate highly purified proteins. *J Assoc Off Anal Chem* , **71** (4), 833-838.
- Smith RD, Hu L, Falkner JA, Benson ML, Nerothin JP and Carlson HA. (2006). Exploring protein-ligand recognition with Binding MOAD. *J Mol Graph Model* , **24** (6), 414-25.
- Smolen JS, Kalden JR, Scott DL, Rozman B, Kvien TK, Larsen A, Loew-Friedrich I, Oed C and Rosenberg R. (1999). Efficacy and safety of leflunomide compared with placebo and sulphasalazine in active rheumatoid arthritis: a double-blind, randomised, multicentre trial. European Leflunomide Study Group. *Lancet* , **353** (9149), 259-66.
- Sperandio O, Reynès CH, Camproux AC and Villoutreix BO. (2010). Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov Today* , **15** (5-6), 220-9.

- Spitzer A, Mitchell LM, Atkinson JP and Hourcade DE. (2007). Properdin can initiate complement activation by binding specific target surfaces and providing a platform for de novo convertase assembly. *J Immunol* , **179**, 2600–2608.
- Sprangers R and Kay LE. (2007). Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* , **445**, 618-622.
- Srinivas V, Bhanuprakash Reddy G and Surolia A. (1999). A predominantly hydrophobic recognition of H-antigenic sugars by winged bean acidic lectin: a thermodynamic study. *FEBS Lett* , **450**, 181-185.
- Stahl PD and Ezekowitz RA. (1998). The mannose receptor is a pattern recognition receptor involved in host defense. *Curr Opin Immunol* , **10**, 50–55.
- Steiner G and Smolen JS. (2002). Autoantibodies in rheumatoid arthritis and their clinical significance. *Arthritis Res* , **4**, S1–S5.
- Stites WE. (1997). Protein-Protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *Chem Rev* , **97**, 1233-1250.
- Szakonyi G, Guthridge JM, Li D, Youn, K, Holers VM and Chen XS. (2001). Structure of complement receptor 2 in complex with its C3d ligand. *Science* , **292**, 1725-1728.
- Szalai AJ, Agrawal A, Greenhough TJ and Volanakis JE. (1997). C-reactive protein: structural biology, gene expression, and host defense. *Immunol Res* , **16**, 127–136.
- Szalai AJ, Digerness SB, Agrawal A, Kearney JF, Bucy RP, Niwas S, Kilpatrick JM, Babu YS and Volanakis JE. (2000). The Arthus reaction in rodents: species-specific requirement of complement. *J Immunol* , **164** (1), 463-8.
- Takeuchi K, Ng E, Malia TJ and Wagner G. (2007). 1-13C amino acid selective labeling in a 2H15N background for NMR studies of large proteins. *J Biomol NMR* , **38** (1), 89-98.
- Taylor PR, Carugati A, Fadok VA, Cook HT, Andrews M, Carroll MC, Savill JS, Henson PM, Botto M and Walport MJ. (2000). A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells in vivo. *J Exp Med* , **192** (3), 359-66.
- Tellinghuisen J. (2005). Optimizing Experimental Parameters in Isothermal Titration Calorimetry. *J Phys Chem B* , **109**, 20027–20035.

ten Wolde S, Breedveld FC, Hermans J, Vandenbroucke JP, van de Laar MA, Markusse HM, Janssen M, van den Brink HR and Dijkmans BA. (1996). Randomised placebo-controlled study of stopping second-line drugs in rheumatoid arthritis. *Lancet* , **347** (8998), 347-52.

Toapanta FR and Ross TM. (2006). Complement-mediated activation of the adaptive immune responses: role of C3d in linking the innate and adaptive immunity. *Immunol Res* , **36** (1-3), 197-210.

Tugarinov V and Kay LE. (2004). An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR* , **28** (2), 165-172.

Tugarinov V and Kay LE. (2003). Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *JACS* , **125** (45), 13868-78.

Tugarinov V, Choy WY, Orekhov VY and Kay LE. (2005). Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc Natl Acad Sci U S A* , **102** (3), 622-7.

Turnbull WB and Daranas AH. (2003). On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry. *J Am Chem Soc* , **125**, 14859–14866.

Turnbull WB. (2011). Divided we fall? Studying low-affinity fragments of ligands by ITC. *GE Healthcare Life Sciences (white paper)* www.gelifesciences.com/microcal , 1-12.

Upadhyay A, Burman JD, Clark EA, Leung E, Isenman DE, van den Elsen JM and Bagby S. (2008). Structure-function analysis of the C3 binding region of Staphylococcus aureus immune subversion protein Sbi. *J Biol Chem* , **283**, 22113–22120.

US Food and Drug Administration. (2007). FDA approves first-of-its-kind drug to treat rare blood disorder. (*USFDA, Rockville, MD, USA*) , <http://www.fda.gov/bbs/topics/NEWS/2007/NEW01589.html>.

van den Elsen JMH and Isenman DE. (2011). A Crystal Structure of the Complex Between Human Complement Receptor 2 and Its Ligand C3d. *Science* , **332** (6029), 608-611.

Van der Merwe PA. (2001). Surface Plasmon Resonance. In *Protein-Ligand interactions: hydrodynamics and calorimetry* (pp. 137-170). Oxford University Press.

Velazquez-Campoy A, Ohtaka H, Nezami A, Muzammil S and Freire E. (2004). Isothermal titration calorimetry. *Curr Protoc Cell Biol* , Chapter 17, Unit 7.

