Mechanisms of Immunoglobulin Deactivation by *Streptococcus pyogenes*

Emma Victoria Dixon

A thesis presented for the degree of
Doctor of Philosophy



Linacre College
Department of Biochemistry
University of Oxford

Trinity Term 2014

To my Grandparents

Abstract

The bacteria *Streptococcus pyogenes* produces a multitude of proteins which interact with and alter the functions of the host immune system. Two such proteins, Endoglycosidase S (EndoS) and Immunoglobulin G-degrading enzyme from *S. pyogenes* (IdeS) are able to specifically alter the effector functions of immunoglobulin G (IgG). EndoS is a glycoside hydrolase which removes the conserved *N*-linked glycan from IgG Fc whereas IdeS is a cysteine protease that cleaves the flexible protein hinge of IgG. The activity of both proteins results in the reduced ability of IgG to elicit immune responses through Fc receptor binding and complement activation. Amongst other applications, both EndoS and IdeS are actively being explored as new therapeutics for IgG-mediated autoimmune diseases. Given the therapeutic potential of EndoS and IdeS, experiments were designed to investigate the structural and functional characteristics of these enzymes in an effort to understand their specificity for and activity against IgG.

Here, bioinformatic and biophysical characterisation of EndoS identified subdomains outside of the catalytic domain which contribute to glycoside hydrolase activity. The substrate specificity of EndoS was also explored and showed that EndoS hydrolyses a broad range of glycans from the IgG scaffold. EndoS was also shown to have activity against alternative glycoprotein substrates, however, this non-specific activity was negligible in the context of whole serum. The effect of EndoS-mediated deglycosylation on the structure of the IgG Fc domain was explored using both X-ray crystallography and small-angle X-ray scattering. Small angle X-ray scattering was also used to characterise both EndoS and IdeS in complex with IgG Fc. Solution-state models of each complex were produced providing preliminary data towards how these enzymes interact with IgG. Overall, the results presented here contribute to our understanding of these enzymes which is of importance as they go forward into clinical applications.

Acknowledgements

First and foremost I would like to thank Nicole Zitzmann for giving me this incredible opportunity. I feel so fortunate to have had such a supportive, generous and enthusiastic supervisor who has been a constant source of strength and guidance these past four years. A huge thanks also goes to both Max Crispin and Chris Scanlan for allowing me to work with them both on EndoS and IdeS. This DPhil seemed an impossible task not so long ago and I am forever grateful for the opportunity to take on these projects. Through this work I've gained incredible experience in the lab and have always felt encouraged to go with my gut and try new techniques. The completion of this thesis represents a great deal to me and a huge part of its success is thanks to them.

Thanks also goes to those who helped me during the first two years of my DPhil whilst working on the hepatitis C virus p7 ion channel. In particular to Phil Biggin and his lab and to Constantina Foutinou, Benjamin Oestringer and J.L. Kiappes for their support and encouragement as fellow members of team p7. Further thanks goes to members of the Structural Genomics Consortium, the Membrane Protein Laboratory and the Research Complex at Harwell for their time and effort towards this work.

I would like to thank Prof. Raymond Dwek for making the Glycobiology Institute what it is and for his encouragement throughout my time here. I would also like to thank all the members of the Zitzmann lab, past and present, who have made every day entertaining and have provided the most fantastic environment in which to work. Particular thanks goes to Pietro Roversi who has given invaluable guidance in all matters X-ray crystallography and who was integral to the analysis of data presented in Chapters 4, 6 and 7. I feel incredibly lucky to have had such a knowledgable and enthusiastic teacher to help me determine my

first X-ray crystal structures.

I have also been fortunate enough to work with the members of the Crispin lab and thank many of them for their contributions to this thesis. In particular to Ben Yu for his work on the cloning and expression of *Mm*IgGs and to Camille Bonomelli for IgG 2G12 production. Thanks also goes to Susan Tickner for her initial efforts on the IdeS project, to Kavitha and Laura help with glycan analysis and to Snezana for her constant wisdom throughout this project.

The chacterisation of the glycans in Chapter 5 would not have been possible without the work of Prof. David Harvey. I am very grateful for his contributions to both the collection and analysis of the mass spectrometry data. A big thanks also goes to Edward Lowe who was responsible for data collection from the SeMet CBM and *Mm*IgG crystals. I would also like to thank Jolyon Claridge for help with NMR experiments and Matt Higgins for generously allowing me to join in on his SAXS beamtime in Hamburg.

To all the friends I've made here in Oxford, I give thanks for all the silly conversations, the nights out, the sports and the antics which have kept me sane all this time. I also wish to thank my family, who have been so very supportive throughout what has been a very challenging and difficult experience at times. Despite being on the other side of the world, they have been such an integral part of my success here and I am so grateful for their love and support.

Lastly, to my fiancé Johnny. To have come to meet you as a result of this DPhil is the biggest gift of all. I can't wait for what's ahead.

Contents

List of Figures **x**

List of Tables **xi**

List of Abbreviations **xii**

1 Introduction **1**
1.1 *Streptococcus pyogenes* . . . 2
1.1.1 *Streptococcus pyogenes* invasion and infection . . . 4
1.1.2 The host response to *Streptococcus pyogenes* infection . . . 4
1.2 Immunoglobulins . . . 6
1.2.1 The structure of IgG . . . 9
1.2.2 The role of IgG in immunity . . . 13
1.2.3 IgG glycosylation . . . 17
1.3 Immune evasion strategies of *Streptococcus pyogenes* . . . 21
1.3.1 Inhibiting AMPs . . . 22
1.3.2 Altering complement activation . . . 22
1.3.3 Disrupting neutrophil recruitment and NETs . . . 24
1.3.4 Intracellular survival and phagocytosis evasion . . . 24
1.3.5 Ig targeting proteins . . . 25
1.3.6 EndoS . . . 25
1.3.7 IdeS . . . 28
1.4 Exploiting immune evasion factors for therapeutic use . . . 34
1.4.1 Autoimmune disease . . . 34
1.4.2 Organ transplantation . . . 36
1.4.3 Receptor refocusing . . . 36
1.4.4 Chemoenzymatic engineering . . . 37
1.5 Aims . . . 38

2 Materials and methods **39**
2.1 Bioinformatics . . . 39
2.2 Molecular cloning and mutagenesis . . . 40
2.2.1 PCR . . . 40
2.2.2 In-Fusion® cloning protocol . . . 41
2.2.3 EndoS and EndoS fragments . . . 42
2.2.4 IdeS . . . 42
2.2.5 Immunoglobulins . . . 44
2.2.6 Control proteins and enzymes . . . 45
2.2.7 Mutagenesis . . . 46
2.3 Protein expression . . . 46
2.3.1 EndoS and EndoS fragments . . . 46
2.3.2 IdeS . . . 49
2.3.3 MBP . . . 50

2.3.4 Immunoglobulins . . . 50
2.3.5 Glycosyltransferases . . . 51
2.3.6 *In vivo* glycoform modulation . . . 51
2.3.7 *In vitro* glycoform modulation . . . 51
2.4 Protein purification . . . 52
2.4.1 EndoS . . . 52
2.4.2 ChitLRR . . . 53
2.4.3 CBM-KO . . . 53
2.4.4 CBM, CBM-CT, CT and MBP . . . 54
2.4.5 IdeS . . . 54
2.4.6 Immunoglobulins . . . 55
2.4.7 Glycosyltransferases . . . 57
2.5 Enzyme substrate complex formation . . . 57
2.5.1 EndoS . . . 57
2.5.2 IdeS . . . 57
2.6 Biophysical methods . . . 58
2.6.1 Determining protein concentration . . . 58
2.6.2 Gel electrophoresis and western blot methods . . . 58
2.6.3 Circular dichroism . . . 60
2.6.4 Small-angle X-ray scattering . . . 61
2.6.5 Nuclear magnetic resonance . . . 64
2.6.6 Enzyme-linked immunosorbent assays . . . 64
2.6.7 Isothermal titration calorimetry . . . 65
2.7 Protein crystallisation . . . 65
2.7.1 CBM . . . 66
2.7.2 Deglycosylated *Mm*IgG1Fc . . . 67
2.7.3 Deglycosylated *Mm*IgG2aFc . . . 68
2.8 X-ray crystallography . . . 68
2.8.1 Data collection . . . 68
2.8.2 Data processing . . . 69
2.9 Glycobiology . . . 69
2.9.1 Enzymatic release of *N*-linked glycans . . . 69
2.9.2 Exoglycosidase digests of *N*-linked glycans . . . 70
2.9.3 Purification . . . 70
2.9.4 Fluorescent labelling . . . 71
2.9.5 Mass spectrometry . . . 71
2.9.6 High performance liquid chromatography . . . 72
2.9.7 Serum glycoproteins . . . 73

3 Structural characterisation of EndoS **75**
3.1 Bioinformatic analysis of EndoS . . . 76
3.1.1 RONN and PONDR . . . 76
3.1.2 FUGUE . . . 77
3.2 Production of EndoS and EndoS fragments . . . 80
3.2.1 Cloning . . . 80
3.2.2 Expression . . . 81
3.2.3 Purification . . . 82
3.3 Production of EndoS glycoprotein substrates . . . 92
3.4 Production of EndoS:substrate complexes . . . 94
3.4.1 EndoS catalytic mutants . . . 95
3.4.2 EndoS-Fc complex . . . 97
3.4.3 EndoS-2G12 complex . . . 97

3.5 Circular dichroism . . . 100
3.5.1 EndoS and EndoS catalytic mutants . . . 100
3.5.2 EndoS fragments . . . 101
3.6 Crystallisation . . . 102
3.7 SAXS . . . 104
3.7.1 EndoS . . . 105
3.7.2 EndoS fragments . . . 109
3.7.3 Models of EndoS in solution using the X-ray crystal structure of EndoS . . . 111
3.7.4 EndoS-Fc complex . . . 113
3.8 Discussion . . . 119

4 Biophysical and structural investigation of the EndoS carbohydrate-binding module 126
4.1 X-ray crystallography . . . 127
4.1.1 Crystallisation . . . 127
4.1.2 Diffraction data collection and processing . . . 129
4.1.3 Phasing . . . 132
4.1.4 Structure building and refinement . . . 137
4.1.5 Model of CBM . . . 138
4.2 Investigating the monosaccharide binding function of CBM . . . 144
4.2.1 Nuclear magnetic resonance . . . 144
4.2.2 Isothermal titration calorimetry . . . 145
4.2.3 X-ray crystallography . . . 148
4.3 Discussion . . . 153

5 Substrate specificity and functional characterisation of EndoS 156
5.1 Investigation of enzyme and substrate components required for activity . . . 157
5.1.1 EndoS activity against IgG fragments . . . 157
5.1.2 The role of the EndoS C-terminus in glycoside-hydrolase activity . . . 162
5.2 Investigation of IgG glycan specificity . . . 167
5.3 Investigation of alternative glycoprotein substrates . . . 170
5.3.1 EndoS activity against IgG Fab glycans . . . 170
5.3.2 Human immunodeficiency virus gp120 . . . 173
5.3.3 Serum glycoproteins . . . 175
5.3.4 EndoS activity against whole serum . . . 177
5.4 Discussion . . . 179

6 Structural effects of EndoS deglycosylation on IgG Fc 188
6.1 Assessing the effects of deglycosylation using SAXS . . . 189
6.2 Structural studies of deglycosylated murine IgG subclasses . . . 194
6.2.1 X-ray crystallography . . . 195
6.2.2 Models of *Mm*IgG1Fc and *Mm*IgG2aFc . . . 201
6.3 Discussion . . . 214

7 Structural characterisation of IdeS in complex with IgG1 Fc 219
7.1 Preparation of the IdeS-Fc complex . . . 220
7.1.1 Expression and purification of IdeS . . . 220
7.2 X-ray crystallography . . . 226
7.2.1 Crystallisation . . . 226
7.2.2 Diffraction data collection and processing . . . 227
7.3 SAXS . . . 229

7.3.1 Data collection and processing 229
7.3.2 Model of IdeS-Fc complex in solution 232
7.4 Discussion 234

8 Conclusions **238**

List of Figures

1.1 Scanning electron micrograph and cell diagram of *S. pyogenes* 2
1.2 Human Ig isotypes 6
1.3 Multiple alignment of the conserved domains of the IgG heavy chain 8
1.4 The structure of human IgG1 10
1.5 The crystal structure of IgG1 Fc 11
1.6 The structure and topology of the IgG Fc CH2 domain 12
1.7 Human Fc receptors 14
1.8 The interaction between Fc and FcγRs 16
1.9 The mammalian *N*-linked glycosylation pathway and typical IgG glycans . 18
1.10 The IgG1 CH2-glycan interface 19
1.11 Immune evasion strategies of GAS 21
1.12 EndoS hydrolysis of IgG glycans 26
1.13 The crystal structure of IdeS 29
1.14 Proteolytic cleavage of IgG by IdeS 32

3.1 Disorder prediction using RONN and PONDR 77
3.2 Predicted EndoS domain architecture and construct boundaries 78
3.3 *E. coli* over-expression of GST-EndoS 81
3.4 *E. coli* over-expression of EndoS fragments 82
3.5 Affinity purification of GST-EndoS 84
3.6 Purification of EndoS using AEC 85
3.7 Purification of EndoS using SEC 85
3.8 Affinity purification of GST-ChitLRR 86
3.9 Purification of ChitLRR using AEC 87
3.10 Purification of ChitLRR using SEC 87
3.11 Purification of CBM-KO using IMAC 89
3.12 Purification of CBM-KO using AEC 89
3.13 Purification of CBM-KO using SEC 90
3.14 Purification of CBM using IMAC and SEC 91
3.15 Purification of CBMCT using IMAC and SEC 92
3.16 Purification of IgG1 Fc and 2G12 IgG using SEC 93
3.17 Affinity purification of EndoS catalytic mutants 96
3.18 Testing activity of EndoS catalytic mutants against IgG1 Fc 96
3.19 Analytical SEC of EndoS^{E235L} and Fc 98
3.20 Analytical SEC of EndoS^{E235L} and 2G12 99
3.21 CD spectra of EndoS and EndoS^{E235L} 100
3.22 CD spectra of CBM and CBMCT 101
3.23 Crystallisation of EndoS 103
3.24 SAXS data and the *P(r)* of EndoS 106
3.25 DAMMIF models of EndoS in solution 107
3.26 DAMMIN refined model of EndoS in solution 108
3.27 SAXS analysis of EndoS fragments 110

3.28 The crystal structure of EndoS . . . 112
3.29 SUPCOMB fit of the crystal structure of EndoS to the SAXS envelope . . . 113
3.30 SUPCOMB fit of EndoS crystal structure subdomains to SAXS envelopes from EndoS fragments . . . 114
3.31 Analysis of EndoS SAXS data using EOM . . . 114
3.32 DAMMIF models of the EndoS^{E235L}Fc complex . . . 116
3.33 DAMMIN models of the EndoS^{E235L}Fc complex . . . 117
3.34 Prediction of EndoSFc complex structures using ClusPro . . . 118
3.35 Preliminary SAXS model of an EndoSFc complex . . . 120

4.1 Purification of CBM . . . 127
4.2 Crystallisation of CBM . . . 130
4.3 CBM linescan data collection . . . 131
4.4 Selection of a space group for the SeMet CBM diffraction data using systematic absences and the SRF . . . 131
4.5 Selection of a space group for the SeMet CBM diffraction data using R_{meas} 132
4.6 Fluorescence scan of SeMet CBM crystals . . . 134
4.7 PP and FOM vs. resolution for SAD phases . . . 136
4.8 Anomalous difference map for SeMet CBM . . . 138
4.9 Ramachandran analysis of the refined CBM model . . . 138
4.10 The asymmetric unit of the CBM crystal structure in $P3_2$ . . . 140
4.11 Analysis of the SRF of CBM calculated using F_o for data scaled in $P3_2$ . . . 141
4.12 The crystal structure and topology of CBM . . . 142
4.13 A representation of the B factors of the three CBM monomers . . . 143
4.14 A ^{15}N HSQC spectrum of CBM . . . 145
4.15 NMR analysis of CBM in the presence of monosaccharides . . . 146
4.16 Non-specific perturbation of CBM chemical shifts induced by monosaccharide titrations . . . 147
4.17 Isothermal calorimetry titrations of the CBM with D-galactose . . . 148
4.18 Crystallisation of CBM using individual monosaccharides . . . 149
4.19 Ramachandran analysis of CBM monosaccharide soaks . . . 150

5.1 Schematic representation of N-linked glycans . . . 157
5.2 Purification of IgG fragments . . . 158
5.3 Glycan profiling of serum IgG and recombinantly expressed IgG fragments . 159
5.4 Fragmentation analysis of released glycans from IgG . . . 161
5.5 Analysis of EndoS-released glycans from serum IgG and IgG fragments . . . 163
5.6 Deletion of the CBM or C-terminal region of EndoS affects glycoside hydrolase activity against serum IgG . . . 164
5.7 HPLC analysis of residual serum IgG glycan species present after a 1 hour incubation with either CBM-KO or ChitLRR . . . 165
5.8 Assessing CBM and CBMCT binding to IgG1 Fc . . . 166
5.9 Glycan profiling of engineered Fc glycoforms . . . 168
5.10 EndoS digestion of engineered Fc glycoforms . . . 169
5.11 Preparation of Fab and Fc fragments from human serum IgG . . . 171
5.12 Glycan profiling of Fab and Fc fragments from human serum IgG . . . 172
5.13 EndoS activity against Fab and Fc fragments from human serum IgG . . . 173
5.14 EndoS activity against HIV gp120 glycans . . . 174
5.15 Protein structures of abundant serum glycoproteins . . . 175
5.16 SDS-PAGE analysis of EndoS treated serum glycoproteins . . . 176
5.17 EndoS activity against serum glycoproteins . . . 178
5.18 EndoS activity against whole human serum . . . 179