- Velyvis A, Yang YR, Schachman HK and Kay LE. (2007). A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proc Natl Acad Sci USA* , **104** (21), 8815-20.
- Venkiteshwaran A. (2009). Tocilizumab. *mAbs* , **1** (5), 432 - 438. <http://dx.doi.org/10.4161/mabs.1.5.9497>.
- Verbrugh HA, Van Dijk WC, Peters R, Van Der Tol ME and Verhoef J. (1979). The role of *Staphylococcus aureus* cell-wall peptidoglycan, teichoic acid, and protein A in the processes of complement activation and opsonization. *Immunology* , **31**, 615–621.
- Voehler MW, Collier G, Young JK, Stone MP and Germann MW. (2006). Performance of cryogenic probes as a function of ionic strength and sample tube geometry. *J. Magn. Reson.* , **183**, 102–109.
- Volanakis JE and Frank M. (1998). *The Human Complement System in Health and Disease*. New York: Marcel Dekker Inc.
- Volk J, Herrmann T and Wüthrich K. (2008). Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR* , **41** (3), 127-138.
- Wagenaar-Bos IG and Hack CE. (2006). Structure and function of C1-inhibitor. *Immunol Allergy Clin North Am* , **26**, 615–632.
- Wagner G. (1993). Prospects for NMR of large proteins. *J Biomol NMR* , **3**, 375-385.
- Walport MJ. (2001). Complement. Second of two parts. *N Engl J Med* , **344**, 1140–1144.
- Walters W and Namchuk M. (2003). Designing screens: how to make your hits a hit. *Nat Rev Drug Discov* , **4**, 259-266.
- Wang Y, Rollins SA, Madri JA and Matis LA. (1995). Anti-C5 monoclonal antibody therapy prevents collagen-induced arthritis and ameliorates established disease. *Proc Natl Acad Sci USA* , **92**, 8955–8959.
- Waterborg JH. (2009). The Lowry Method of Protein Quantiation. In *The Protein Protocols Handbood, 3rd edition, Chapter 2* (p. 7). Humana Press.
- Weber PC and Salemme FR. (2003). Applications of calorimetric methods to drug discovery and the study of protein interactions. *Curr Opin Struct Biol* , **13**, 115–121.
- Weiler JM, Daha MR, Austen KF and Fearon DT. (1976). Control of the amplification convertase of complement by the plasma protein beta1H. *Proc Natl Acad Sci USA* , **73**, 3268–3272.

Weinblatt ME, Coblyn JS, Fox DA, Fraser PA, Holdsworth DE, Glass DN, and Trentham DE. (1985). Efficacy of low-dose methotrexate in rheumatoid arthritis. *N Engl J Med* , **312**, 818-822.

Weinblatt ME, Keystone EC, Furst DE, Moreland LW, Weisman MH, Birbara CA, Teoh LA, Fischkoff SA and Chartash EK. (2003). Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. *Arthritis Rheum* , **48** (1), 35-45.

Weinblatt ME, Kremer JM, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Fox RI, Jackson CG, Lange M and Burge DJ. (1999). A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. *N Engl J Med* , **340** (4), 253-9.

Weis JJ, Toothaker LE, Smith JA, Weis JH, Fearon DT. (1988). Structure of the human B lymphocyte receptor for C3d and the epstein-barr virus and relatedness to other members of the family of C3/C4 binding proteins. *J Exp Med* , **167** (3), 1047-66.

Wells JA. (1990). Additivity of mutational effects in proteins. *Biochemistry* , **29**, 8509-8517.

Wells JA and McClendon CL. (2007). Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* , **450** (7172), 1001-9.

Whaley K and Ruddy S. (1976). Modulation of C3b hemolytic activity by a plasma protein distinct from C3b inactivator. *Science* , **193**, 1011–1013.

White J, Lukacik P, Esser D, Steward M, Giddings N, Bright JR, Fritchley SJ, Morgan BP, Lea SM , Smith GP. (2004). Biological activity, membrane-targeting modification, and crystallization of soluble human decay accelerating factor expressed in E. coli. *Protein Sci* , **13**, 2406-2415.

Wider G. (1998). Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. *Prog Nucl Magn Reson Spectros* , **32**, 193-275.

Wider G, Wüthrich K. (1999). NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr Opin Struct Biol* , **9** (5), 594-601.

Wilkinson BJ, Kim Y, Peterson PK, Quie PG and Michael AF. (1978). Activation of complement by cell surface components of *Staphylococcus aureus*. *Infect Immun* , **20**, 388–392.

Wiseman T, Williston S, Brandts JF and Lin LN. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* , **179**, 131-137.

Wolfe FE and Hawley DJ. (1998). The longterm outcome of rheumatoid arthritis. Work disability: a prospective 18 year study of 816 patients. *J Rheumatol* , **25**, 2108–2117.

Wüthrich K. (1998). The second decade – into the third millennium. *Nat Struct Biol* , **5**, 492-495.

Yelin E and Wanke LA. (1999). An assessment of the annual and long-term direct costs: the impact of poor function and functional decline. *Arthritis Rheum* , **42**, 1209–1218.

Zhang L, Mallik B, Morikis D. (2007). Immunophysical exploration of C3d-CR2(CCP1-2) interaction using molecular dynamics and electrostatics. *J Mol Biol* , **369** (2), 567-83.

Zhang YL and Zhang ZY. (1998). Low-Affinity Binding Determined by Titration Calorimetry Using a High-Affinity Coupling Ligand A Thermodynamic Study of Ligand Binding to Protein Tyrosine Phosphatase 1B. *Anal Biochem* , **261**, 139–148.

Zuiderweg ER. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* , **41**, 1-7.