5.19 The crystal structure of IgA Fc in complex with FcαRI 186

6.1 Purification of Fc and Fc^{GlcNAc} for SAXS 190
6.2 The solvent corrected SAXS data for Fc and Fc^{GlcNAc} 191
6.3 The *P(r)* of Fc and Fc^{GlcNAc} 192
6.4 Dimensionless Kratky plot of Fc and Fc^{GlcNAc} 193
6.5 Glycosylation profiles of *Mm*IgG1Fc and *Mm*IgG2aFc 196
6.6 Purification of deglycosylated *Mm*IgG1Fc and *Mm*IgG2aFc 196
6.7 Crystallisation of *Mm*IgG1Fc 197
6.8 Crystallisation of *Mm*IgG2aFc 198
6.9 Ramachandran analysis of *Mm*IgG1Fc and *Mm*IgG2aFc 202
6.10 The crystal structure of *Mm*IgG1Fc 203
6.11 The crystal structure of *Mm*IgG2aFc 204
6.12 The open conformation of the *Mm*IgG2aFc structure 205
6.13 A representation of the *B* factors of *Mm*IgG1Fc and *Mm*IgG2aFc structures 205
6.14 Electron density for the CH2 loops of *Mm*IgG1Fc and *Mm*IgG2aFc 206
6.15 Structural alignment of IgG Fc crystal structures with *Mm*IgG1Fc and *Mm*IgG2aFc 208
6.16 Structural alignment of the CH2 domain of IgG1 Fc crystal structures with *Mm*IgG1Fc and *Mm*IgG2aFc 209
6.17 Multiple sequence alignment of the Fc domains from the IgG subclasses of *M. musculus* 211
6.18 Multiple sequence alignment of the Fc domains from the IgG subclasses of *Homo sapiens* 212
6.19 Mapping residues unique to the *Mm*IgG2aFc amino acid sequence 213

7.1 Multiple alignment of IdeS from GAS serotypes M1 and M59 222
7.2 *E. coli* over-expression of IdeS 223
7.3 Purification of IdeS-SS-M59 and IdeSFL-M59 using IMAC and SEC 224
7.4 Analytical SEC of IdeSFL-M59 and Fc 225
7.5 SDS-PAGE analysis of IdesFc complexes 226
7.6 Putative crystals of an IdeSFL-M59:Fc complex 228
7.7 X-ray diffraction data from putative crystals of an IdeSFL-M59:Fc complex 228
7.8 DAMMIN models of the IdeS-Fc complex 231
7.9 Prediction of IdeSFc complex structures using ClusPro 233
7.10 Preliminary model of the IdeS-Fc complex 234

List of Tables

1.1 Classification of Group A Streptococcal Infection . . . 3
1.2 Relative proportions of human serum IgG *N*-glycans . . . 19

2.1 PCR protocol for Q5® 2X MasterMix . . . 41
2.2 Primers for EndoS cloning . . . 43
2.3 Primers for IdeS cloning . . . 43
2.4 Primers for IgG cloning . . . 45
2.5 Primers for MBP cloning . . . 46
2.6 Primers for EndoS mutagenesis . . . 47
2.7 Primers for IdeS mutagenesis . . . 47
2.8 Summary of protein constructs . . . 48
2.9 Extinction coefficients used for determining protein concentration . . . 58
2.10 Parameters for determining serum glycoprotein protein concentration . . . 59
2.11 $I(0)$ values of BSA standards . . . 64
2.12 Conditions used for CBM crystal optimisation . . . 67
2.13 X-ray diffraction data collection parameters . . . 68
2.14 X-ray diffraction data collection parameters for CBM crystals . . . 69
2.15 HPLC gradients . . . 73

3.1 Bioinformatic analysis of EndoS and EndoS fragments using FUGUE . . . 79
3.2 Analysis of CD spectra from EndoS and EndoS^{E235L} using CDSSTR . . . 101
3.3 Analysis of CD spectra from CBM and CBMCT using CDSSTR . . . 102
3.4 SAXS analysis EndoS . . . 105
3.5 SAXS analysis of EndoS fragments . . . 109
3.6 SAXS analysis of the EndoS^{E235L}Fc complex . . . 115

4.1 CBM X-ray diffraction data collection statistics . . . 133
4.2 Phaser molecular replacement statistics . . . 134
4.3 SHARP refinement of SeMet residue positions . . . 135
4.4 SAD phasing statistics from SHARP and after density modification with SOLOMON . . . 137
4.5 CBM refinement statistics . . . 139
4.6 DALI results . . . 144
4.7 Diffraction data collection statistics for CBM monosaccharide soaks . . . 151
4.8 Refinement statistics for CBM monosaccharide soaks . . . 152

6.1 SAXS analysis of Fc and Fc^{GlcNAc} . . . 191
6.2 *Mm*IgG1Fc and *Mm*IgG2a diffraction data collection statistics . . . 199
6.3 *Mm*IgG1Fc and *Mm*IgG2a refinement statistics . . . 201

7.1 SAXS analysis of putative IdeSFL-M59:Fc complex . . . 230

List of Abbreviations

2AA	2-aminobenzoic acid
2AB	2-aminobenzamide
α_2M	α_2-macroglobulin
ADCC	Antibody-dependent cell-mediated cytotoxicity
AEC	Anion exchange chromatography
ARF	Acute rheumatic fever
ATCC	American Type Culture Collection
B4GalT1	β-1,4-galactosyltransferase
BBFNN	Bio-Basis Function Neural Network
bp	Base pairs
BSA	Bovine serum albumin
C4BP	C4b-binding protein
CBM	Carbohydrate-binding module
CDC	Complement dependent cytotoxicity
CD	Circular dichroism
CID	Collision-induced decomposition
CII	Collagen II
CMP-sialic acid	Cytidine-5′-monophospho-*N*-acetylneuraminic acid
CV	Column volumes
DAM	Dummy atom model
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DESY	Deutsches Elektronen-Synchrotron
DHB	2,5-dihydroxybenzoic acid
DLS	Diamond Light Source
DMEM	Dulbecco's modified Eagle's Medium
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNaseI	Deoxyribonuclease I

DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EndoS	Endoglycosidase S
EOM	Ensemble optimisation method
ESI MS	Electrospray ionisation mass spectrometry
ESRF	European Synchrotron Radiation Facility
FBS	Fetal bovine serum
FcγR	Fcγ receptor
Fc^{GlcNAc}	IgG1 Fc deglycosylated with EndoS
FcRn	Neonatal Fc receptor
FH	Factor H
FHL-1	Factor H-like protein 1
FOM	Figure of merit
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
GAS	Group A streptococcus
GnTI	GlcNAc transferase I
GPI	Glycosylphosphatidylinositol
GST	Glutathione S-transferase
GU	Glucose units
HDX	Hydrogen-deuterium exchange
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hIg	Hybrid immunoglobulin
HRP	Horseradish peroxidase
HSQC	Heteronuclear single-quantum coherence
HT	High-throughput
IdeS-SS	IdeS with no signal sequence
IdeSFL	Full-length IdeS
IdeS	Immunoglobulin G-degrading enzyme of *S. pyogenes*
IEF	Iso-electric focusing
IgG	Immunoglobulin G

Ig	Immunoglobulins
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITAM	Immunoreceptor tyrosine-based activation motif
ITIM	Immunoreceptor tyrosine-based inhibition motif
ITP	Immune thrombocytopenic pupura
IVIg	Intravenous immunoglobulin
LB	Lysogeny broth
LRR	Leucine rich repeats
LTA	Lipoteichoic acid
MAC	Membrane attack complex
MALDI-TOF	Matrix-assisted laser desorption/ionisation time-of-flight
MALLS	Multi-angle laser light scattering
mAU	Milli absorbance units
MBP	Maltose binding protein
MME	Monomethyl ether
*Mm*IgG1Fc	*Mus musculus* 6A6-IgG1 Fc
*Mm*IgG2aFc	*Mus musculus* 6A6-IgG2a Fc
MOPS	3-(*N*-morpholino)propanesulfonic acid
MS	Mass spectrometry
M_w	Molecular weight
MWCO	Molecular weight cut-off
NCS	Non-crystallographic symmetry
NEB	New England Biosciences
Ni-NTA	Ni-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NSD	Normalised spatial discrepancy
OD	Optical density
P/S	Penicillin/streptomycin
PBS-T	PBS, 0.05% v/v Tween 20
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCT	Pre-crystallisation test
PEI	Polyethyleneimine
PNGase F	Peptide *N*-glycosidase F

PP	Phasing power
PSGN	Post-streptococcal glomerulonephritis
Q-TOF	Quadrupole-time-of-flight
QELS	Quasi-electron light scattering
RF	Relative fluorescence
RONN	Regional Order Neural Network
rpm	Rotations per minute
SAD	Single-wavelength anomalous diffraction
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
SBD	Substrate binding domain
scIgG	Single cleaved IgG
SDS-PAGE	Sodium-dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SeMet	Selenomethionine
Sib35	Streptococcal immunoglobulin-binding protein 35
SIgA	Secreted IgA
SIRAS	Single isomorphous replacement with anomalous scattering
SLE	Systemic lupus erythematosus
SLO	Streptolysin O
SLPI	Secretory leukocyte proteinase inhibitor
SLS	Streptolysin S
SpeB	Streptococcal erythrogenic toxin B
SRF	Self-rotation function
ST6Gal1	α-2,6-sialyltransferase
TLS	Translation/libration/screw
TRIM	Tripartite motif protein 21
UDP-galactose	Uridine 5′-diphosphogalactose
WHO	The World Health Organisation

1 Introduction

The design and development of new therapeutic drugs is of huge importance for the continuing health and wellbeing of the human population. In recent years, research has turned towards mining the proteins produced by human pathogens for potential therapeutic agents. Due to the evolutionary pressure for a given pathogen to avoid the human immune system, many of these pathogen-derived proteins are able to specifically modulate the host immune response. As such, these proteins possess therapeutic potential as drugs which alter various aspects of the immune system.

The bacteria *Streptococcus pyogenes* produces a number of proteins which affect both the innate and adaptive arms of immunity in humans. Two of these proteins, Endoglycosidase S (EndoS) and Immunoglobulin G-degrading enzyme of *S. pyogenes* (IdeS), are currently being channelled towards therapeutic use. The ability of these two enzymes to specifically alter the effector functions of immunoglobulin G (IgG) has led to research into their use as treatments for IgG-mediated autoimmune diseases and several other applications. The structural and functional characterisation of these two proteins forms the focus of this thesis.

1.1 *Streptococcus pyogenes*

Streptococcus pyogenes is a Gram-positive, non-motile, catalase-negative bacterium and is member of the genus *Streptococcus*. *S. pyogenes* is a β-haemolytic bacterium and is classified into Lancefield group A [1] based on its cell wall C-carbohydrate. As such, *S. pyogenes* is commonly referred to as group A streptococcus (GAS). Group A streptococci are further classified into M serotypes based on the sequence of the *emm* gene. The *emm* gene encodes the important cell surface M protein, with approximately 200 different *emm* types having now been identified [2]. Individual *S. pyogenes* cells are spherical in shape and can exist as diplococci (Fig. 1.1A) or in chains. In addition to the peptidoglycan layer surrounding the *S. pyogenes* plasma membrane, *S. pyogenes* is capable of producing a capsule of hyaluronic acid (Fig. 1.1B) [3] and, as a facultative anaerobe, *S. pyogenes* is able to survive in both the presence and absence of oxygen.

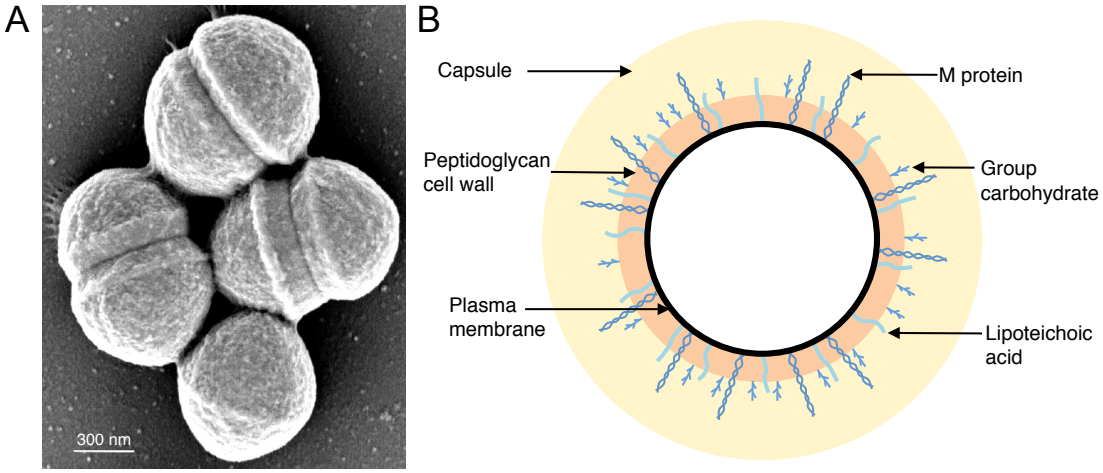


Figure 1.1: Scanning electron micrograph and cell diagram of *S. pyogenes*. (A) *S. pyogenes* pictured using cryo field emission scanning electron microscopy reproduced from Severin *et al.* [4]. (B) Schematic diagram of key cell components of *S. pyogenes*. Components are labelled accordingly with the hyaluronic acid capsule shown in yellow, the peptidoglycan layer in orange and the plasma membrane in black.

The World Health Organisation (WHO) estimates that GAS infections are responsible for 500, 000 deaths worldwide each year [5]. Infections target mucosal membranes and the skin but can also cause disease in deeper tissues. GAS infections cause a number of diseases ranging from mild illnesses, including pharyngitis or impetigo, through to severe, potentially fatal diseases like toxic shock syndrome and necrotising fasciitis (reviewed in [6]). In addition to these suppurative (pus-forming) infections, GAS infection can lead to

non-suppurative sequelae including the autoimmune-mediated disorders of acute rheumatic fever (ARF) and post-streptococcal glomerulonephritis (PSGN). Table 1.1 summarises the different diseases caused by GAS infection as described by Breiman *et al.* [7].

Table 1.1: Classification of Group A Streptococcal Infection as proposed by Breiman *et al.* [7].

Infection class	Description	Examples
I	Streptococcal toxic shock syndrome	
II	Other invasive infections	Meningitis, pneumonia, peritonitis, puerperal sepsis, osteomylitis, septic arthritis, necrotising fasciitis, erysipea, cullulitis
III	Scarlet fever	
IV	Non-invasive infections	Pharyngitis, tonsilitis, otitis media, sinusitis, vaginitis, impetigo
V	Non-suppurative sequelae	ARF, PSGN

The cell surface M protein is a well established protective antigen against GAS [8] and as such, has been heavily pursued as a vaccine target. Despite this, the development of a vaccine for GAS has not yet been successful. This is due to two key factors, first, that the M protein-induced response is serotype specific and does not confer generic immunity against GAS. Second, that epitopes of the M protein were revealed to be cross reactive with cardiac myosin, tropomyosin, vimentin, laminin and keratin (reviewed in [9]). This cross-reactivity and molecular mimicry observed in the M protein of GAS is an important factor in the autoimmune sequelae (ARF and PSGN) observed after GAS infections. Indeed, the effects of this cross-reactivity were first observed in a report describing the increased prevalence of ARF in children given a crude M protein vaccine compared to control children [10]. Exploration of GAS vaccine targets is ongoing with some M protein-based approaches still being pursued amongst a host of alternative GAS antigens [11].

Non-invasive infections such as pharyngitis and impetigo are highly susceptible to penicillin and generally respond well to treatment. Treatment with penicillin is not always successful however, with treatment failure likely occurring due to a number of factors including penicillin tolerance in GAS, the ability of GAS to survive intracellularly and the presence of cohabiting β-lactamase producing bacteria [12]. For severe infections, protein synthesis inhibitors such as clindamycin are used in addition to penicillin and surgical removal of infected tissues may be required. Intravenous immunoglobulin (IVIg) therapy is also used

to treat invasive GAS infections, as a number of clinical studies suggest that IVIg therapy reduces mortality rates in patients suffering from streptococcal toxic shock syndrome or necrotising fasciitis. The success of IVIg treatment is potentially a result of providing the patient with neutralising antibodies against disease-inducing streptococcal superantigens and exotoxins [13, 14, 15].

1.1.1 *Streptococcus pyogenes* invasion and infection

As described in the previous section, *S. pyogenes* primarily infects the throat, causing pharyngitis or targets the skin resulting in a number of cutaneous and subcutaneous infections (Table 1.1). To cause infection, *S. pyogenes* must first colonise these areas of the body by adhering to cells of these target tissues. Group A streptococci attach themselves to the extracellular matrix (ECM) of eukaryotic cells or directly to cells themselves via a number of protein and non-protein components collectively called adhesins. These GAS cell surface molecules include the hyaluronic acid capsule, lipoteichoic acid (LTA), fibrils or pili, M proteins and a number of other fibronectin-binding proteins (for a review see [16]). Until recently, GAS was thought to exist solely as an extracellular pathogen but experiments by LaPenta *et al.* in 1994 successfully identified the intracellular occurrence and survival of GAS [17]. This internalisation occurs via host-cell integrins and a collection of GAS proteins called invasins [16].