APPENDIX I

Primers used for site-directed mutagenesis

Primers for C3d mutants

D36A forward	5'-gctgtgcattacctggctgaaacggagcagtgg-3'
D36A reverse	5'-ccactgtccgtttcagccaggtaatgcacagc-3'
E42A forward	5'-aacggagcagtgaggcgaagttcggcctag-3'
E42A reverse	5'-ctaggccgaacttcgccactgctccgtt-3'
E117A forward	5'-atggctgatcctggcgaagcagaagcccg-3'
E117A reverse	5'-cgggcttctgcttcgccaggatcagccat-3'
E166A forward	5'-ggctaaagatatttgcgaggagcaggtcaacagcc-3'
E167A reverse	5'-ggctgttgacgtgctccgcgcaaatatctttagcc-3'
E167A forward	5'-agatatttgcgaggcacaggtcaacagcc-3'
E167A reverse	5'-ggctgttgacgtgctccgcgcaaatatct-3'
D292A forward	5'-ggctcaatacaaaaggccgcccctgaccac-3'
D292A reverse	5'-gtggtcaggggcgcccttttgattagcc-3'
K78A/R79A forward	5'-ctttgcgcttctggtgagcgccacccagcacctgg-3'
K78A/R79A reverse	5'-ccaggtgctgggtgctgctgacgaaggccgcaaag-3'
D163A/E166A forward	5'-gctaaagctatttgcgaggagcaggtcaacagc-3'
D163A/E166A reverse	5'-gctgttgacgtgctccgcgcaaatatctttagc-3'
E37A/E39A forward	5'-gtgcattacctggatgcaacggcgagtgaggagaagttc-3'
E37A/E39A reverse	5'-gaacttctccactgcgcttgcatccaggtaatgcac-3'

Primers for CD21 SCR1-2 mutants

Y16A forward	5'-tcctaaatggccgattagtgttattctacccccattgctg-3'
Y16A reverse	5'-cagcaatggggtagaataagcactaatccggccatttagga-3'
R28A forward	5'-cattgctgttggtaccgtgatagcgtacagttgttcaggt-3'
R28A reverse	5'-acctgaacaactgtacgtatcacggtaccaacagcaatg-3'
K41A forward	5'-ggtaccttcgcctcattggagaagcaagtcattatgcataact-3'
K41A reverse	5'-agttatgcataatagacttgcttctccaatgaggcggaaggtacc-3'
K48A forward	5'-ggccttcgtggcagcgccacccagcacct-3'
K48A reverse	5'-aggtgctgggtgcccgtgccacgaaggcc-3'
K57A forward	5'-aaagtggatggaacctgggatgcacctgctcctaaatgtga-3'
K57A reverse	5'-tcacatttaggagcaggtgcatcccaggttccatccacttt-3'
R83A forward	5'-agcccatagtagcaggaggatacaaaattgagggtctacaccctac-3'
R83A reverse	5'-gtagggtgtagagccctcaattttgtatcctcctggtactatgggct-3'
R83E forward	5'-cccatagtagcaggaggatacaaaattgaaggctctacaccct-3'
R83E reverse	5'-agggtgtagagccttcaattttgtatcctcctggtactatggg-3'
R122E forward	5'-aataatatgtggggccgacagaactaccaacctgtgaag-3'
R122E reverse	5'-cttacacaggttggtagttctgtcgccccacatattatt-3'
R83E/G84Y forward	5'-catagtagcaggaggatacaaaattgaatactctacaccctacag-3'
R83E/G84Y reverse	5'-ctgtagggtgtagagtattcaattttgtatcctcctggtactatg-3'

APPENDIX II

Protein construct of wild type C3d

MLDAERLKHL IVTPSGAGEQ NMIGMTPTVI AVHYLDETEQ WEKFGLEKRQ GALELIKKGY
TQQLAFRQPS SAFAAFVKRA PSTWLTAYVV KVFS LAVNLI AIDSQVLCGA VKWLILEKQK
PDGVFQEDAP VIHQEMIGGL RNNNEKDMAL TAFVLISLQE AKDICEEQVN SLPGSITKAG
DFLEANYMNL QRSYTVAIAG YALAQMGR LK GPLLNKFLT AKDKNRWEDP GKQLYNVEAT
SYALLALLQL KDFDFVPPVV RWLNEQRYYG GYGSTQATF MVFQALAQYQ KDAPDHQELN
LDVSLQLPSR

Number of amino acids: 310

Molecular weight: 34695.0

Theoretical pI: 5.89

Amino acid composition:

Ala (A)	31	10.0%
Arg (R)	11	3.5%
Asn (N)	13	4.2%
Asp (D)	15	4.8%
Cys (C)	2	0.6%
Gln (Q)	22	7.1%
Glu (E)	19	6.1%
Gly (G)	20	6.5%
His (H)	4	1.3%
Ile (I)	13	4.2%
Leu (L)	38	12.3%
Lys (K)	20	6.5%
Met (M)	8	2.6%
Phe (F)	13	4.2%
Pro (P)	13	4.2%
Ser (S)	14	4.5%
Thr (T)	15	4.8%
Trp (W)	5	1.6%
Tyr (Y)	12	3.9%
Val (V)	21	6.8%

Total number of negatively charged residues (Asp + Glu): 34

Total number of positively charged residues (Arg + Lys): 31

Protein construct of wild type CD21 SCR1-2 with 6xHis-tag

MSCGSPPPII NGRISYYSTP IAVGTVIRYS CSGTFRLIGE KSLLCITKDK VDGTDWKPAP KCEYFNKYSS
 CPEIIVPGGY KIRGSTPYRH GDSVTFACKT NFSMNGNKS SV WCQANNMWGP TRLPTCVS ID
 GRVEHHHHHH HVD

Number of amino acids: 143

Molecular weight: 15902.1

Theoretical pI: 8.94

Amino acid composition:

Ala (A)	4	2.8%
Arg (R)	7	4.9%
Asn (N)	7	4.9%
Asp (D)	6	4.2%
Cys (C)	8	5.6%
Gln (Q)	1	0.7%
Glu (E)	4	2.8%
Gly (G)	13	9.1%
His (H)	8	5.6%
Ile (I)	9	6.3%
Leu (L)	5	3.5%
Lys (K)	9	6.3%
Met (M)	3	2.1%
Phe (F)	4	2.8%
Pro (P)	12	8.4%
Ser (S)	14	9.8%
Thr (T)	10	7.0%
Trp (W)	3	2.1%
Tyr (Y)	7	4.9%
Val (V)	9	6.3%

Total number of negatively charged residues (Asp + Glu): 10

Total number of positively charged residues (Arg + Lys): 16

APPENDIX III

NMR assignment of C3d

Residue in C3d	Arbitrary spin system number	Chemical shifts (ppm)							
		HN	N	C α i	C α i-1	C β i	C β i-1	CO i	CO i-1
ALA197	2	8.30	121.61	53.28	54.73	14.76	25.40	178.02	178.61
	6	6.55	113.02	51.83	55.35	25.60	36.02	0.00	178.22
GLN126	7	8.09	119.44	50.86	54.87	29.25	38.54	177.44	174.43
LEU250	8	7.85	117.95	52.40	55.02	39.26	24.23	177.04	177.18
	9	7.54	118.66	55.47	54.65	29.13	23.92	0.00	177.78
	11	8.10	113.25	53.25	53.73	26.58	38.18	178.90	179.71
VAL196	13	8.04	114.53	54.76	65.08	25.40	66.51	178.49	174.85
THR220	14	7.04	117.05	61.76	60.82	65.71	66.51	0.00	175.36
MET281	15	8.45	110.90	52.60	59.37	24.51	36.05	179.88	177.30
GLY109	16	7.91	110.69	44.09	51.77	0.00	31.92	176.56	177.70
ILE 11	20	7.61	116.19	54.17	53.82	35.46	37.09	178.48	177.88
TYR 60	21	8.57	123.62	58.11	44.81	35.47	0.00	0.00	176.14
GLY174	24	8.26	105.89	43.43	63.55	0.00	26.90	176.59	180.02
	25	8.35	125.40	53.16	59.18	0.00	29.25	0.00	183.96
GLU135	26	10.47	121.80	54.36	54.98	23.03	25.68	177.04	175.63
	27	7.26	118.24	56.38	61.51	37.51	28.09	177.83	173.14
TYR268	28	7.74	124.80	50.36	52.07	36.15	29.41	172.84	175.34
GLY270	29	9.13	112.35	42.23	51.23	0.00	32.42	174.24	176.72
ASN300	30	8.21	120.03	50.19	51.82	35.50	38.42	178.93	176.77
GLY 51	31	7.70	106.12	43.64	55.94	0.00	24.28	176.63	178.98
LYS 48	32	7.87	119.76	54.45	55.15	28.95	24.24	179.38	178.04
	34	8.71	118.82	54.15	55.29	36.28	36.66	0.00	178.49
ILE137	35	6.06	110.92	56.47	53.27	32.21	29.83	0.00	173.71
	36	8.32	110.94	56.77	51.51	69.05	33.39	0.00	0.00
	40	7.85	120.46	41.16	51.31	0.00	37.82	0.00	0.00
	41	7.85	106.79	41.16	50.05	36.07	36.13	172.73	176.36
	42	6.56	114.21	63.28	0.00	65.34	0.00	0.00	0.00
PHE253	44	7.63	119.40	57.14	48.80	34.85	35.10	0.00	178.43
GLY271	45	7.60	106.37	41.63	42.16	0.00	0.00	174.50	174.92
SER309	46	8.26	116.98	55.26	59.98	60.61	28.63	0.00	0.00
ARG310	47	7.84	128.12	54.32	55.25	27.76	60.60	180.82	173.75
LYS232	50	7.55	120.95	52.86	41.79	28.06	0.00	178.03	174.33
	52	9.55	120.94	55.48	0.00	36.06	0.00	0.00	0.00
LYS178	53	7.77	120.21	56.01	63.97	28.95	65.33	179.43	175.68
ARG 67	55	7.09	123.01	52.41	54.26	27.29	35.45	177.99	174.70
	56	8.41	117.69	54.81	56.20	38.11	0.00	0.00	0.00
GLU 54	57	7.74	118.48	56.10	55.12	25.41	37.54	178.45	180.35
	60	7.85	114.52	52.13	59.48	61.19	29.50	178.89	177.12

ASP302	61	8.21	121.54	51.29	52.15	37.81	38.70	175.66	174.84
VAL 12	65	8.15	123.30	56.28	54.09	28.68	35.44	177.95	176.38
	67	8.38	111.78	42.40	58.25	0.00	66.83	174.29	177.08
ASN225	68	7.92	113.10	49.81	54.59	37.19	27.17	0.00	174.50
GLU 19	70	7.08	114.68	54.25	44.65	27.47	0.00	178.46	174.70
GLU166	71	8.35	116.37	55.23	57.75	25.84	25.39	177.84	177.19
GLN 50	72	7.71	117.92	55.95	57.26	24.28	24.81	178.44	177.49
LEU 53	73	8.32	118.31	55.15	52.09	37.45	14.51	0.00	179.82
LEU305	74	7.93	123.82	52.03	55.85	38.42	60.33	0.00	0.00
ALA 74	75	7.27	118.73	46.95	54.67	22.73	40.53	175.65	177.94
	77	8.27	120.31	51.54	51.59	37.93	38.11	0.00	178.46
ASP223	78	8.66	120.25	52.37	51.77	36.06	29.20	175.86	175.86
	79	7.27	119.36	0.00	55.28	0.00	25.10	178.94	178.07
	81	8.91	122.88	0.00	55.99	0.00	27.24	0.00	175.02
VAL169	82	7.80	121.86	57.55	53.45	0.00	27.44	175.17	176.03
MET188	83	8.52	110.84	52.51	59.74	25.40	36.27	178.44	176.91
THR219	84	8.21	109.25	60.87	54.64	66.52	38.43	175.62	181.05
ALA101	85	7.59	121.31	48.62	55.52	14.15	39.01	176.59	173.30
	87	8.16	124.72	49.70	52.51	38.40	25.69	0.00	176.38
GLY272	88	8.05	106.66	40.38	41.64	0.00	0.00	175.64	174.41
ASP128	90	8.65	127.23	0.00	51.71	39.63	28.19	0.00	174.49
ALA239	91	9.22	120.84	52.63	56.85	16.81	27.16	179.44	177.31
LYS 78	92	7.45	115.35	52.64	62.38	28.58	28.06	176.33	176.91
VAL131	94	8.19	114.06	58.28	59.10	28.35	29.24	177.98	176.57
ARG208	95	7.59	114.47	52.08	43.47	29.26	0.00	117.07	175.55
ILE176	97	7.93	121.45	61.94	58.44	34.64	60.32	0.00	178.44
ALA185	98	7.65	118.76	51.03	56.27	15.04	26.58	178.92	176.40
GLY207	99	7.57	108.95	43.45	51.55	0.00	30.76	175.63	176.25
PHE 73	100	7.56	112.20	54.68	46.88	40.50	15.39	0.00	174.51
	101	7.56	123.58	56.67	55.74	36.55	60.03	173.78	176.37
GLN105	102	8.83	122.01	56.01	58.95	24.50	25.68	180.79	176.40
	103	9.31	121.61	0.00	59.12	35.62	0.00	176.58	176.86
VAL106	105	7.33	120.14	61.87	56.02	28.35	24.51	117.53	178.95
	106	8.30	130.02	52.66	52.45	22.00	27.21	172.26	175.25
	110	8.71	122.28	54.37	44.28	36.35	0.00	179.44	174.94
ASP163	112	8.46	118.02	54.02	58.00	36.60	28.94	178.92	177.10
	113	8.21	123.54	51.90	50.37	38.55	35.73	177.07	175.03
ASP252	114	7.46	119.25	48.77	53.50	35.06	25.35	177.03	178.85
VAL 90	115	8.81	119.37	57.08	58.84	25.09	26.29	172.36	178.61
LYS251	117	7.49	114.28	53.51	52.39	0.00	39.11	176.58	0.00
	118	7.92	117.62	52.54	55.45	38.72	29.06	175.88	177.15
	120	7.93	128.41	54.35	55.61	27.95	60.60	180.95	172.14
	121	7.65	126.94	50.96	55.20	38.16	35.14	0.00	175.87
CYS108	123	9.10	114.06	51.75	55.77	31.88	40.18	178.05	177.60
	124	8.52	117.36	55.63	59.46	60.71	31.03	117.07	180.50

	126	8.07	115.87	58.81	50.70	66.51	40.48	0.00	0.00
GLN 68	127	8.34	129.97	0.00	52.41	27.37	27.19	181.27	175.30
LEU307	128	8.12	125.36	49.93	52.25	37.84	25.70	176.12	176.57
PHE 66	131	7.71	116.19	54.26	50.68	35.44	14.16	0.00	178.07
	133	7.82	123.52	49.30	52.57	13.57	38.77	0.00	177.17
	135	7.99	117.79	55.89	51.35	26.71	13.85	177.38	181.88
	136	8.02	123.43	0.00	53.01	38.41	15.63	0.00	0.00
GLY231	137	8.33	112.28	41.79	60.63	0.00	27.76	174.24	177.07
ALA 75	138	9.62	120.36	51.97	47.02	15.91	22.74	174.72	177.51
SER304	140	8.26	118.94	55.88	59.38	60.30	28.65	173.81	178.38
HIS133	141	10.33	122.24	52.39	61.83	22.47	34.85	176.59	180.71
ALA221	143	6.07	123.70	48.80	61.76	12.55	65.90	177.98	173.99
	145	7.63	112.55	50.02	62.24	36.09	28.35	177.11	176.55
	146	8.38	123.59	52.10	59.99	39.00	28.36	0.00	175.91
	147	8.03	118.77	63.32	50.54	26.28	15.62	175.70	174.64
GLN306	148	8.02	121.06	52.29	52.01	25.68	38.42	180.14	176.53
	151	7.97	119.45	0.00	49.34	0.00	15.81	0.00	180.82
VAL 32	152	7.60	116.78	63.95	51.88	27.18	14.65	177.51	177.87
	154	8.29	117.34	55.30	53.12	0.00	29.32	0.00	0.00
GLU 39	155	7.39	119.99	53.86	58.41	23.45	65.97	177.52	172.95
LYS224	158	7.60	108.90	54.61	52.37	27.23	36.06	174.67	173.22
	159	8.85	124.10	55.14	50.38	35.75	40.78	0.00	0.00
GLY138	160	7.35	114.39	42.95	56.45	0.00	32.20	176.09	176.58
GLN249	161	7.85	122.53	55.09	54.51	24.27	36.64	178.46	180.92
ARG226	162	6.50	108.82	52.03	49.84	24.53	37.19	172.83	172.59
GLN 62	163	8.51	120.64	55.61	63.55	24.52	65.62	179.41	178.79
	164	9.03	129.01	54.73	49.75	36.60	0.00	0.00	0.00
THR177	165	7.93	115.97	63.98	61.93	65.34	34.56	176.59	178.27
LEU301	167	8.10	122.98	52.16	50.22	38.69	35.47	0.00	0.00
SER 82	169	7.80	114.46	52.15	59.49	61.23	29.56	176.58	0.00
ILE164	170	7.44	118.61	60.31	54.01	35.15	36.65	178.46	178.96
	172	6.74	112.85	51.53	55.89	30.99	25.40	0.00	176.68
	176	8.29	123.23	51.41	41.67	27.01	0.00	0.00	0.00
ASN186	180	6.73	112.47	50.32	51.01	38.69	15.03	0.00	178.91
LYS 58	182	8.07	121.54	56.50	57.30	38.01	28.77	178.91	178.95
	183	8.30	125.05	49.34	59.99	0.00	28.64	175.72	174.72
SER 70	185	6.86	109.77	55.15	62.52	60.80	27.73	174.71	176.55
LEU190	186	7.07	119.91	52.62	49.62	40.19	35.74	177.50	176.21
GLY 45	187	8.54	108.07	42.92	56.70	0.00	38.03	174.71	175.72
ASP254	188	8.64	119.70	53.79	57.13	37.38	34.82	179.40	177.69
LEU107	189	7.36	120.67	55.82	61.87	40.28	28.37	177.62	178.96
	192	8.42	118.20	56.17	43.92	30.70	0.00	176.60	174.46
ALA 52	194	7.67	125.34	52.05	43.65	14.46	0.00	179.87	176.73
	195	8.51	122.70	53.84	52.69	0.00	29.84	0.00	178.83
VAL303	196	7.93	120.66	59.40	51.31	28.67	37.78	0.00	176.75