1.1.2 The host response to *Streptococcus pyogenes* infection

Upon infection by GAS, the host ellicits a number of innate immune responses in an effort to clear GAS from the body. Innate immune responses are initially stimulated by the interaction of GAS-derived pathogen-associated molecular patterns with host pattern recognition receptors. However, the specific ligands and receptors involved in these interactions for GAS have not yet been resolved [18]. One of the first stimulated lines of defence is host-production of antimicrobial peptides (AMPs) including α-definsin, β-defensins (HBD1–3) and cathelicidin (LL-37) [19]. AMPs can function as chemotactic stimuli to draw phagocytic cells towards the bacteria, with defensins also having the ability to target GAS directly by inhibiting cell wall synthesis [20].

As reviewed by Kovarik and Fieber [18], recruitment of macrophages, neutrophils and dendritic cells is an integral step in the clearance of group A streptococci as it leads to phagocytosis and killing of GAS. An additional immune response to GAS is the production of neutrophil extracellular traps (NETs). NETs are produced by neutrophils during an apototic-like process called NETosis [21]. NETs are composed primarily of chromatin with additional protein elements including histones, granule proteins and antimicrobial peptides. These components function together to bind and kill microorganisms [22].

The activation of both the classical and alternative complement (C) pathways forms another important host response to GAS infection. Indeed, many immune-evasion tactics of GAS (discussed in Section 1.3) inhibit pathogen clearance by targeting the complement pathway. In general, opsonisation of microorganisms can occur via the deposition of the complement opsonin C3b on GAS cell surfaces. Complement receptors (CR) expressed on the cell surface of phagocytes are then able to recognise and bind the surface-deposited C3b leading to phagocytosis and killing of the tagged microorganism. The complement cascade may also progress towards the production of membrane-attack complex (MAC) causing cell lysis of the microorganism. Alternatively, the cascade may continue towards the production of chemotactic complement factors including C5a, which recruit phagocytes to the site of infection [23].

The adaptive immune response to GAS also plays a role clearance of GAS infections where serotype specific antibodies to the M protein of GAS are protective against GAS [8]. Specifically, antibodies to the N-terminal region of the M protein are able to induce effective opsonisation and phagocytosis of GAS. This illustrates an important route of GAS clearance, albeit a serotype specific response. Indeed, antibodies against the M protein have been investigated thoroughly as a target for a GAS vaccine but due to various complications, have not yielded a commercially available vaccine, as discussed earlier in Section 1.1.

Despite the efforts of the immune system, GAS is able to specifically and effectively circumvent host immune responses using a variety of immune evasion factors which are described in Section 1.3. These immune evasion tactics of GAS lead to severe infectious disease, characterised by extensive tissue damage, multiple pathologies and in extreme cases, death. A number of the immune evasion factors from GAS, including EndoS and

IdeS, interact with and alter the function of immunoglobulin G (IgG). To appreciate the effect of these IgG-targeting immune evasion factions, IgG and its role in the immune system are described in detail in the following section.

1.2 Immunoglobulins

Immunoglobulins (Ig), or antibodies, play a crucial role in the human immune system. Antibodies are large glycoproteins produced by B cells which are found primarily in the blood and extracellular spaces. There are a number of different classes or isotypes of immunoglobulins which are grouped according to their structure, function and localisation.

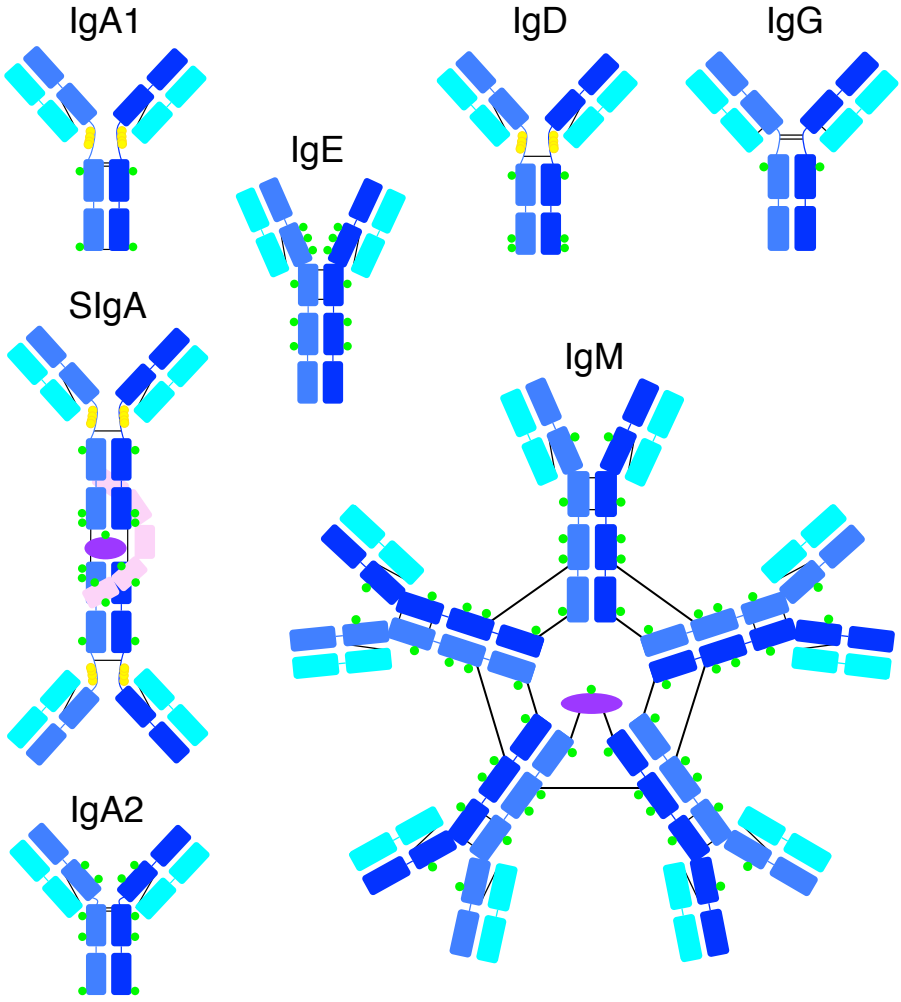


Figure 1.2: Human Ig isotypes. A schematic representation of human Ig isotypes shows heavy chains in blue with light chains in cyan. Although IgG exists in four distinct subclasses, only the IgG1 subclass is depicted here. For SIgA and IgM, the J chain is shown in purple and the secretory component of SIgA is shown in pink. Disulphide bonds are represented as black lines and glycosylation sites indicated by small circles (green, *N*-linked; yellow, *O*-linked).

All of the human Ig isotypes follow a similar structural pattern where two heavy chains and two light chains come together to form a homodimer of heterodimers linked by disulphide bonds. In humans, there are five of these isotypes, namely IgA, IgD, IgE, IgG and IgM (Fig. 1.2). Each Ig isotype has a unique heavy chain ($\alpha, \delta, \epsilon, \gamma$ and μ, respectively) which determines which isotype it belongs to, whereas the light chain is classified into only two types, kappa (κ) and lambda (λ) [24]. B cells use a highly regulated isotype switching system to respond to external stimuli to ensure the production of the appropriate isotype during an immune response [25].

IgG is the most abundant immunoglobulin in serum and is found at concentrations of approximately 10 mg mL^{-1} in healthy adults [26]. IgG exists in four subclasses, IgG1, IgG2, IgG3 and IgG4 which are numbered according to their abundance in human serum with IgG1 being the most prevalent subclass [23]. The differences between IgG subclasses manifest primarily in the hinge region, where the subclasses show variation in the number of bridging disulphide bonds (Fig. 1.3). Further to this, the position of the disulphide bonds which link the heavy and light chains together also differs between subclasses. The four IgG subclasses show differing affinities for Fc receptors resulting in subclass-dependent effector functions [27]. IgG forms much of the adaptive immune response in humans and will be discussed in detail in the following sections.

IgA is split into two subclasses, IgA1 and IgA2, which have evolved as a result of gene duplication. IgA exists as a monomer in serum (approximately 90% IgA1 and 10% IgA2) and circulates at about 2 mg mL^{-1} [23]. IgA is able to form disulphide bond-linked dimers via an additional 15 kDa protein subunit, the J chain, and the addition of a further component, the secretory chain, gives rise to secretory IgA (SIgA). SIgA is the primary form of IgA antibody found on mucosal surfaces [31] and as such, IgA acts as an important molecule in preventing the adherence of pathogens, like GAS, to these surfaces [32]. The role of IgD in the immune system remains under investigation where the exact function of this isotype is yet to be ascertained. It appears that IgD is involved in maintenance of B cell homeostasis with a potential role in general immunomodulation [33]. IgE is present in the serum at low concentrations ($\sim$150 ng mL^{-1}) and is sequestered to the tissues to provide local antibody-mediated immunity [34, 35]. IgE binds to the high affinity IgE receptor FcϵRI on both mast cells and basophils and plays a central role in hypersensitivity

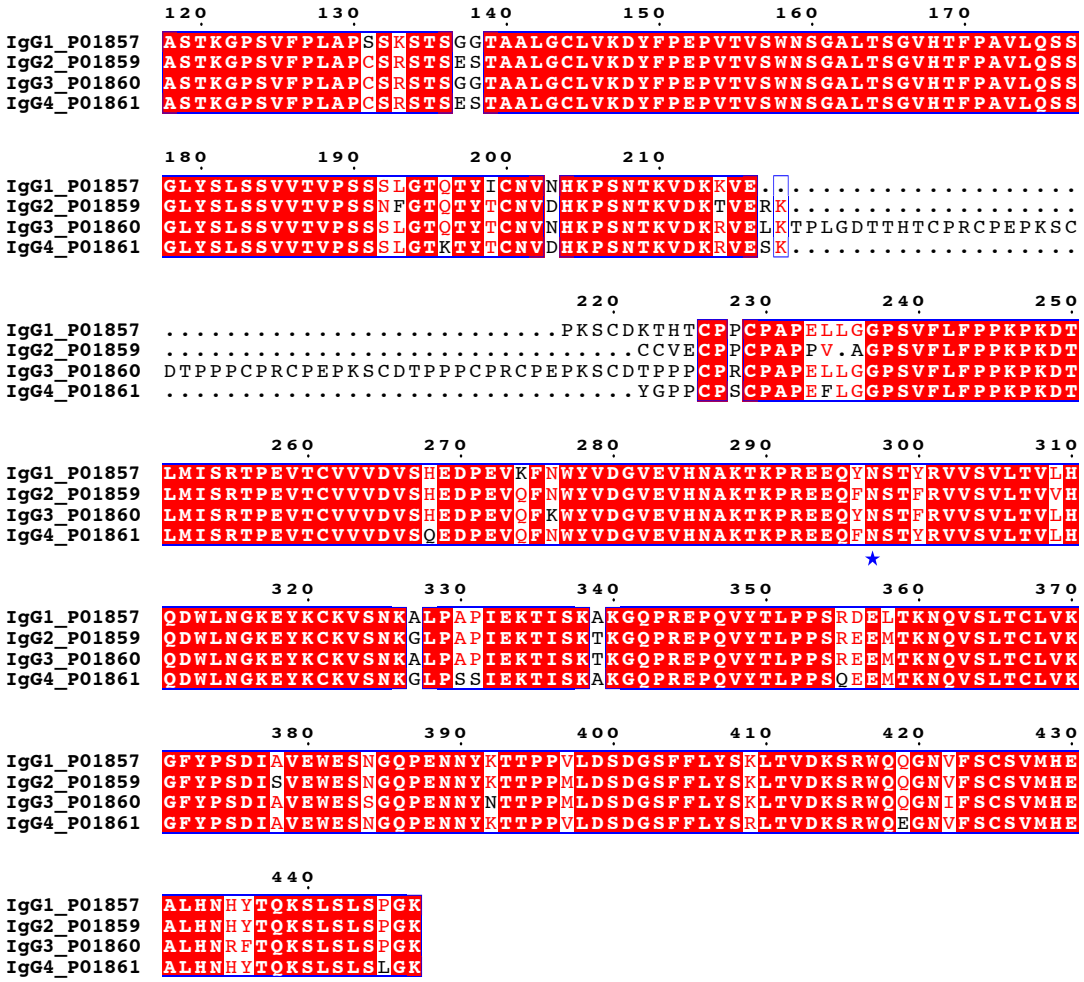


Figure 1.3: Multiple alignment of the conserved domains of the IgG heavy chain. The alignment of generic sequences for human IgG1 (Uniprot ID: P01857), IgG2 (P01859), IgG3 (P01860) and IgG4 (P01861) was made using ESPript 3.0 [28] and a multiple alignment file generated using Clustal Omega [29]. Residues are numbered according to Eu antibody numbering convention [30]. Conserved residues are shown in white text and are highlighted in red. Non-conserved residues are shown in black text on a white background. Residues in red text indicate conservative changes or partial conservation across different subclasses. The position of the *N*-linked glycan is indicated by a blue star at position 297.

and allergic reactions [36]. IgM is produced by naive B cells without the need for antigenic stimuli and is present in human serum at concentrations of approximately 1.5 mg mL^{-1} [37, 23]. IgM exists as a pentamer in serum, oligomerised by a network of disulphide bonds and an additional J chain. IgM is a potent activator of the classical complement pathway through C1q binding and is an important protein in innate immunity [38].

1.2.1 The structure of IgG

The crystal structure of a human IgG1 antibody is shown in Figure 1.4A and reveals the typical 'Y' shape of immunoglobulins. IgG is a heterodimeric protein comprising two covalently linked subunits each containing a heavy chain and a light chain. This multi-domain structure contributes to the two key functions of IgG. Firstly, IgG has antigen-binding capability manifested in the variable Fab (fragment antigen binding) domains of the protein. Secondly, IgG has the capacity to elicit immunological responses through the conserved Fc (fragment crystallisable) domain. These Fc-mediated responses include binding to various Fcγ receptors (FcγRs) to instigate antibody-dependent cell-mediated cytotoxicity (ADCC) and phagocytosis, as well as activation of the complement pathway [39]. Within the Fc domain, there is an *N*-linked glycosylation site on the IgG heavy chain which contributes heavily to the structural integrity and function of the IgG Fc and is discussed in more detail in Section 1.2.3 [40, 41].

IgG Fc

The Fc domain of IgG is involved in eliciting the effector functions of the immune response. The crystal structure of Fc was first described by Huber *et al.* in 1976 [43], and since then over 40 additional structures of Fc, both in complex with Fc binding proteins and alone, have been published [44]. Fc exists as a homodimer of the C-terminal half of the IgG heavy chain (Fig. 1.5). Together, the contributing heavy chain monomers join to form a characteristic horse-shoe shape. Each monomer comprises two subdomains, CH2 and CH3 (or Cγ2 and Cγ3), which both exhibit an Ig fold. Indeed, all of the subdomains of IgG exhibit these Ig folds in both the constant and variable regions. In the IgG domains, these Ig folds consist of either seven or nine β-strands which come together to form two β-sheets. The β-sheets arrange into a greek-key β-barrel domain which is stabilised by