	197	8.53	121.32	52.91	54.62	15.04	0.00	172.73	178.76
CYS165	198	7.82	112.40	51.28	60.32	36.36	35.18	176.98	178.79
THR 61	199	8.47	113.75	63.55	58.11	65.64	35.47	178.93	175.52
GLY139	200	9.42	113.89	41.76	43.00	0.00	0.00	0.00	175.82
TYR187	201	7.49	124.57	59.76	50.33	36.31	38.71	177.06	175.91
THR 83	202	11.72	135.53	63.44	52.15	65.94	61.22	0.00	178.54
VAL 77	204	9.02	119.44	62.35	50.49	28.04	39.02	177.04	177.76
ALA110	205	7.15	125.36	52.17	44.13	14.46	0.00	178.90	176.48
GLY 59	207	8.32	106.71	44.82	56.49	0.00	28.95	176.14	178.63
LYS 91	210	7.49	117.81	54.75	57.07	29.10	25.10	177.96	179.12
ALA 17	211	7.71	121.53	48.33	40.94	12.98	0.00	176.60	173.65
GLY 16	212	7.93	104.13	40.93	54.86	0.00	62.37	173.77	172.02
ASP103	213	7.65	126.45	50.89	55.24	38.10	35.15	0.00	176.69
	218	6.83	109.69	0.00	62.53	0.00	27.75	0.00	176.56
ALA129	219	8.67	124.93	47.12	51.73	15.91	39.60	172.38	176.55
ALA72	220	6.52	120.83	46.86	59.66	15.34	58.26	174.72	171.26
SER 15	221	8.78	118.99	54.85	59.88	62.38	28.06	178.93	176.90
	222	8.55	119.51	0.00	51.48	0.00	32.78	0.00	0.00
	223	7.87	123.56	49.41	52.41	39.46	25.96	0.00	0.00
ILE 100	224	7.62	111.32	55.54	52.88	39.01	40.39	173.30	176.54
LYS 57	225	8.41	121.23	57.29	62.99	28.65	33.97	178.96	177.93
	227	8.35	121.05	51.53	41.27	26.31	0.00	0.00	0.00
ILE132	229	10.82	129.30	61.84	58.28	34.95	28.32	180.81	177.72
	230	7.65	123.57	56.79	55.70	36.64	59.83	174.23	179.36
	232	7.02	116.14	52.35	54.65	23.31	25.52	176.11	177.38
	233	8.38	116.29	43.28	56.82	0.00	0.00	175.65	178.31
	234	8.52	120.18	57.95	52.83	24.49	28.92	0.00	0.00
	236	7.82	117.34	51.04	60.98	37.82	28.07	177.05	177.23
	237	9.07	125.63	58.87	50.67	0.00	37.81	0.00	0.00
VAL256	239	7.19	115.57	64.60	54.86	28.06	38.08	0.00	177.00
	240	8.10	120.43	0.00	51.51	0.00	38.01	178.97	174.41
	241	8.32	122.32	53.37	52.19	0.00	25.70	0.00	0.00
ARG 49	244	7.13	122.91	57.29	54.42	24.82	28.99	177.94	178.56
GLU 47	245	9.52	119.75	55.15	55.49	24.22	38.42	178.00	179.71
	246	8.61	116.00	60.70	46.98	0.00	16.52	0.00	0.00
	247	8.37	127.40	52.07	52.63	27.75	40.85	0.00	172.40
MET136	249	7.96	121.41	53.22	54.40	29.54	23.08	178.05	176.72
PHE255	250	7.60	118.67	54.84	53.78	37.84	37.26	0.00	178.41
	251	7.89	126.33	52.89	41.98	38.95	0.00	0.00	176.43
	254	7.19	117.97	56.84	52.29	37.58	28.04	0.00	0.00
	255	7.88	118.28	52.30	51.08	0.00	37.82	0.00	178.43
SER175	256	7.70	119.11	58.46	43.45	60.31	0.00	0.00	178.30
GLU166	257	8.62	124.70	57.75	51.30	25.38	36.27	172.37	176.54
	259	8.18	119.07	53.02	51.59	25.77	37.82	178.93	177.05
	260	8.27	120.69	0.00	53.06	0.00	25.76	0.00	0.00