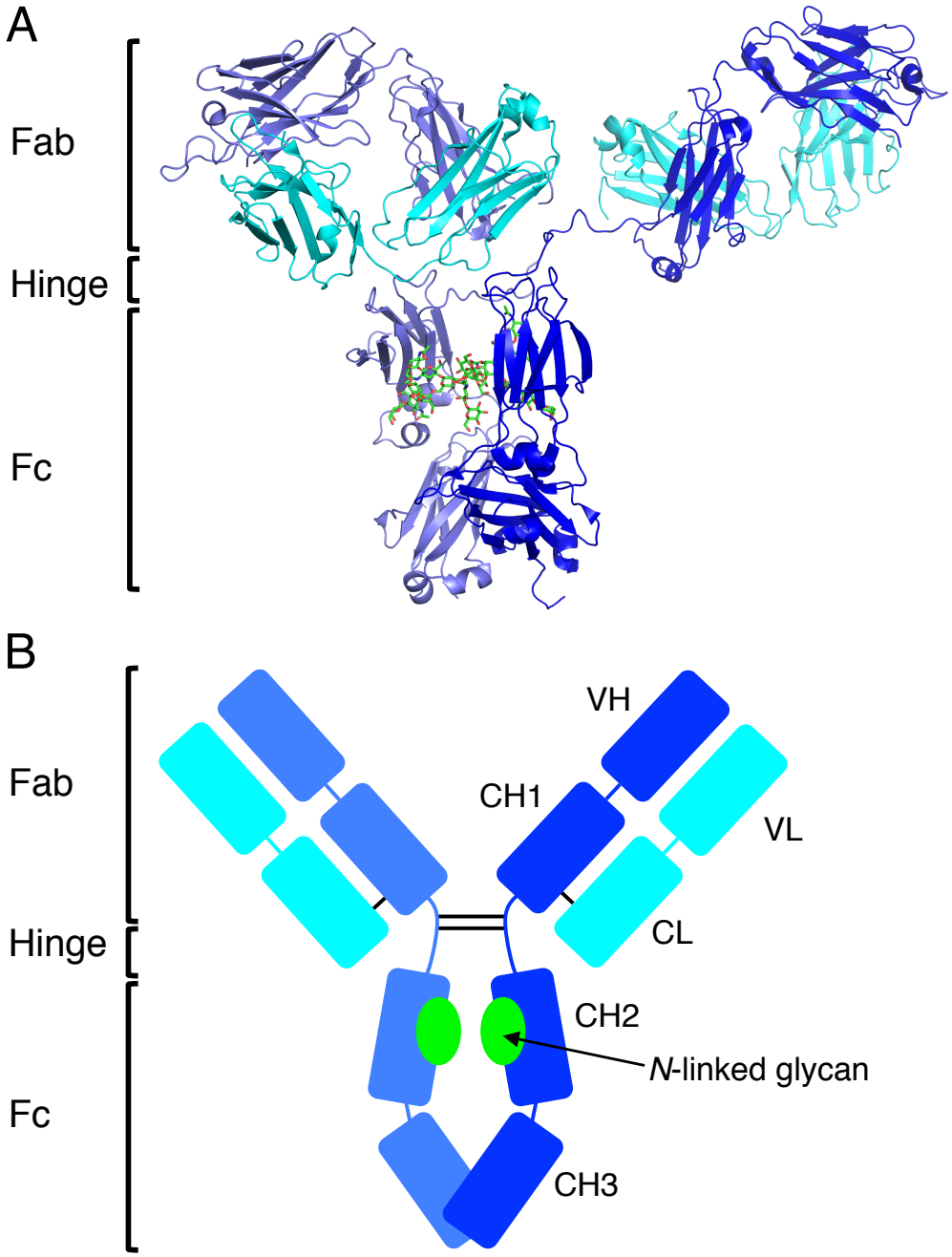


Figure 1.4: The structure of human IgG1. (A) The crystal structure of monoclonal human IgG1 b12 (PDB ID: 1HZH, [42]), shows two Fab domains connected by an extended hinge region to the Fc domain. The protein structure is shown in cartoon representation with the heavy chains coloured blue and marine and both light chains shown in cyan. The *N*-linked glycan at Asn297 is shown in stick representation in green. The oxygen atoms of the glycan are coloured red and the nitrogens are coloured blue. (B) A schematic representation showing the key structural features is shown, using the same colour scheme as in panel A. In addition, the position of the disulphide bonds which link the heavy chains to each other and to their respective light chains are indicated by black lines. The subdomains of one of the heterodimers of IgG1 are labelled to show the location of variable domains of the heavy and light chains (VH and VL), the constant domain of the light chain (CL) and the three constant domains of the heavy chain (CH1, CH2 and CH3).

the formation of an inter-sheet disulphide bond [45]. The CH2 and CH3 domains of Fc both form seven-stranded Ig folds, with two sheets of three and four strands each. The strands are conventionally labelled from A to G and are described in detail with respect to the CH2 domain in Figure 1.6.

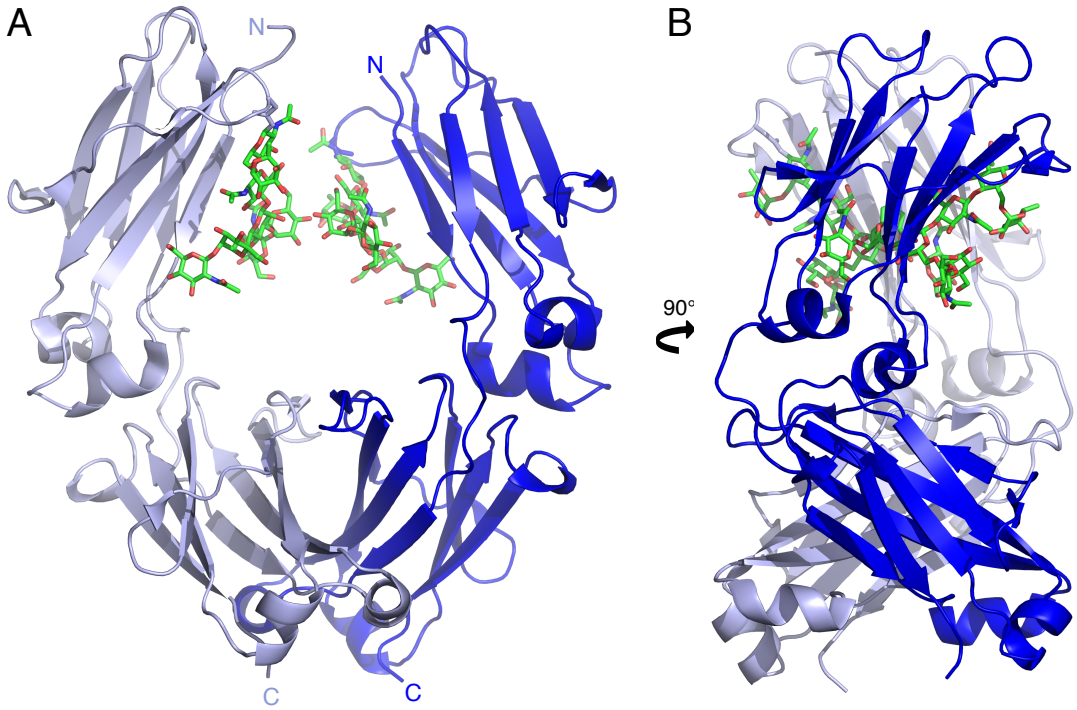


Figure 1.5: The crystal structure of IgG1 Fc. The Fc dimer (PDB ID: 3AVE [46]) is shown in two orientations where panel A is rotated 90° to give rise to the view in panel B. The protein is shown in cartoon representation with the individual chains coloured in slate and blue and the N- and C-termini labelled accordingly. The *N*-linked glycans at position Asn297 of each monomer are shown in stick representation in green with oxygen atoms coloured red and nitrogen atoms coloured blue.

The major protein–protein contacts of the Fc dimer are formed between the CH3 domains. Analysis of this interaction using PISA [47] gives a buried surface area of approximately 1100 Å^2 with ΔG = -11.6 kcal mol^{-1}. The dimer is also linked by a disulphide bond in the hinge region. The hinge region containing this disulphide linkage is generally not well resolved, if at all, in crystal structures of IgG Fc. This is attributed to the lack of secondary structure in the hinge and its potential for flexibility. The cavity between the CH2 domains of Fc is filled by the *N*-linked glycan at Asn297, which forms extensive contacts with the inner β-sheet of the CH2 domain [43]. Further discussions on the glycosylation of IgG and the glycan types found at this position are presented in Section 1.2.3.

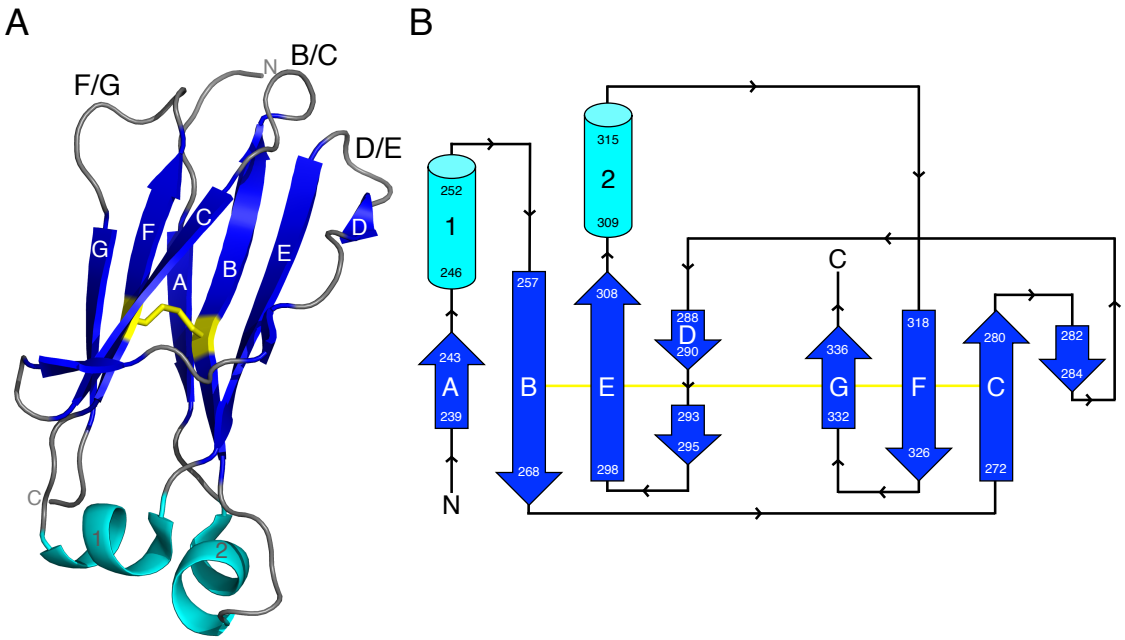


Figure 1.6: The structure and topology of the IgG Fc CH2 domain. (A) The crystal structure of the CH2 domain from IgG Fc taken from a monomer of PDB ID: 3AVE [46] with glycans omitted for clarity. The orientation follows that of the foreground CH2 domain (blue) in Fig 1.5B. The strands are labelled A-G according to convention and the helices are labelled 1-2. The CH2 loops are labelled (F/G, B/C and D/E) as are the N- and C-termini. The disulphide bond between residues Cys261 and Cys321, which bridges the two β-sheets is shown in yellow. (B) A topology map of the CH2 domain of Fc was produced based on an output from PDBsum [48]. The secondary structure elements and termini are labelled as in A.

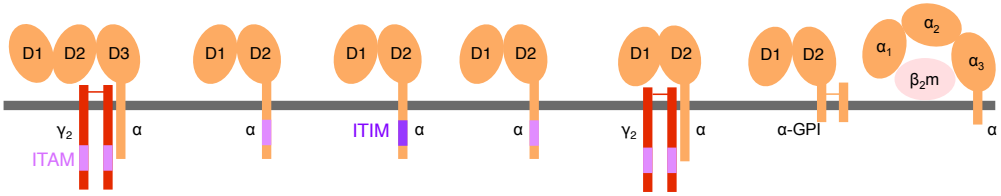
1.2.2 The role of IgG in immunity

IgG has a central role in the adaptive immune response. Antigen binding capabilities of the Fab domains combined with the effector functions of the Fc domains allows the body to selectively recognise and destroy pathogens. The functions of IgG are described briefly here with respect to bacterial infection and can be divided into three sections namely; neutralisation, opsonisation and complement activation.

During the process of neutralisation, IgG is able to target pathogenic bacterial toxins by binding them directly via antigenic epitopes. Binding of such toxins in this way reduces their ability to execute their pathogenic potential as well as labelling them for phagocytosis [23]. Opsonisation involves the effective 'tagging' of pathogens through binding of the Fab domains to antigenic moieties on the cell surface of the pathogen. Coating the target with IgG enables phagocyte recognition of the pathogen via the Fc domains of the opsonising antibodies. The Fc domains are recognised by the FcγRs present on phagocytic cells [49], which, upon binding the Fc of opsonising antibodies, signal phagocytosis of the target pathogen.

IgG can also mediate bacterial killing by activation of the classical complement pathway through binding of C1q [50]. In essence, the complement system is a complex enzymatic cascade which leads to the destruction of the target cell and is reviewed in depth by Rother *et al.* [51]. Fc binding of C1q instigates a series of enzymatic cleavages of various complement factors resulting in the deposition of complement C3b on the surface of the target pathogen. Phagocytes not only recognise the Fc fragments of opsonising antibodies via FcγRs but also recognise complement deposited on the surface of the pathogen via the complement receptors CR1 and CR3 resident on the surface of phagocytes. Binding of complement and Fc in this way again results in phagocytosis and killing of the tagged pathogen. Activation of complement via the classical pathway can also lead to the production of chemotactic signals which aim to recruit further neutrophils to the site of infection. Finally, the complement cascade can lead to the formation of the membrane attack complex (MAC) which forms pores in the outer membrane of target cells causing lysis and killing.

Interaction with FcγRs



Receptor	FcγRI (CD64)	FcγRIIA (CD32a)		FcγRIIB (CD32b)	FcγRIIC (CD32c)	FcγRIIIA (CD16a)		FcγRIIIB (CD16b)	FcRn
Allele	-	H131	R131	-	-	V158	F158	-	-
Affinity	High	Low	Low	Low	Low	Low	Low	Low	High
Kd	15 nM	0.2 μM	0.3 μM	8.3 μM	8.3 μM	0.5 μM	0.9 μM	5 μM	0.1μM
Subclass preference	1>3>4	1>>3>2>4		4>3>1>2	4>3>1>2	3>>1>4>2		3>1	NA

Figure 1.7: Human Fc receptors. A schematic representation of the Fc receptors is given where receptors (orange) are shown inserted into a plasma membrane (grey). FcγRI and FcγRIIIA are shown with the potentially associated γ_2 chains (red). The intracellular signalling motifs are shown in magenta (ITAM, activatory) and purple (ITIM, inhibitory) with FcγRIIIB shown with a GPI-anchor. FcRn is shown with an associated β_2-macroglobulin subunit (pink). The tabulated values are derived from Bruhns *et al.* [52] with Kd values reported for binding to human IgG1. NA; not available.

These phagocytic processes rely on the successful interaction of the IgG Fc domain with cell surface FcγRs. FcγRs are a family of single-pass transmembrane proteins of which six have been identified in humans. A schematic diagram of these receptors is illustrated in Figure 1.7 along with the measured affinity of each receptor to human IgG1. These receptors are grouped into activating and inhibitory receptors. Activating receptors (FcγRI, FcγRIIA, FcγRIIC and FcγRIIIA) usually associate with a transmembrane signalling component containing the putative activator immunoreceptor tyrosine-based activation motif (ITAM) [49]. The inhibitory receptor FcγRIIB contains an immunoreceptor tyrosine-based inhibition motif (ITIM) [49]. The remaining receptor FcγRIIIB shows 99% homology to the activatory FcγRIIIA receptor but has a glycosylphosphatidylinositol (GPI) anchor and is found only on neutrophils [53, 54], where although it has been shown to be important in the neutrophil response to immune complexes, its role in the immune system remains largely unclear [55, 56].

X-ray crystallographic analysis of complexes of Fc:FcγRIIIB [57, 58], Fc:FcγRIIIA [59], Fc:FcγRIIA [60] and, more recently, Fc:FcγRIIB [61] reveal that the Fc:FcγR interaction is mediated by the CH2 and lower hinge regions of the Fc domain. Although the same overall site on IgG Fc is occupied by the FcγRs, small differences between the binding

modes of the different FcγR classes have been exploited to generate Fc mutants with specific affinities to certain receptors [62]. Figure 1.8 shows the crystal structure of the glycosylated soluble domains of FcγRIIIA with a glycosylated IgG1 Fc domain (PDB ID: 3SGJ [59]). A critical contact is made by the receptor residues Trp87 and Trp110 which form a tryptophan sandwich about the Fc residue Pro329. A salt bridge is formed between Lys117 of the receptor and Asp265 of IgG Fc, and then several further hydrophobic and electrostatic contacts are made along the CH2 loops and the lower hinge (Fc residues 234–238). In addition, the most recent crystal Fc:FcγRIIIA structure utilised a glycosylated FcRIIIA fragment and revealed that extensive contacts are also made between the Fc glycan and the glycan of the receptor [59].

Other binding partners of IgG Fc

In addition to the FcγRs described above, other Fc binding partners of IgG Fc include the neonatal Fc receptor (FcRn), C1q, tripartite motif protein 21 (TRIM21) and potentially dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN). FcRn binds to a site between the CH2 and CH3 domains of IgG distinct from the binding site of the FcγRs and is implicated in controlling the serum half life of IgG antibodies [63]. C1q, as mentioned earlier in this section, is part of the classical complement pathway and binds to Fc in a similar region to the FcγRs. Although no X-ray crystal structure exists for this complex, biochemical and biophysical analysis suggests that C1q makes contacts to the lower hinge and CH2 domains [50, 64].

TRIM21 is a more recently discovered binding partner of IgG Fc, which may be involved in activation of immune signalling of antibodies that have been internalised into cells [65]. The cytoplasmic TRIM21 protein binds to IgG Fc with sub nanomolar affinity at the CH2-CH3 interface [66, 67]. Finally, binding of DC-SIGN to α-2,6-sialylated IgG Fc has been reported to mediate the anti-inflammatory properties of these IgG Fc glycoforms [68, 69]. However, this finding is controversial and is not supported by direct biophysical characterisation of this potential interaction [70].