SER 71	261	7.98	108.75	59.62	55.14	58.19	60.62	0.00	174.79
	262	7.23	119.53	61.26	63.03	28.71	28.04	0.00	179.79
LEU 64	264	7.54	116.50	54.15	54.25	36.44	23.62	180.77	179.66
	265	8.29	105.46	44.11	60.67	0.00	23.92	0.00	180.02
GLY 18	266	7.70	103.26	44.67	48.42	0.00	12.98	174.71	176.56
TRP 84	268	8.97	122.78	61.00	63.46	26.69	65.94	178.90	178.38
ALA 65	269	7.03	120.94	50.68	54.18	14.14	36.38	177.97	179.57
	270	6.55	116.15	56.49	51.51	30.65	0.00	175.23	174.49
	271	8.46	121.86	54.31	50.10	28.02	38.15	0.00	175.35
LYS162	272	7.25	122.24	58.01	48.87	28.95	16.18	177.46	178.98
LYS120	273	8.82	128.77	51.65	50.96	27.17	25.70	177.13	176.31
GLN168	275	7.51	115.16	53.45	55.21	27.46	25.86	176.13	177.42
	276	8.02	121.89	51.92	53.87	29.23	15.64	0.00	176.57
	277	8.67	127.39	52.62	58.80	35.46	66.23	0.00	174.48
	278	8.12	111.68	41.67	55.45	0.00	61.46	0.00	0.00
	280	9.21	117.59	54.81	52.92	36.95	15.32	0.00	179.09
TYR269	281	6.85	122.45	51.22	50.37	32.94	36.06	176.63	172.60
LEU 55	282	7.44	120.59	54.70	56.09	38.12	25.40	0.00	180.16
	283	7.34	121.56	54.41	0.00	0.00	0.00	177.32	181.71
HIS 33	284	8.07	117.40	56.80	63.90	26.98	27.27	117.01	177.40
ASN189	287	7.30	115.15	49.59	52.52	35.75	25.42	176.13	178.30
	288	8.03	108.44	43.94	0.00	0.00	0.00	0.00	0.00
	289	7.31	125.64	53.75	56.25	26.86	61.34	0.00	0.00
	290	8.74	125.15	58.40	50.31	35.16	32.50	0.00	174.53
GLN119	291	6.90	120.18	50.93	50.40	25.69	28.05	176.37	174.80
	292	7.94	108.71	59.79	55.13	58.24	60.84	0.00	174.75
	293	7.78	120.96	0.00	52.31	0.00	38.21	175.16	175.89
TYR273	295	8.68	111.85	0.00	40.34	40.37	0.00	178.27	0.00
LEU 46	296	9.02	131.48	55.50	42.91	38.43	0.00	179.89	174.62
LYS222	298	8.42	126.10	51.79	48.83	29.07	12.57	0.00	177.82
SER104	300	9.25	126.72	58.88	50.99	0.00	38.14	0.00	178.14
ARG 79	302	7.53	123.80	51.86	52.65	27.92	28.35	177.01	175.36
	303	8.32	120.26	53.48	49.76	26.57	15.81	0.00	178.45
GLN 63	311	8.15	120.59	54.23	55.61	23.60	24.47	180.16	179.24
	312	8.21	126.16	49.92	52.24	37.82	29.30	181.29	173.11
GLU127	313	9.17	126.96	51.71	50.88	28.35	29.31	174.66	174.67
ALA 31	316	8.46	126.23	51.90	57.08	14.75	35.18	179.81	177.24
	317	7.45	121.82	51.50	59.88	33.56	66.53	0.00	0.00
ILE 102	319	8.10	126.15	55.24	48.65	35.14	14.15	176.06	176.56
	322	8.16	126.35	49.29	59.24	15.64	29.23	0.00	176.90
	323	8.24	125.99	51.88	51.39	0.00	0.00	178.88	176.92
	324	8.31	120.49	50.69	55.36	31.28	28.06	179.18	173.46
GLN134	325	7.73	123.68	55.00	52.38	25.69	22.44	175.63	176.37
	329	7.62	126.97	50.96	55.20	30.11	38.37	0.00	182.82
ALA 80	334	8.21	129.42	47.75	51.86	14.15	28.05	175.14	173.78