Other non-human Fc binding proteins of note include the bacterial protein A from *Streptococcus aureus* (PDB ID: 1FC2 [71], PDBI ID: 1L6X) and protein G from group G *Streptococcus* (PDB ID: 1FCC, [72]) which both bind to the CH2-CH3 interface of Fc. Both

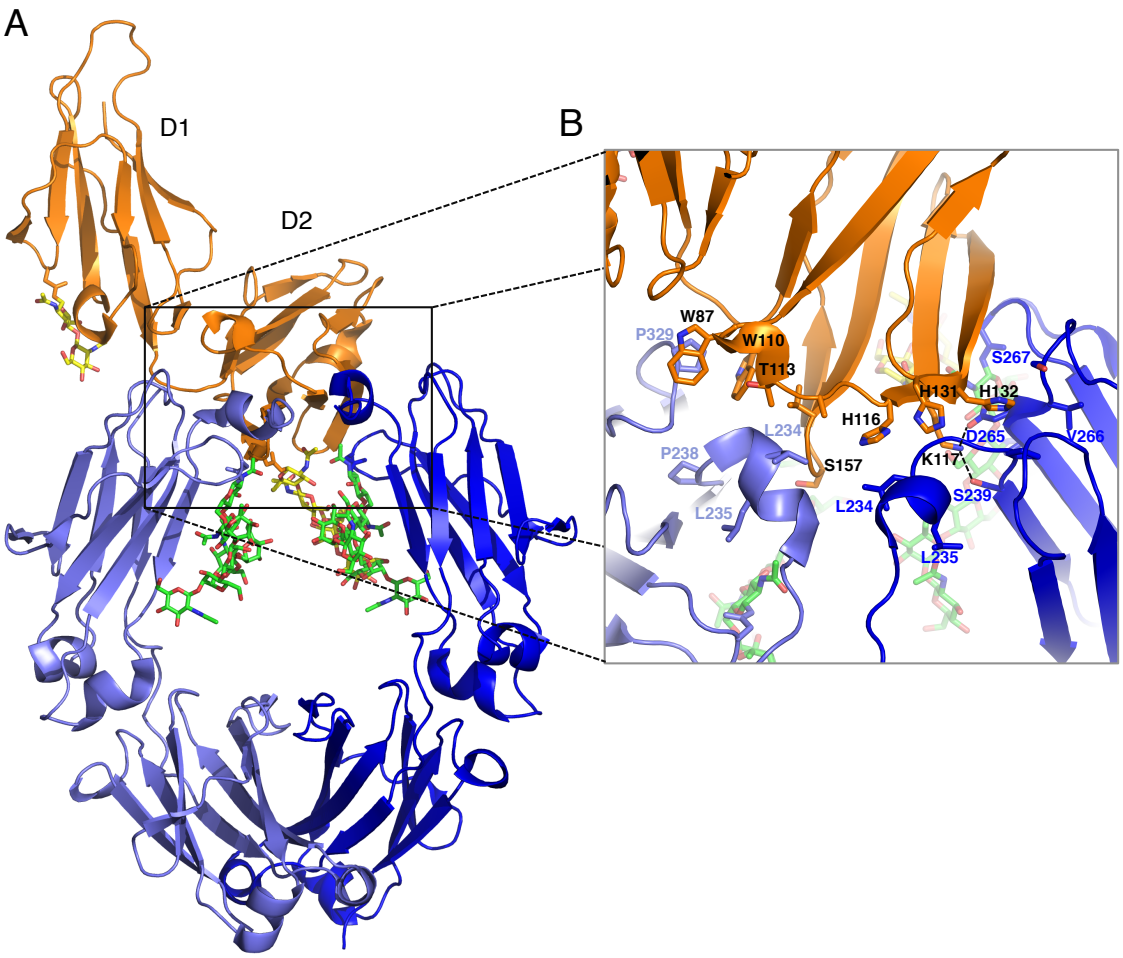


Figure 1.8: The interaction between Fc and FcγRs. (A) The X-ray crystal structure of IgG1 Fc bound to the soluble fragment of the glycosylated FcγRIIIA receptor reveals the 1:1 binding stoichiometry. Both proteins are shown in cartoon representation with the Fc monomer chains coloured slate and blue and the FcγRIIIA receptor shown in orange with domains D1 and D2 labelled accordingly. The glycans of both structures are shown in stick representation with glycans of the Fc in green and glycans of the receptor in yellow. In both cases, glycan oxygens are shown in red with nitrogens in blue. (B) A close up of the boxed section from panel A reveals important residue contacts between Fc and the receptor. Residues at the interface are numbered with receptor residues in black font, and Fc residues labelled in the colour corresponding to each monomer chain.

proteins are involved in enhancing the immune evasion strategies of these bacteria by preventing phagocytosis [73, 74, 75] but have also been exploited as sepharose conjugates for the purification of IgG antibodies.

1.2.3 IgG glycosylation

All four subclasses of human IgG share a conserved glycan at Asn297 which contributes to both IgG structure and function. In addition to this conserved Fc glycan, the variation in Fab sequences means that glycosylation sites and thus *N*-glycans are present on approximately 15–20% of normal human IgG Fab domains [76, 77, 78]. In total, the glycosylation of IgG accounts for between 2–3% of its molecular weight (reviewed in [79]).

The glycosylation pathway

Mammalian *N*-linked glycosylation occurs as the protein traffics through the endoplasmic reticulum and golgi apparatus of the cell [80, 81]. The well-characterised eukaryotic *N*-linked glycosylation pathway employed to glycosylate IgG is described in Figure 1.9A.

In healthy human adults, as many as 32 different glycoforms have been detected [77, 79, 85]. Human serum IgG glycans are predominantly of the biantennary complex type. Neutral glycoforms account for approximately 81% with the remaining 19% of glycans present as either mono- or disialylated forms (Table 1.2). Glycans present on the Fab domains of IgG more commonly contain a bisecting GlcNAc residue and are proportionately more sialylated when compared to their Fc counterparts [77]. Despite differences in the amount of terminal glycan processing, both IgG Fc and Fab glycans show similar levels of core fucosylation (approximately 91%). Commonly isolated Fc glycoforms are shown in Figure 1.9B, whilst a summary of the proportion of specific glycoforms found on IgG Fc and Fab domains can be found in Table 1.2.

The *N*-linked glycan of IgG Fc forms extensive contacts with the inward facing β-sheet (strands A, B, E, D) of the CH2 domain (Fig. 1.10). Important residues which make hydrophobic contacts with the glycan include Phe241, Phe243, Val262 and Val264. Indeed, the limited glycan processing of the Fc glycans is attributed to these interactions which appear to limit glycan flexibility [79, 86]. Substitution of these residues significantly

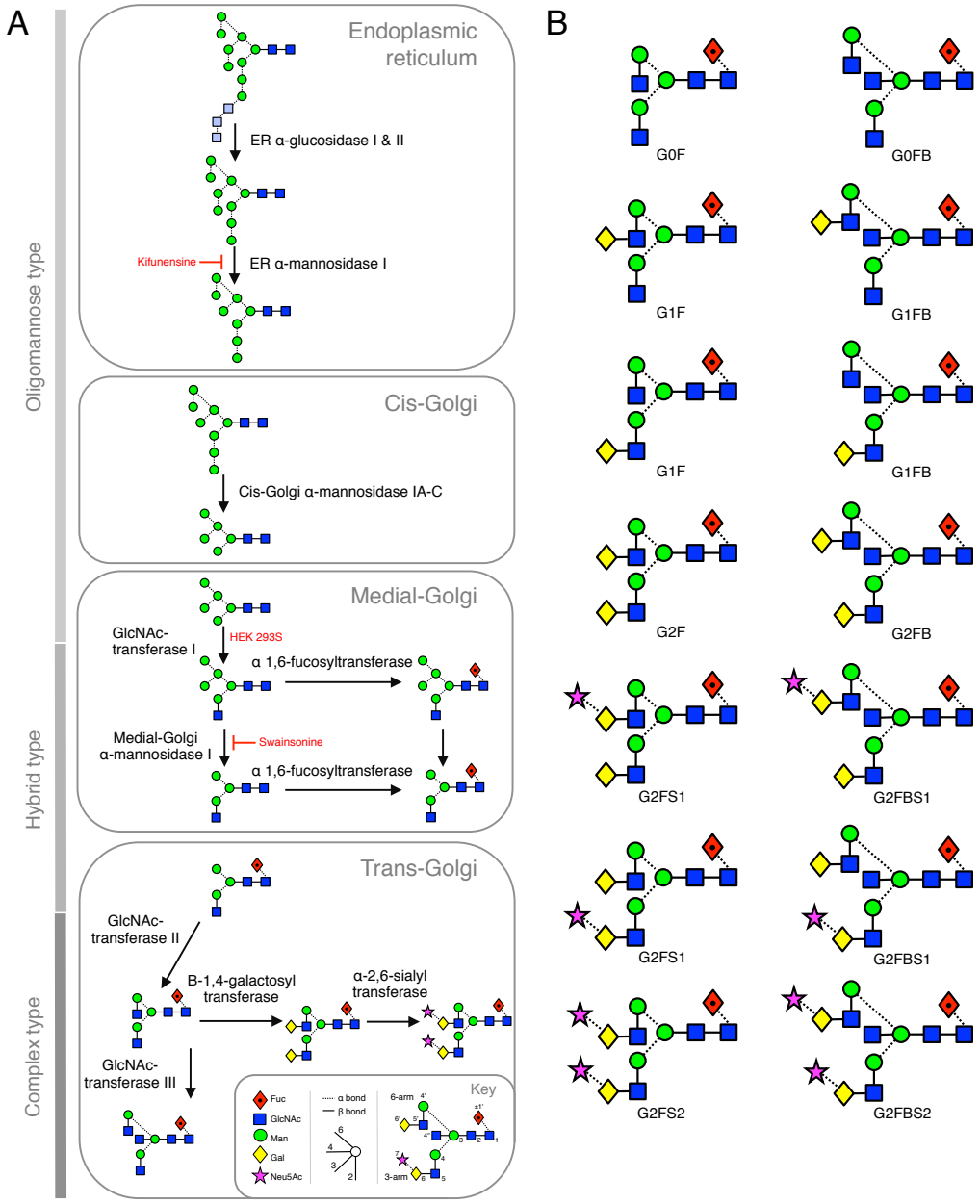


Figure 1.9: The mammalian *N*-linked glycosylation pathway and typical IgG glycans. (A) A schematic representation is shown for key glycan processing steps of the *N*-linked glycan as it passes from the endoplasmic reticulum through various compartments of the golgi. Glycan types are indicated to the left of the panels. Enzymes targeted by the glycan pathway inhibitors kifunensine and swainsonine are indicated in red, as is the GlcNAc-transferase I enzyme which is deficient in HEK 293S cells [82] (B) Complex biantennary glycans commonly found on human serum IgG are shown in schematic representation. G0, agalactosylated; G1, monogalactosylated; G2, digalactosylated; S1, monosialylated; S2, disialylated; F, fucosylated; B, bisected. In both panels glycans are represented following the notation described by Harvey *et al.* [83] with the colour scheme described by the Centre for Functional Glycomics. The key in panel A reveals the notations used to describe linkages and anomericity and shows residue labelling consistent with Vliegenthart *et al.* [84].

Table 1.2: Relative proportions of human serum IgG *N*-glycans*

%	G0	G1	G2	Monosial.	Disial.	Fucosylated	Bisected
IgG	25.5	39.4	17.8	11.6	6.95	91.55	14.5
Fc	26.3	41.85	20.05	8.7	4.35	92.35	11.65
Fab	6.45	7.8	2.8	35	48.85	89.7	43.45

*Values were calculated from data given in Holland *et al.* [77] where the average values for IgG glycans from the two control patients is quoted as a percentage of the total glycan pool. Monosial. and Disial. refer to mono- and disialylated species, respectively.

increases glycan processing towards more branched and sialylated glycans, potentially due to a increased flexibility and thus accessibility of these glycans to golgi processing enzymes [87]. Overall, these glycans contribute strongly to both the structural and functional integrity of IgG Fc.

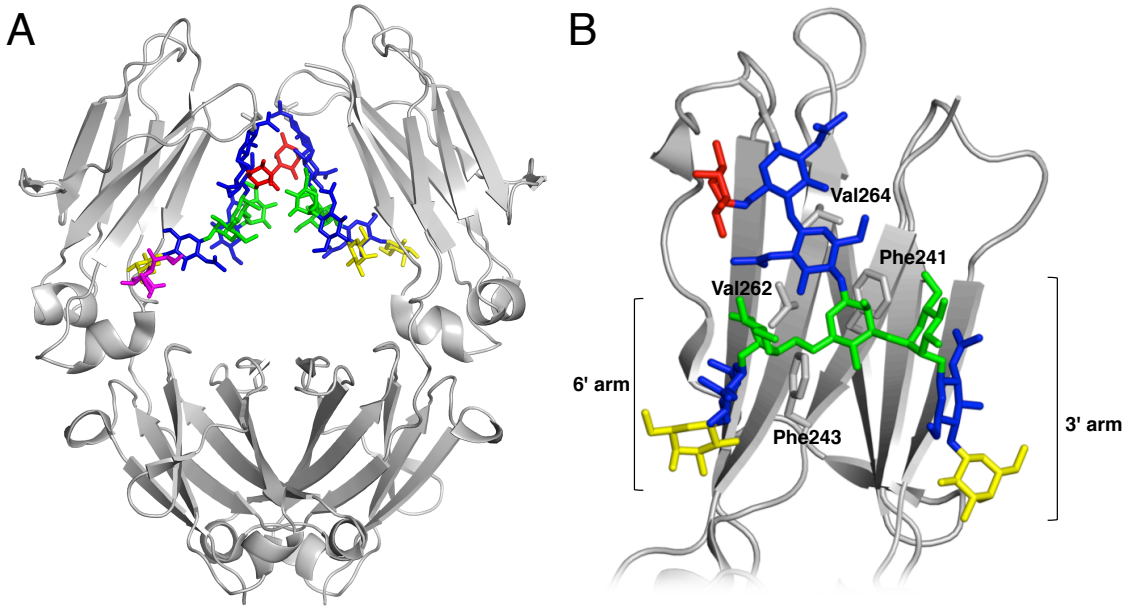


Figure 1.10: The IgG1 CH2-glycan interface. (A) The crystal structure of IgG Fc (PDB ID: 4BYH [88]) shows the protein in cartoon representation (grey) and the glycans (G2F and G1FS1) in stick representation following the colour scheme described by the Centre for Functional Glycomics: Fuc, red; GlcNAc, blue; Man, green; Gal, yellow; Neu5Ac, magenta. (B) A close up of the inner, glycan binding face of the CH2 domain shows the G2F glycan and the hydrophobic ladder comprising Phe241, Phe243, Val262 and Val264 which are shown as sticks.

Functional effects of IgG glycosylation

The importance of the Fc glycan to IgG can be demonstrated by discussing the effects of deglycosylation on Fc structure and function. The effect of deglycosylation on the

structure of Fc has been explored using thermal stability assays [89], X-ray crystallography [90, 91], small angle X-ray scattering (SAXS) [41] and Förster resonance energy transfer (FRET) [92]. In all cases, the loss of the *N*-linked glycan alters the structural properties of the Fc domain.

Using differential scanning microcalorimetry, Mimura *et al.* showed that increasingly truncated glycans at position Asn297 resulted in corresponding decreases in thermal stability of the CH2 domain, consistent with the proposed role of these glycans in conferring structural integrity to the Fc. In addition, the minimal glycan unit required for Fc stability and function was found to be the core trisaccharide $Man_1GlcNAc_2$ [89]. Although crystallographic analyses suggest that loss of the glycan results in a collapse of the two CH2 domains [90, 91], measurements of deglycosylated Fc domains in solution by both SAXS [41] and FRET [92] suggest that deglycosylation renders the Fc domain more flexible and able to access more 'open' CH2 conformations. Further discussions on the structural effects of Fc deglycosylation are explored in Chapter 6.

Critically, the *N*-linked glycan is involved in maintaining the effector functions of IgG Fc. The aglycosylated Asn297Ala mutant loses binding capability to FcγRI, FcγRIIA, FcγRIIB, and FcγRIIIA, presumably through loss of structural elements critical for the interaction [62, 93]. However, specific mutagenesis of Fc residues restores some binding of aglycosylated Fc to selective FcγRs [94, 95, 96]. Deglycosylation of IgG Fc also disrupts effective binding of Fc to C1q showing that Fc glycosylation is also critical for successful activation of the complement pathway [50, 89, 40].

The glycoforms present on serum IgG are not only critical for outright binding of Fc to FcγRs and C1q, but variation in these glycoforms also modulates the binding affinities and selectivities for these binding partners (reviewed in [97]). Different glycosylation states can therefore dictate the type of immune response generated by IgG making glycosylation an important consideration for the design and manufacturing of therapeutic antibodies. Glycosylation of IgG also changes with various disease states [79]. Given this, glycan profiling has been used as a tool for both understanding and diagnosing various diseases with the most notable example being rheumatoid arthritis [98].

1.3 Immune evasion strategies of *Streptococcus pyogenes*

S. pyogenes produces a vast array of immune evasion factors to help it circumvent the immune system. Having briefly discussed the human immune response to GAS in Section 1.1.2, the mechanisms by which GAS attempts to evade these immune responses are described here and are summarised in Figure 1.11. These GAS mechanisms target the host's attempts at inhibiting bacterial adherence as well as various aspects of the complement system and the host mechanisms of phagocytosis and phagocyte recruitment. There are also several secreted GAS proteins which target antibody-mediated immune responses. These include the endo-β-N-acetylglucosaminidase, EndoS and the cysteine protease IdeS. These two enzymes form the focus of this thesis and are discussed individually in Sections 1.3.6 and 1.3.7, respectively.

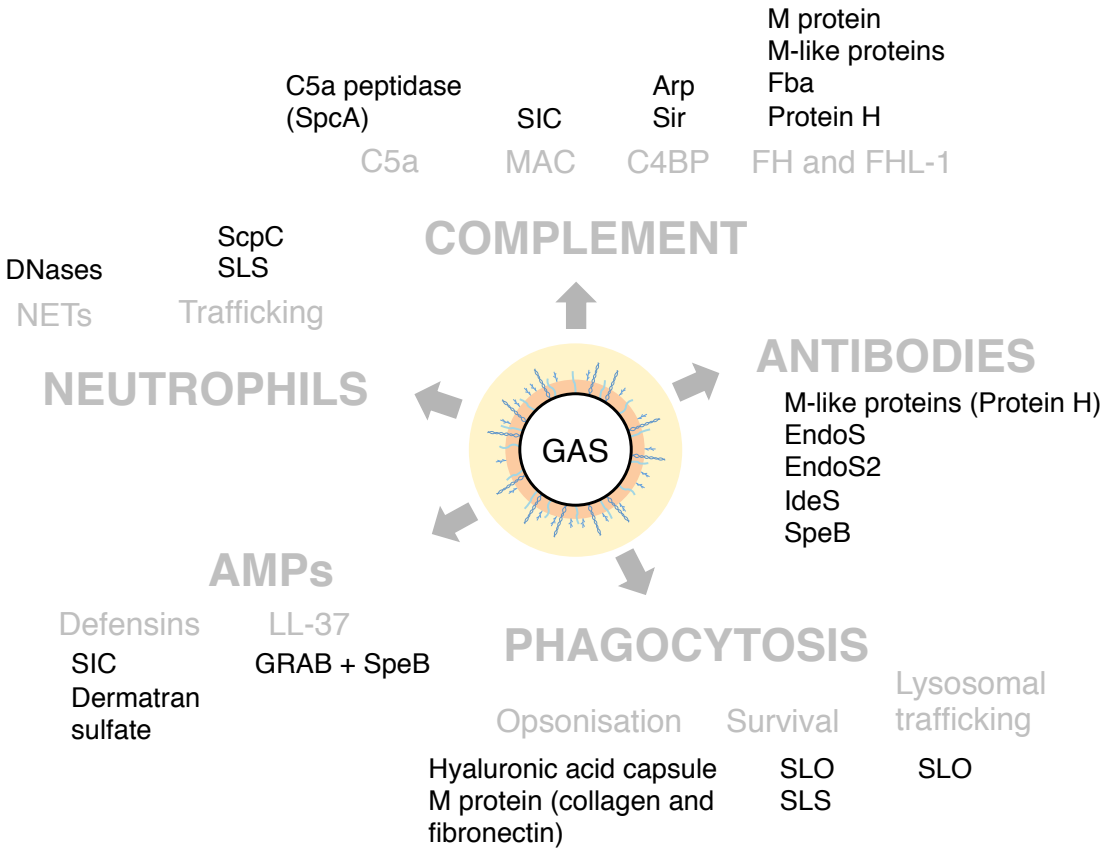


Figure 1.11: Immune evasion strategies of GAS. Key aspects of the host immune response which are targeted by GAS are shown in grey font. The GAS-produced immune evasion factors which interfere with or block these immune responses are shown in black font.

1.3.1 Inhibiting AMPs

As mentioned in Section 1.1.2, an initial element of the innate immune response to GAS is the production of AMPs. Streptococcal inhibitor of complement-mediated lysis (SIC) is a protein which interacts with and blocks the antibacterial activity of AMPs including α-defensin, the β-defensins (HBD1 to 3) and LL-37 [99, 100]. SIC blocks the antimicrobial and enzymatic function of lysozyme and the antimicrobial (but not antiprotease) activity of secretory leukocyte proteinase inhibitor (SLPI) [101]. SIC, as its name suggests, also has a potential role in inhibiting the complement pathway as described later in this section.

GAS further targets LL-37 using an intricate system mediated by a cell surface protein called GRAB. GRAB binds human α_2-macroglobulin (α_2M), localising α_2M to the cell surface of GAS. α_2M functions to bind and imprison proteases to inhibit their activity against other host proteins, and in the case of GAS, α_2M successfully binds and encapsulates streptococcal erythrogenic toxin B (SpeB); a GAS-produced cysteine protease. Although the α_2M-bound SpeB is no longer free in solution to target other proteins, small peptides, such as AMPs, are still able to enter the α_2M-SpeB complex to be proteolysed by SpeB. Essentially, by binding α_2M, GRAB creates an effective shield of proteolytic activity against LL-37 around group A streptococci [102]. The SpeB protease is a multifunctional protein which contributes as a virulence factor of GAS as well as an immune evasion factor [103, 104]. In addition to the protease functions described above, SpeB is able to metabolise host proteoglycans to produce dermatran sulphate which, in turn, inhbitis the antibacterial activity of α-defensin [105].

1.3.2 Altering complement activation

The complement pathway is heavily targeted by the defence mechanisms of GAS [106] where recruited proteins include the host regulatory components factor H (FH) and factor H-like protein 1 (FHL-1). Both FH and FHL-1 bind cell surface-bound C3b and recruit the serine protease factor I. Factor I is then able to cleave C3b to give iC3b resulting in no further activation of complement, preventing downstream opsonisation and phagocytosis. GAS recruitment and binding of FH and FHL-1 is associated with an inhibition of complement activation and prevention of C3b-mediated phagocytosis. FH and FHL-1 binding

and antiphagocytic capability has been attributed to the cell surface M protein of GAS [107] but this has recently been contested [108]. FH binding has also been associated with a more recently identified protein called Fba (fibronectin-binding protein) [109] and the M-like protein from M1 serotypes of GAS, protein H [110].

The M protein of some GAS serotypes is also able to bind human fibrinogen. Fibrinogen binding reduces the deposition of complement on the cell surface of GAS leading to a reduction in complement-mediated phagocytosis [111, 112].

The complement regulatory factor C4b-binding protein (C4BP) is also recruited by GAS. C4BP prevents the assembly of C3- and C5-convertases (C4b2a and C4b2a3b, respectively) on the surface of cells and facilitates the degradation of assembled convertases by displacing C2a. In addition, C4BP is a cofactor for the proteolytic degradation of C4b by factor I [23]. Two cell surface members of the M protein family, Arp and Sir, both bind to and enhance the function of C4BP, preventing activation of the classical complement pathway and again inhibiting phagocytosis of GAS [113, 114].

An additional GAS immune evasion factor, the cell wall-associated C5a peptidase (ScpA), is able to proteolytically inactivate the chemotaxin, C5a, reducing phagocyte recruitment to the site of infection [115].

Another potential immune evasion mechanism of GAS is to target the MAC, which comprises the terminal complement components C5b, C6, C7, C8 and C9. In a series of steps, the MAC associates to form a large pore in the membrane of target cells leading to lysis and cell death. Although the thick peptidoglycan cell-wall of Gram-positive bacteria is generally considered to confer resistance to lysis in this way [116], recent studies indicate that the dynamic cell wall of GAS may not be protective against the MAC. It was shown that C5b67 complexes do localise to the cell surface of Gram-positive bacteria, furthermore, that C5b67 complexes specifically localise to the division septum between dividing GAS cells where the peptidoglycan layer may be thinnest [117]. The defence mechanisms of GAS against the MAC also suggest that MAC-mediated cell lysis may occur in Gram-positive bacteria. The streptococcal SIC protein (described earlier in this section) is able to bind C5b67 complexes preventing formation of the MAC in sensitised erythrocytes, thus inhibiting complement-mediated lysis [118, 119]. However, the role of SIC in preventing potential MAC-mediated lysis of GAS has not been confirmed.

1.3.3 Disrupting neutrophil recruitment and NETs

Neutrophil trafficking and NETs are the target of several GAS immune evasion strategies. Similar to the C5a peptidase mentioned above, the GAS protein ScpC causes a reduction in chemotaxis and neutrophil trafficking by cleaving CXC chemokines, including human IL-8 and mouse CXCL1 and CXCL2 [115]. In addition to C5a peptidase and ScpC, GAS produces streptolysin S (SLS) which also reduces neutrophil recruitment to the site of infection through potential inhibition of chemotactic signals [120]. Lastly, NETs (described in Section 1.1.2), which function to trap and kill microorganisms, are disrupted by GAS production of deoxyribonucleases (DNases). GAS produces a selection of DNAses, some encoded by prophage DNA, which have been implicated in breaking down the chromatin scaffold of NETs allowing GAS to avoid this host mechanism of defence, leading to increased GAS survival [121, 122, 123].

1.3.4 Intracellular survival and phagocytosis evasion

In addition to the strategies used to evade the complement-mediated pathways of phagocytosis (Section 1.3.2), the antiphagocytic mechanisms of GAS can also be attributed to the capsule of the bacteria. The capsule of GAS comprises the glycosaminoglycan hyaluronic acid, which is a polymer of *N*-acetylglucosamine and glucuronic acid. In humans, hyaluronic acid is a major component of the ECM and is found in high concentrations in soft connective tissues [124]. The hyaluronic acid capsule of GAS confers resistance to phagocytosis, where this capsule is thought to mask the bacteria from the host immune system [125]. GAS also recruits collagen fibres and fibronectin to its surface to form large bacterial aggregates which are resistant to antibody mediated phagocytosis [126].

GAS also directly induces both lysis and apoptosis of phagocytes via the production of streptolysins. Streptolysin O and to a lesser extent, streptolysin S (SLS) induce lysis and cell death of phagocytic cells and as a result, protect GAS from phagocytic killing [127, 128, 129]. SLO also contributes to the resistance of phagocytosed GAS to lysosomal killing and is likely involved in the mechanism by which GAS survives intracellularly [130]. Indeed, GAS have been found to internalise successfully into macrophages [17] where they are able to survive and replicate in a mechanism that may be dependent on the M protein [131,

132]. These intracellular reservoirs of GAS help to explain the sometimes poor response to antibiotics during invasive infections where viable GAS have been found inside cells from patient samples from the site of infection up to 20 days after treatment with intravenous antibiotics [133].

1.3.5 Ig targeting proteins

GAS expresses a number of proteins which interact with antibodies. This includes both M and M-like proteins, which bind to the Fc domain of both IgG and IgA [134, 135, 136]. Although the relevance of some of these interactions are yet to be resolved, some show potential anti-phagocytic mechanisms [137]. For instance, the M-like protein, protein H, binds to IgG Fc resulting in a inhibition of complement activation [138].

In addition to these Ig binding proteins, GAS produces a cysteine protease called SpeB which was described previously in Section 1.3.1. Amongst other functions, this largely non-specific protease cleaves the hinge region of IgG, separating the antigen binding Fab domains from the Fc domain. In addition, SpeB degrades all four remaining classes of antibody, IgA, IgD, IgE and IgM [139]. SpeB mediated hydrolysis of IgG lead to a reduction in antibody-mediated phagocytosis suggesting a role for SpeB in evading the adaptive immune response against GAS [140]. Two further immune evasion factors of GAS which interact with IgG include the endoglycosidase, EndoS and the cysteine protease, IdeS, which are discussed in Section 1.3.6 and 1.3.7, respectively.

1.3.6 EndoS

EndoS (endo-β-N-acetylglucosaminidase of streptococci) is an enzyme secreted by GAS which is able to cleave specifically the β-1,4-di-N-acetylchitobiose core of the N-linked glycan residue Asn297 of human IgG [141]. EndoS was first isolated by Collin and Olsén in 2001 from the supernatant of cultures of the serotype M1 GAS, AP1 [141]. In this study, the ability of EndoS to cleave at the chitobiose core of the IgG Fc glycan was identified using a combination of SDS-PAGE and western blots probing for various parts of the IgG N-linked glycan with specific lectins. This glycolytic activity against IgG was then shown to be specific for native IgG, where heat-denatured IgG was resistant to glycan hydrolysis

by EndoS [139]. The ability of EndoS to cleave the chitobiose core of IgG *N*-linked glycans has since been confirmed by a number of other publications using both MS and HPLC based methods [142, 143, 76].

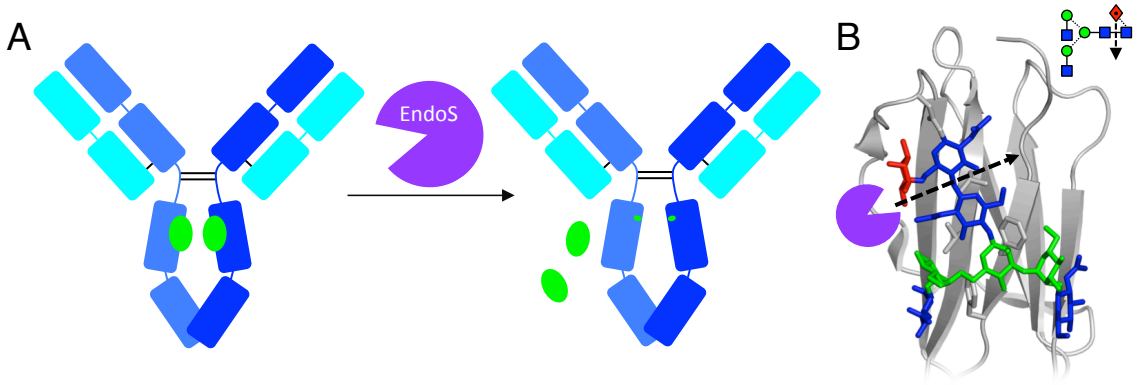


Figure 1.12: EndoS hydrolysis of IgG glycans. (A) A schematic representation of EndoS deglycosylation of IgG1 shows IgG (heavy chains, dark blue; light chains, cyan; glycans, green) being deglycosylated by EndoS (purple). (B) The cleavage site of EndoS (black dotted line) is shown on the glycosylated CH2 domain of IgG1 Fc (PDB ID: 3AVE [46]). This panel shows an identical view to that in Figure 1.10B with an additional schematic diagram of the G0F glycan found in the X-ray crystal structure. The schematic diagram of the glycan follows the notation described by Harvey *et al.* [83] with the colour scheme described by the Centre for Functional Glycomics.

EndoS shows no glycolytic activity against human IgM [144] and has also been reported to have no activity against IgA. However, data supporting this latter statement were not shown in the relevant publication [141]. EndoS has not been tested for its activity against the remaining Ig subclasses, nor has its activity been tested against alternative glycoprotein substrates. Investigations into the glycan specificity of EndoS initially indicated that EndoS was not able to cleave glycans with a bisecting GlcNAc from human IgG [76]. However, EndoS has since been shown to be capable of hydrolysing these glycoforms from Fc in addition to the remaining complex biantennary type glycans (both fucosylated and afucosylated) commonly found on serum IgG Fc [143]. EndoS activity has also been tested against PNGaseF-released *N*-glycans free in solution. In these experiments, neither oligomannose type glycans nor hybrid type glycans could be digested by EndoS. Futhermore, from a pool of *N*-linked glycans released from the human immunodeficiency virus 1 protein gp120, EndoS selectively hydrolysed complex biantennary type glycans found on IgG, but did not show any activity against glycans with a bisecting GlcNAc in this context. Altogether, these data suggest that protein–protein interactions as well as protein–glycan interactions are contributing to the specificity and activity of EndoS.

IgG Fc effector functions are strongly influenced by the *N*-linked glycan at Asn297, including FcγR binding and activation of complement (discussed in Section 1.2.3). In line with this, deglycosylation of IgG by EndoS causes a reduction in both FcγR binding and complement activation via the classical pathway. No effect of EndoS on the activation of complement via the MBL or alternative pathways has been observed. Importantly, the glycolytic activity of EndoS was able to inhibit IgG-mediated opsonophagocytosis *in vitro* of GAS, resulting in increased survival of GAS in human blood. Together, these data suggest that EndoS contributes to the immune-evasion strategy of GAS by targeting the adaptive immune response [140]. Notably, subsequent investigation of the ability of EndoS to disrupt FcγR binding revealed that this mechanism was somewhat IgG subclass dependent. Although all murine and human subclasses of IgG are effectively deglycosylated, human IgG2 as well as mouse IgG2a appear resistant to the *effects* of EndoS-mediated IgG deglycosylation [145, 144]. Finally, although EndoS shows antiphagocytic activity in both neutrophil- and monocyte-killing assays, it may not significantly contribute to the virulence of GAS *in vivo*[146].

EndoS is a 110 kDa protein which shows significant homology in its N-terminal region to endo-β-*N*-acetylglucosaminidase F2 (EndoF2) from *Flavobacterium meningosepticum* with further sequence analysis of EndoS confirming the presence of a chitinase family 18 active site between residues 227–235 [141]. Subsequent homology modelling of EndoS gave further support for the presence of a family 18 chitinase domain in the first 446 residues of EndoS [147]. Downstream of the proposed catalytic domain, the EndoS sequence shows homology to leucine-rich repeat (LRR) proteins called internalins [147]. LRRs are structurally well characterised and are commonly involved in protein–protein interactions (for a review see [148]). The C-terminal region of EndoS remained largely uncharacterised by the homology modelling carried out by Allhorn *et al.* [147] and is a point of discussion and research in this thesis (Chapters 3 and 4). Nevertheless, in addition to identifying homologous domains in EndoS, Allhorn *et al.* also identified a number of residues which are important for EndoS activity: Glu235 is crucial for EndoS catalytic activity and a glutamate to glutamine mutation at this residue inhibits EndoS activity against IgG. In the same study, the 12 tryptophan residues of EndoS, but not the lysine or cysteine residues, were also shown to be critical for the function of EndoS. Finally, the cysteine protease SpeB was shown to proteolytically inactivate EndoS by cleaving between residues Lys445 and

Lys446, suggesting a potential mechanism for regulation of EndoS activity during GAS infection [147].