	338	8.72	125.10	49.18	50.67	40.26	31.31	0.00	0.00
	342	7.94	123.84	0.00	42.16	0.00	0.00	0.00	0.00
	344	8.82	122.25	53.88	53.40	0.00	0.00	0.00	176.38
	345	8.49	122.44	49.41	57.40	66.22	32.20	172.81	175.33
	346	8.43	121.21	57.04	52.46	25.57	28.34	0.00	0.00
	353	8.27	109.57	42.24	52.52	0.00	38.42	0.00	0.00
	355	8.13	108.33	42.14	49.71	0.00	15.64	0.00	0.00
	356	8.16	109.25	42.22	52.51	0.00	38.42	0.00	0.00
	373	8.31	122.59	0.00	55.37	0.00	60.88	0.00	0.00
	374	8.37	122.67	53.50	52.13	37.82	38.48	175.12	174.86
	377	8.25	123.94	59.29	52.63	29.21	26.27	0.00	0.00
GLU238	378	7.45	123.33	56.88	64.22	26.91	28.26	177.52	178.39
	381	8.27	119.66	0.00	59.34	0.00	28.74	178.48	173.39
PHE280	382	8.06	119.34	59.37	64.86	36.01	66.02	177.99	174.73
	383	7.62	119.05	0.00	43.46	0.00	0.00	0.00	177.60
	386	8.08	120.51	0.00	53.75	0.00	26.56	0.00	179.30
	388	8.05	120.81	51.43	42.23	37.94	0.00	0.00	179.06
	389	7.99	121.44	54.49	0.00	42.19	0.00	173.76	178.29
	390	8.22	122.62	52.97	53.48	29.40	26.58	0.00	0.00
	391	8.95	129.87	47.98	54.65	19.38	29.29	0.00	0.00
LEU 95	392	8.33	125.63	52.01	54.89	38.71	60.53	0.00	0.00
	393	8.15	125.64	54.68	54.18	15.63	27.14	0.00	0.00
	394	8.15	125.36	49.39	59.21	38.48	29.53	0.00	0.00
	395	7.79	126.29	52.82	51.87	39.27	38.70	0.00	176.50
	397	7.50	120.28	52.75	42.43	27.87	0.00	0.00	0.00
VAL124	398	8.72	120.05	58.10	41.36	31.33	0.00	0.00	0.00
	399	8.23	120.74	53.21	49.29	26.63	15.79	0.00	0.00
	401	8.27	118.80	58.24	55.12	34.56	24.50	176.11	179.79
	403	6.96	118.92	52.73	58.69	36.33	29.25	0.00	0.00
	404	8.33	117.80	55.25	57.03	37.47	0.00	0.00	0.00
	405	8.74	117.41	52.52	50.03	36.63	0.00	0.00	0.00
	406	9.23	116.03	57.87	52.13	68.00	39.00	0.00	0.00
	409	8.05	116.07	0.00	42.30	0.00	0.00	0.00	0.00
	410	7.99	115.34	54.52	40.89	0.00	0.00	0.00	0.00
	415	7.21	125.05	0.00	44.15	15.31	0.00	0.00	0.00
	416	8.75	126.11	48.03	51.61	0.00	0.00	0.00	0.00
	417	8.12	125.47	0.00	59.11	15.66	66.82	0.00	0.00
	418	8.01	117.90	63.34	40.40	26.21	0.00	174.70	176.21
	419	8.30	110.76	42.12	53.63	0.00	26.67	0.00	0.00
	420	7.27	119.05	0.00	49.83	0.00	14.44	175.67	178.63
	421	7.88	118.26	63.04	57.22	28.39	25.78	0.00	175.98
	422	8.02	121.90	55.97	49.14	38.42	26.23	0.00	178.31
	423	8.33	121.17	0.00	42.12	42.07	0.00	0.00	178.62
	424	7.83	120.60	52.78	54.20	29.02	26.30	180.00	178.37
	425	8.32	120.37	0.00	60.42	0.00	28.42	0.00	176.19

	426	8.33	122.59	56.60	59.11	66.49	29.22	0.00	174.85
	427	8.44	122.42	52.83	40.79	0.00	0.00	0.00	175.86
	428	8.24	122.26	0.00	52.08	0.00	38.70	0.00	177.28
ALA 96	429	8.13	121.39	51.32	52.01	0.00	38.70	0.00	0.00
	430	8.08	121.69	0.00	52.19	0.00	29.21	0.00	0.00
	431	8.10	122.58	0.00	53.56	0.00	0.00	0.00	0.00
	432	8.10	121.70	0.00	49.22	0.00	15.64	0.00	0.00
THR195	724	7.76	123.05	65.03	0.00	66.17	0.00	0.00	176.49
	726	8.35	121.10	52.75	60.12	0.00	28.50	0.00	0.00
ILE 30	727	7.82	118.41	57.14	0.00	0.00	0.00	177.07	175.96
VAL 89	738	8.19	120.05	0.00	53.47	0.00	0.00	0.00	0.00
	744	9.10	109.16	40.39	58.50	0.00	66.79	0.00	0.00
	746	9.00	109.60	42.75	51.02	0.00	0.00	0.00	0.00
PHE 76	748	6.17	102.56	50.49	51.87	39.02	15.86	177.94	174.66
	750	8.98	104.50	42.51	0.00	0.00	0.00	0.00	0.00
GLY123	752	7.33	108.01	41.40	49.17	0.00	36.59	0.00	0.00
	754	8.52	106.60	41.39	51.52	0.00	51.62	174.71	179.80
	756	8.65	112.26	64.86	53.28	0.00	0.00	0.00	178.26
THR 38	758	7.44	104.82	58.43	55.22	65.92	0.00	173.25	178.44
	760	7.52	128.84	54.86	48.43	0.00	15.95	0.00	0.00
	762	9.14	126.02	58.94	50.81	0.00	38.10	174.70	177.88
	764	8.92	126.15	54.15	51.72	0.00	30.45	0.00	0.00
	766	7.42	109.47	42.53	0.00	0.00	0.00	0.00	0.00
	767	7.04	106.76	55.88	53.50	0.00	0.00	0.00	0.00
	769	7.23	112.39	51.09	41.92	37.84	0.00	0.00	0.00
	771	8.41	113.05	54.55	57.05	0.00	68.78	0.00	0.00
	774	6.68	114.10	50.39	53.41	0.00	0.00	0.00	178.69
	778	7.27	119.43	48.80	54.67	16.16	25.10	178.93	179.73
	781	8.15	123.16	56.11	51.99	15.65	26.71	0.00	178.28
	785	9.55	120.56	55.88	0.00	36.04	0.00	178.44	178.79
	801	8.08	127.52	63.31	51.80	34.25	0.00	173.33	181.36
	812	8.03	108.43	43.92	0.00	0.00	0.00	174.67	179.59
	833	8.09	128.82	57.37	56.48	32.21	30.71	0.00	0.00
	835	7.68	124.30	52.02	51.39	15.33	36.62	0.00	173.94
	837	7.18	119.43	61.26	53.49	15.98	14.73	0.00	0.00
	841	7.55	114.24	50.53	53.77	0.00	0.00	176.11	177.35
	845	7.22	120.83	51.54	52.54	37.28	0.00		176.04