During the course of this project, an X-ray crystal structure of EndoS (residues 90–995) was published [149] revealing key structural features of the enzyme. Given the relevance of an X-ray crystal structure to the work described in this thesis, a full description of this structure is given in Chapter 3 in reference to the structural data collected here.

Recently, a protein homologous to EndoS from GAS serotype M49 was described which also exhibits endoglycosidase activity against IgG and can inhibit IgG-mediated phagocytosis. The so called EndoS2 enzyme also showed additional endoglycosidase activity against other substrates and differs significantly in the C-terminal region compared to EndoS [150].

1.3.7 IdeS

IdeS is a cysteine protease capable of specifically cleaving the flexible hinge region of IgG. The discovery of IdeS is somewhat confused by the fact that three groups, each independent of the others, described the discovery and characterisation of this protein between the years 2000 and 2002. To begin, this protein was originally identified through proteomic analysis of immunogenic proteins from GAS cell culture supernatants and due to its sequence homology to the α subunit of the human Mac-1 protein (CD11b/CD18) was designated the name Mac [151]. Lei *et al.* furthered their studies from the proteomic identification of Mac, showing that Mac was able to inhibit phagocytosis and killing of GAS *in vitro* potentially by mechanisms involving binding to CD11b and FcγRIIIA [152]. The second group identified and named the same protein IdeS, since the protein was recognised as a GAS secreted cysteine protease capable of digesting IgG [153]. Finally, Kawabata *et al.* identified this protein as an unanchored surface protein from GAS capable of binding IgG, thus naming it streptococcal immunoglobulin-binding protein 35 (Sib35) [154]. Therefore, in the literature, Mac (Mac-1 and Mac-2), IdeS, Mac/IdeS and Sib32 all refer to the same protein. In the remainder of this thesis, this protein will be discussed using the name IdeS.

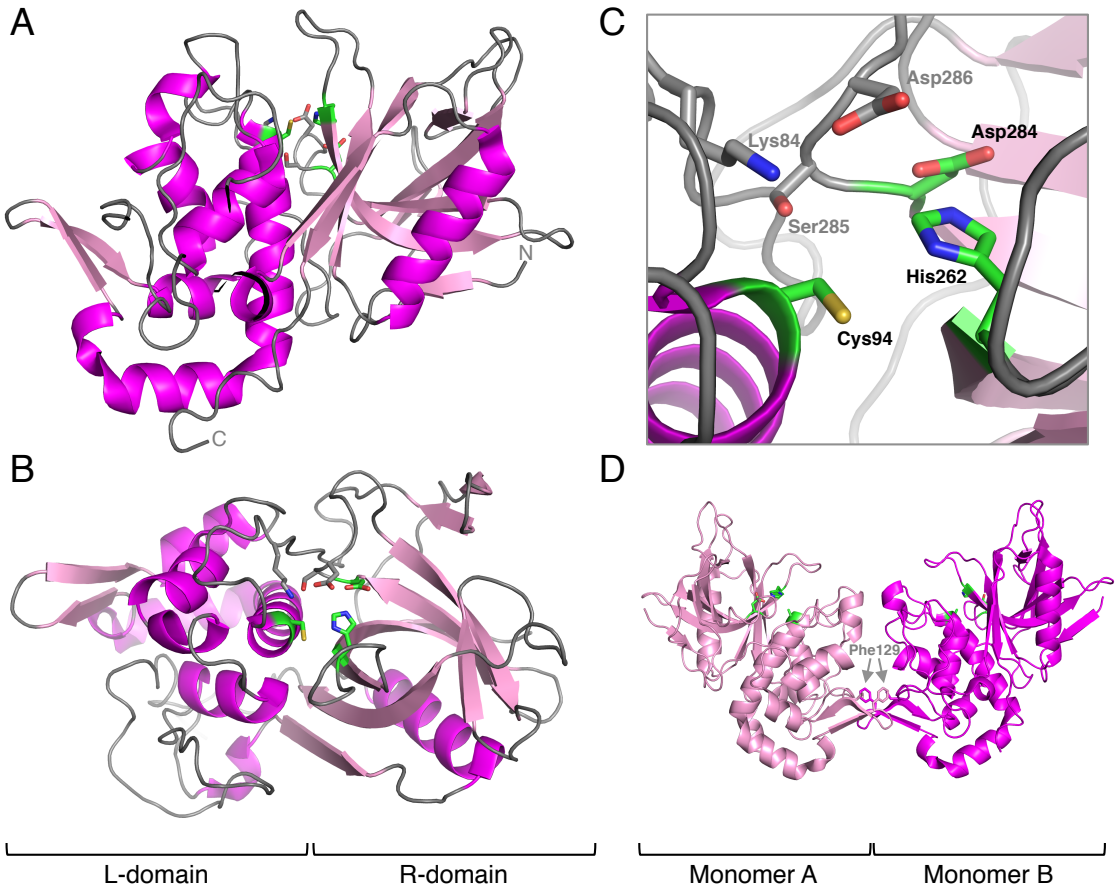


Figure 1.13: The crystal structure of IdeS. (A) A single monomer from the asymmetric unit of the crystal structure of IdeS is shown in cartoon representation (PDB ID: 2AU1 [155]). The secondary structure is coloured to show α-helices in magenta, β-sheets in pink and unstructured regions in grey. The catalytic triad residues Cys94, His262 and Asp284 are shown in stick representation in green with oxygen atoms in red, nitrogen atoms in blue and sulphur atoms in yellow. (B) The monomer from A was rotated through 90° to give a top-down view of the active site. (C) A close up of the active site reveals the catalytic triad residues (green) and the position of other closely packed side-chains important for catalytic activity. (D) The crystallographic dimer observed by Agniswamy *et al.* shows two monomers from the asymmetric unit, one in pink and one in magenta. The Phe129 residues implicated in potential dimerisation are shown in stick representation.

IdeS structure

IdeS is a 38 kDa protein with a secretion signal sequence from amino acids 1–29. The signal sequence is cleaved upon secretion to leave a soluble 35 kDa protein [152, 153]. Three crystal structures of IdeS have been published including the inactive IdeS C94S mutant (PDB ID: 1Y08, [156]), the inactive C94A mutant (PDB ID: 2AVW, [155]) and finally, the catalytically active IdeS protein (PDB ID: 2AU1 [155]). In all cases, the structure reveals a globular protein belonging to the papain superfamily (Fig. 1.13). IdeS comprises two globular domains, one with primarily α-helical structure (the L-domain) and the other comprising mostly β-sheet secondary structure (R-domain).

The active site is found in a pocket at the interface between the two domains where the critical residues have been identified through both structural and mutational analysis. Cys94, His262 and Asp 284 form the catalytic triad with contributions from Asp286 required for full activity. Lys84 in combination with the amide nitrogen of Cys94 form the oxyanion hole of the enzyme [157, 156, 155].

Although dimerisation of IdeS has been implicated by X-ray crystallographic (Fig. 1.13D) [155] and kinetic analyses [158], further evidence suggests it is more likely to function as a monomer [159]. Having observed a crystallographic dimer, Agniswamy *et al.* made mutations of F129A and M134A at the apparent IdeS dimer interface. These caused the protein to form inclusion bodies, at which point the more conservative mutations of F129I and M143L were designed to disrupt the apparent dimer. Both F129I and M143L were minimally soluble suggesting, together with the previous alanine mutagenesis, that these residues are likely important for the overall integrity of the protein. Both the F129I and M143L mutations reduced the amount of apparent dimer IdeS populations by SEC and although the F129I mutation negatively affected the cysteine protease activity of IdeS, the M143L mutation, somewhat counterintuitively, caused an increase in kinetic activity [155]. Furthermore, the affinity of the apparent dimer as calculated by analytical ultracentrifugation gave a low Kd of 250 μM [155]. In the monomeric structure by Wenig *et al.* the residues of the apparent dimeric interface of IdeS did not give defined density and could not be modelled, suggesting potential flexibility of these residues in solution. In addition, the reequilibration of the dimer-monomer mixture of IdeS when the IdeS monomer peak from a SEC was reloaded onto a SEC column, as observed by Agniswamy

et al., was not observed in subsequent publications and indeed, was not observed during this work [159].

IdeS substrate specificity

IdeS exists as two main isoforms across different GAS serotypes. The two isoforms (commonly referred to as allelic variants) share approximately 85% sequence similarity across the entire protein sequence where the differences between variants are concentrated to the middle third of the protein ($\sim$ 110–202), which exhibits sequence identity as low as 50% [160, 154]. Nevertheless, both variants maintain IgG protease activity, with the exception of IdeS from serotype M28, which loses activity due to an additional Cys residue at position 257 which disulphide bonds to the active site Cys94 [160, 161]. IdeS specifically cleaves between Gly236 and Gly237 of the IgG hinge, one heavy chain at a time, to first give single cleaved IgG (scIgG) and subsequently, $F(ab)_2$ and the Fc separated into its two monomeric components (Fig. 1.14A) [158, 162].

Other human Ig isotypes (IgA, IgD, IgE and IgM) are not proteolysed by IdeS but all IgG subclasses are susceptible to cleavage by IdeS. IgG2 is more resistant to IdeS cleavage, presumably due to the unique sequence of the lower hinge region (Fig. 1.14B)[163]. IdeS is unable to cleave peptides identical to the lower hinge region of IgG, suggesting a role for exosite binding to IgG in IdeS activity [158]. In addition, IdeS retains the ability to cleave the hinge region of Fc produced through digestion of IgG with papain, suggesting the Fab domains of IgG are dispensable for IdeS activity. IdeS also retains activity against reduced, monomeric Fc fragments as well as deglycosylated Fc fragments suggesting that dimerisation and glycosylation of Fc are not required for IdeS activity [158].

By cleaving IgG and separating the Fab domains, which bind antigen from the Fc domain, which ellicits effector functions, IdeS is able to reduce IgG-mediated opsonophagocytosis of GAS [153, 152, 160]. In addition, complete cleavage of IgG and cleavage of only a single IgG heavy chain, yielding scIgG, are both sufficient to disrupt ADCC and complement dependent cytotoxicity (CDC) *in vitro* and *in vivo* [164]. The reduced ability of scIgG to instigate ADCC and CDC was attributed to its reduced affinity for FcγR receptors compared to intact IgG [164].

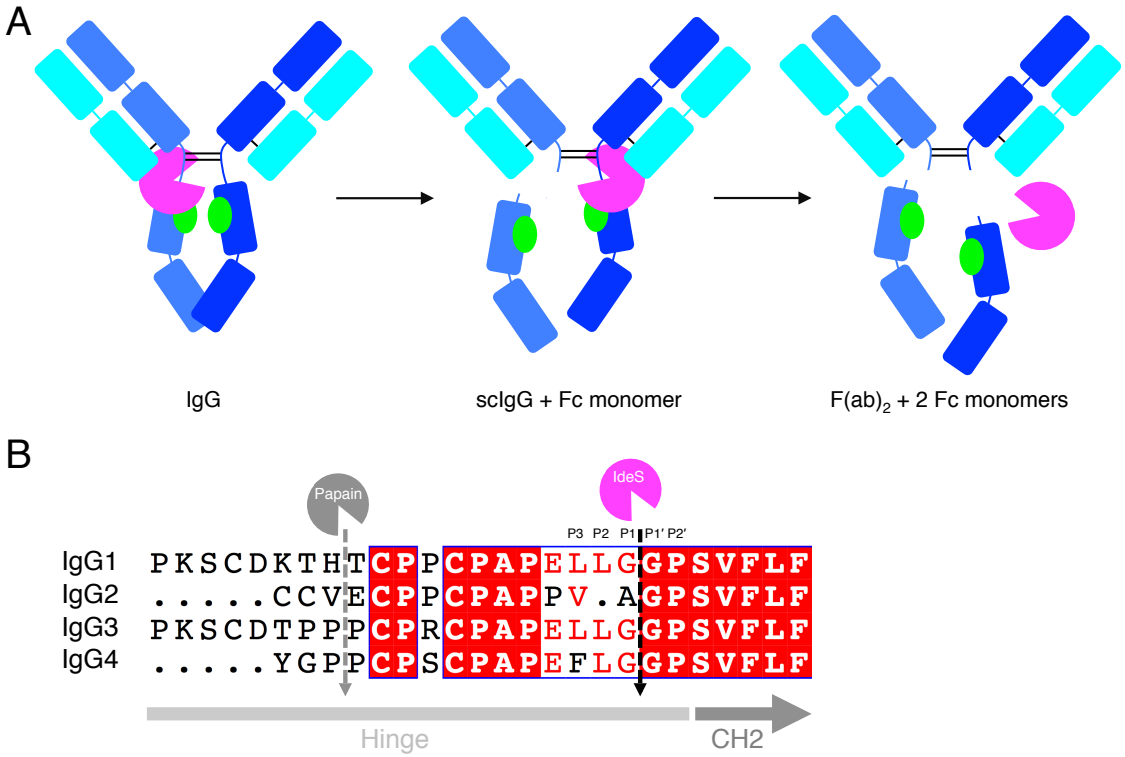


Figure 1.14: Proteolytic cleavage of IgG by IdeS. (A) A schematic representation of the sequential cleavage of IgG by IdeS shows formation of scIgG followed by $F(ab)_2$ and 2 Fc monomers. (B) An alignment of the lower hinge region of human IgG1 (Uniprot ID: P01857), IgG2 (P01859), IgG3 (P01860) and IgG4 (P01861) was made using ESPript 3.0 [28] from a multiple alignment file generated using Clustal Omega [29]. Conserved residues are shown in white text and are highlighted in red. Non-conserved residues are shown in black text on a white background. Residues in red text indicate conservative changes or partial conservation across different subclasses. The IdeS cleavage site is indicated by a black dotted line with P3, P2, P1, P1′ and P2′ residues of the substrate labelled accordingly. The cut site for papain is indicated by a grey dotted line.

Alternative or additional functions of IdeS

In addition to its role as an IgG protease, other potential functions of IdeS have been suggested to contribute to its antiphagocytic activity. This is a controversial subject where in some instances, the proteolytic activity of IdeS was critical in conferring antiphagocytic activity [161, 165] whereas in others, even a catalytically inactive mutant was able to inhibit IgG-mediated phagocytosis [160]. Potential alternative mechanisms draw on the observed binding of IdeS to FcγRIIIB and IdeS sequence homology of IdeS to human Mac-1 (CD11b/CD18), a surface integrin receptor, which shows potential interactions with FcγRIIIB [152]. Although sequence homology to human Mac-1 was observed by Lei *et al.*, the subsequent crystal structures of IdeS [155, 156] show that this homology does not translate to secondary or tertiary structure of IdeS, making this feature of IdeS less likely to be of any importance. In addition, the majority of the observations made by Lei *et al.* attributed to interactions with CD11b and FcγRIIIB can be explained by the IgG protease activity of IdeS. Nevertheless, surface plasmon resonance analysis showed that IdeS from serotype M28, but not serotype M1, bound to both FcγRII and FcγRIII (class not specified) competing with IgG Fc for binding [163]. Therefore, blocking FcγRs from binding to Fc may yet be an additional mechanism by which IdeS inhibits phagocytosis.

IdeS also has an RGD motif between amino acids 214–216. These motifs are identifiable as cell-adhesion motifs which have the ability to bind human integrins (for a review see [166]). IdeS has been shown to bind to cells expressing various integrins, where antibodies against IdeS, integrins or simply an RGD peptide are able to mitigate IdeS binding [160]. The RGD motif is located on the opposite side of the protein to the active site, and may be responsible for the ability of IdeS to mediate binding to neutrophils. Although the RGD motif may not contribute to the antiphagocytic activity of IdeS, it may facilitate an alternative, as yet undescribed function of IdeS, perhaps serving a similar function to the SpeB:GRAB complex by concentrating IdeS protease activity to the surface of GAS cells. Given the multifunctional properties of other GAS immune evasion factors discussed earlier in this section (e.g. M protein, SIC, SpeB), it is reasonable to hypothesise that IdeS may have additional mechanisms of action which help GAS evade various immune responses.

1.4 Exploiting immune evasion factors for therapeutic use

Bacterial or pathogen derived proteins are becoming increasingly exploited and researched for therapeutic purposes. For instance, the small (14 kDa) protein azurin and an azurin-like protein called Laz from *Pseudomonas aeruginosa* have both shown potential as anticancer drugs, inhibitors of HIV-1 growth and inhibitors of the growth of the malarial parasite *Plasmodium falciparum* [167, 168, 169]. Considering the necessity for pathogens to evade and modulate the immune system during infection, the majority of other pathogen derived proteins with therapeutic potential often show the most promise for the treatment of various immunological diseases. This topic is the subject of a number of recent reviews [170, 171, 172, 173, 174] and is of particular relevance to this thesis. Due to their ability to modulate IgG effector functions, both EndoS and IdeS from GAS have shown promise in a number of therapeutic applications resulting in significant interest in these two streptococcal immune evasion factors.

1.4.1 Autoimmune disease

Autoimmunity occurs when the immune system elicits a response as a result of stimulation by self-antigens. Although many autoimmune diseases are considered to be mediated by cellular T-lymphocyte mediated responses, it is becoming increasingly apparent that B-lymphocytes and autoantibodies play a central role in the pathology of many of these diseases [175]. The current treatment regime for these diseases includes reducing the levels of circulating pathogenic antibodies by means of plasmapheresis, corticosteroids or immunosuppressant drugs [176], reducing antibody production using monoclonal antibodies against B-cells [177] or by removing the spleen (splenectomy) to reduce unwanted phagocytic clearance of cells [178]. Given the contribution of IgG to autoimmunity, the use of the IgG targeting enzymes EndoS and IdeS, has been explored as an alternative or additional method for clearing pathogenic IgG in the treatment of autoimmune disease [179]. Both EndoS and IdeS have shown success in the treatment of various mouse models of autoimmunity [180, 181].

EndoS was first implicated as a treatment for autoimmune disease using a mouse model of rheumatoid arthritis. Here, an acute form of rheumatoid arthritis was induced in mice

using anti-collagen II (CII) antibodies. Deglycosylation of anti-CII antibodies prior to administration, rendered these antibodies significantly less arthrogenic, effectively abrogating the development of rheumatoid arthritis in these mice [182]. Since this initial study which explored EndoS as a treatment for rheumatoid arthritis, EndoS has been shown to successfully treat various mouse models of immune disease including immune thrombocytopenic pupura (ITP) [183], autoimmune haemolytic anemia [184], glomerulonephritis [185], epidermolysis bullosa acquisita [186] and neuromyelitits optica [187]. In addition, EndoS has shown potential as a treatment for systemic lupus erythematosus (SLE) where it was able to abolish the proinflammatory properties of immune complexes derived from SLE patients *in vitro* [188].

Of note are the IgG subclass specific effects of EndoS observed by Albert *et al.* in the mouse model of ITP. ITP is caused by autoantibodies which react with platelets in the blood resulting in the characteristic phenotypes of a low platelet count and mucocutaneous bleeding [189]. In the ITP mouse model, the effect of anti-platelet IgG antibodies could be significantly reduced when using IgG1 and IgG2b to stimulate pathogenesis, whereas IgG2a, although successfully deglycosylated, maintained the capacity to induce ITP disease through FcγR binding functions. The structural properties which may confer this unique behaviour to mouse IgG2a are investigated further in Chapter 6.

IdeS has also shown promise as a possible treatment for various autoimmune diseases. Much like EndoS, IdeS was successful in the prevention of rheumatoid arthritis [190], ITP [181], glomerulonephritis [185] and neuromyelitits optica [191] when assessed using mouse models of these autoimmune diseases. Importantly, in the mouse model for rheumatoid arthritis, IdeS showed no effect on disease progression when IdeS-resistant antibodies were used to instigate the disease. This finding gives support for the importance of the proteolytic activity of IdeS when used in these applications [190].

Both EndoS and IdeS are well tolerated by both mice and rabbits during *in vivo* treatment [192, 183]. Despite the generation of anti-EndoS and anti-IdeS antibodies during treatment, no adverse effects of EndoS or IdeS treatment in these animals has yet been reported [183, 181]. However, a thorough characterisation of the specificity of these enzymes is desirable and necessary given the potential of these enzymes for therapeutic use. The specificity of IdeS has been well researched and no activity has been observed against

a vast array of other human proteins including alternative Ig isotypes [158]. EndoS has not been as extensively characterised, but has been reported to be inactive against IgM [144] and IgA [141]. The characterisation of EndoS specificity is therefore investigated in Chapter 5.

1.4.2 Organ transplantation

IdeS has gained traction as a possible immunosuppressant for organ transplant patients. Here, IdeS can be given prophylactically with an aim to reduce organ rejection in sensitised patients by eliminating cross-reactive serum IgG. Phase I trials of IdeS in this setting have been completed with Phase II trials due to finish in March 2015. An initial press report indicates that a successful kidney transplant following IdeS pretreatment has been completed, indicating promise for IdeS as a tool for desensitisation of transplant organ recipients [193].

1.4.3 Receptor refocusing

The concept of receptor refocusing using EndoS was first described by Baruah *et al.* and involves using EndoS to reduce levels of serum IgG when treating with therapeutic antibodies [142]. This idea was explored with an aim to reduce the effective concentration of therapeutic antibodies required to achieve FcγR binding in a background of competing serum IgG. Essentially, high concentrations of serum IgG in the blood can compete with therapeutic antibodies for FcγRs during treatment, requiring high doses of therapeutic antibody to elicit an effect. To reduce the required therapeutic antibody to achieve a therapeutic response, it was proposed that EndoS could be used to eliminate competing levels of serum IgG.

As a control, a competitive assay testing therapeutic antibody binding to FcγRIIIA in a background of serum was carried out to show that serum IgG indeed inhibits therapeutic antibody binding to FcγRIIIA. To overcome this effect, the therapeutic antibody was expressed as an apparently EndoS-resistant $Man_5GlcNAc_2$ glycoform. The serum and EndoS-resistant therapeutic antibody were then pretreated with EndoS prior to the assay. In this circumstance, serum IgG competition was significantly reduced such that binding

of the therapeutic antibody was restored to levels near those observed in the absence of serum.

Although the $Man_5GlcNAc_2$ glycoform was sufficiently resistant to EndoS treatment in these assay conditions, experiments conducted during the course of this thesis reveal this glycoform to be susceptible to EndoS cleavage. In order to further pursue this receptor refocusing strategy, structural studies of EndoS alone and in complex with IgG Fc were pursued, as were investigations into glycan specificity. This was done in an effort to identify important features of the deglycosylation reaction with an aim to generate a completely EndoS-resistant IgG antibody.

1.4.4 Chemoenzymatic engineering

Considering the effect that the glycosylation state can have on Fc effector functions (Section 1.2.3), EndoS has been pursued as an IgG specific glycosyltransferase. Engineering glycosidases into glycosyltransferases has been explored with success in the past. For example, the endo-β-*N*-acetylglucosaminidase from *Streptococcus pneumonia*, EndoD, was engineered successfully into a transglycosidase enzyme by way of a single point mutation of an asparagine residue, such that it could be used to glycosylate IgG Fc with a specific glycoform [194].

In 2012, Goodfellow *et al.* showed that by using an oxazoline substrate, EndoS could function as a transglycosidase. Here, EndoS was used to synthesise a 95% pure $Man_3GlcNAc_2Fuc$ glycoform of human IgG [143], however, alternative native IgG glycoforms were not produced. Later in 2012, Huang *et al.* described the mutagenesis of active site residues of EndoS required to convert the enzyme to a glycosyltransferase. Point mutation of Asp233 to either a glutamine or alanine residue was sufficient to make this transition. The authors were able to show successful transglycosylation of IgG to give specific and homogenous glycoforms.

Although use of EndoS to generate IgG glycoforms for therapeutic use has not yet been reported, EndoS has successfully been used as a transglycosidase to derive an α-2,6-sialylated glycoform of Fc for crystallisation experiments [44]. Here, crystals of the engineered glycoform were able to provide information on the relationship between the terminal galactose

and sialyic acid residues of the Fc glycan with respect to the CH2 domain [44].

1.5 Aims

Considering the therapeutic potential of both EndoS and IdeS, this project aimed to further characterise these enzymes and their interactions with IgG both structurally and functionally. A better understanding of these interactions was pursued with a view to exploit new information for therapeutic purposes.

Given the lack of any atomic resolution structural information on EndoS, structural studies of this enzyme were explored with an aim to better understand the functional relevance of different regions of the protein. Experiments addressing this aim are described in Chapters 3 and 4.

Although EndoS is described as an IgG specific enzyme, explicit experiments addressing this property have not been described in detail in the literature. Furthermore, a systematic study of which IgG Fc glycoforms are susceptible to EndoS cleavage has not been carried out. Experiments which address these points are presented in Chapter 5.

Experiments which aimed to understand the properties of mouse IgG2a Fc which allow it to maintain FcγR binding capabilities even when deglycoyslated with EndoS are described in Chapter 6. Further experiments in Chapter 6 also address how EndoS deglycosylation of IgG Fc in general affects the structure of the Fc domain.

Finally, this project aimed to determine a structure of IdeS in complex with its substrate, IgG Fc, in order to appreciate residues pertinent to the specificity this interaction. Experiments which approach this aim are presented in Chapter 7.

2

Materials and methods

‘

All solutions and media were prepared with sterile 18.2 MΩ water from a Milli-Q Plus water purification device (Millipore). Solutions were 0.22 μm vacuum-filtered using nitrocellulose membranes (Merck Millipore) unless otherwise stated.

2.1 Bioinformatics

Using the online server for the programme FUGUE [195], the protein sequence of *S. pyogenes* serotype M1 EndoS (GenBank accession number AF296340) was analysed for regions of homologous protein structure. The full-length sequence (residues 1–995) was submitted to the server as well as the following fragments; residues 446–995, residues 730–995 and residues 927–995.

To assess regions of disorder, the full-length sequence of EndoS was submitted to the Regional Order Neural Network (RONN) online server [196]. Here, regions of disorder are detected at the amino-acid level by way of a pattern recognition algorithm which utilises a modified version of the Bio-Basis Function Neural Network (BBFNN)[197, 198].

2.2 Molecular cloning and mutagenesis

In this section, general molecular biology methods are described, followed by subsections detailing the design and synthesis of specific constructs used throughout this thesis. Mutagenesis reactions are described at the end of this section.

All deoxyribonucleic acid (DNA) primers were purchased from Sigma. Enzymes were purchased from New England Biosciences (NEB) unless otherwise stated. Purification of plasmid DNA from *Eschericia coli* cells was performed using the QIAprep Spin Miniprep Kit (Qiagen).

Agarose gels for analysis of DNA were made using 1% w/v agarose in 40 mM Tris, 20 mM acetic acid and 1 mM ethylenediaminetetraacetic acid (EDTA). Gels were typically run at 110 V for 45 minutes. Purification of DNA from agarose gels was carried out using the NucleoSpin® Gel and PCR Clean-up kit (Clontech).

All polymerase chain reaction (PCR) products were run on agarose gels, excised and purified prior to restriction digests. After restriction digest reactions, PCR products were purified using the NucleoSpin® Gel and PCR Clean-up kit (Clontech). Linearised vectors were also run on agarose gels where bands corresponding to fully digested vector were excised and purified prior to ligation or In-Fusion® cloning.

DNA sequencing was carried out by Source BioScience, Oxford.

2.2.1 PCR

PCR was carried out using Q5® High-Fidelity DNA Polymerase 2X Master Mix (NEB). Forward and reverse primers were added to a give a final concentration of 0.5 μM. A total of 100 ng template DNA was used per reaction. Details of the reaction itself are shown in Table 2.1.

Table 2.1: PCR protocol

Cycle	Temp. (°C)	Time
Cycle 1	98	30 s
Cycle 2 (× 30)	98	10 s
	60	30 s
	72	4 min
Cycle 3	72	2 min

2.2.2 In-Fusion® cloning protocol

A number of the constructs described in this section were made using the ClonTech In-Fusion® cloning protocol. This system relies on Gibson Assembly [199] to insert a gene of interest into a linearised plasmid vector. All inserts were amplified by PCR from the relevant template DNA, with the exception of the IgG1 CH2-H and CH2 constructs. These inserts were synthesised as linear DNA fragments as detailed in Section 2.2.5. All inserts were designed to contain 15 base pairs (bp) at both the 3′ and 5′ end which have complementarity to the terminal 15 bp of the appropriately digested destination vector. In all cases, vectors were digested with the relevant restriction enzymes to linearise the plasmid DNA prior to Gibson Assembly of the insert into the plasmid as detailed below.

A 1:1 mass ratio of vector to insert was used in all instances. The total mass of DNA used for each reaction varied between 100–200 ng. Insert and vector were combined before water was added to give a final reaction volume of 10 μL. The entire reaction volume was used to resuspend the lyophilised components of the In-Fusion® HD EcoDry Cloning Plus Kit (ClonTech). The reaction was incubated for 37 °C for 15 min then 50 °C for 15 min.

A 2.5 μL volume of the resultant reaction was then used to transform Stellar™ competent cells (ClonTech) as per the manufacturer's instructions. Transformed *E. coli* were selected for by carbenicillin resistance after growth overnight at 37 °C. Individual colonies were picked and grown overnight in liquid Miller lysogeny broth (LB) [200] prior to plasmid DNA extraction. Positive clones were checked for correct insertion of the desired gene by DNA sequencing.

2.2.3 EndoS and EndoS fragments

An N-terminally glutathione S-transferase (GST) tagged construct of the full-length EndoS protein was kindly provided by Prof. Ben Davis of the Chemistry Research Laboratory, University of Oxford [143]. The vector (pGEX-4T-1, GE Healthcare) contains the *S. pyogenes* serotype M1 *ndoS* nucleotide sequence (GenBank accession number AF296340) which has been codon-optimised for *E. coli* expression.

The pGEX-4T-1-*ndoS* vector was used as a template for generating various EndoS constructs using primers detailed in Table 2.2. Chit-LRR (residues 1–760), CBM (residues 761–924), CBM-CT (residues 761–995) and CT (residues 924–995) were amplified by PCR to be cloned into bacterial expression vectors. The CBM-KO construct was generated via overlap PCR to remove residues 761–924. For this construct, the products of two reactions — one using primers pGEX EndoS F and Overlap 760 R, the other using Overlap 927 F and pGEX EndoS R — were combined. Using this as template DNA, a second reaction was then run using primers pGEX EndoS F and R resulting in the truncated CBM-KO construct.

Constructs were cloned into the GST fusion vector pGEX-4T-1 (GE-Healthcare) using 5′ BamHI and 5′ NotI restriction sites and into the C-terminal hexahistidine tag vector Champion^{TM} pET303 (Invitrogen) using 3′ XbaI and 5′ XhoI restriction sites. In addition, a construct of the full-length EndoS was amplified by PCR to be cloned into the Champion^{TM} pET303 (Invitrogen) using 3′ XbaI and 5′ XhoI restriction sites.

For all constructs, the amplified PCR product was digested with the appropriate restriction enzymes, as was the destination vector. Constructs were then ligated into their respective vectors using T4 Ligase. After ligation, the plasmids were transformed into DH5-α *E. coli* cells (NEB). Successfully transformed cells were selected for by carbenicillin resistance. Positive colonies were grown overnight in Miller LB before plasmid DNA extraction. Successful ligation and gene insertion were verified by sequencing.

2.2.4 IdeS

IdeS nucleotide sequences from *S. pyogenes* serotype M1 (GenBank accession number AAK33786.1) and serotype M59 (GenBank accession number AFC66043.1) were codon-

Table 2.2: EndoS cloning primers

Primer	Sequence (5′ – 3′)
pGEX EndoS F	GCGTGGATCCATGGACAAAC
5′ pGEX EndoS R	CGATGCGGCCGCTCATTTTTTCAG
pGEX EndoS 761 F	TATCGGATCCGAAAAAACCATGATGGTGAAC
pGEX EndoS 926 R	TATTGCGGCCGCTCAATTCGGCAGCGGATAGCCC
pGEX EndoS 760 R	ATAAGCGGCCGCTCAATCGCCAACAATCACTTTGGC
pET303 EndoS F1	GATCAGTCTAGAATGGACAAACATCTG
pET303 EndoS R1	TTCAGTCTCGAGTTTTTTCAGCAGTTG
pET303 EndoS F2	GAGGTCTAGAATGGACAAACATC
pET303 EndoS R2	GGTGCTCGAGTTTTTTCAGCAG
pET303 EndoS 761 F	TATCTCTAGAATGGAAAAAACCATGATGGTGAAC
pET303 EndoS 926 R	ATAGCTCGAGATTCGGCAGCGGATAGCCC
pET303 EndoS 760 R	ATAGCTCGAGATCGCCAACAATCACTTTGGC
pET303 EndoS 924 F	AAGAAGGAGGTCTAGAATGCTGCCGAATGCTGATA
pET303 EndoS 995 R	GGTGGTGGTGCTCGAGTTTTTTCAGCAGTTGGCGTT
Overlap 927 F	GTGATTGTTGGCGATGCTGATACGATCATG
Overlap 760 R	CATGATCGTATCAGCATCGCCAACAATCAC

optimised for expression in *E. coli*. Details of primers used for IdeS construct design can be found in Table 2.3.

Table 2.3: IdeS cloning primers

Primer	Sequence (5′ – 3′)
IdeSM1 F	CTTTAAGAAGGAGGTCTAGAATGCGTAAACGTTGTTATA
IdeSM1 39 F	CTTTAAGAAGGAGGTCTAGAATGATCCGTTATAGCGAAGTT
IdeSM1 R	TGGTGGTGGTGGTGCTCGAGATTGGTCTGATTCC
IdeSM59 F	GATCAGTCTAGAATGCGTAAGCGTTGT
IdeSM59 39 F	CTTTAAGAAGGAGGTCTAGAATGATCCGTTATAGTGAAGTCA
IdeSM59 R1	TTCAGTCTCGAGGTTGGTTTGGTTCCA
IdeSM59 R2	TGGTGGTGGTGGTGCTCGAGGTTGGTTTGGTTC

The *E. coli* codon-optimised *ideS* serotype M59 gene was synthesised by GenScript into the pUC57-Simple vector. The gene was then amplified by PCR from this vector using primers designed to append 3′ XbaI and 5′ XhoI sites (Table 2.3, IdeSM59 F and IdeSM59 R2). Both the PCR product and empty Champion^{TM} pET303 (Invitrogen) vector were digested with XbaI and XhoI. The IdeSM59 insert was then ligated into the Champion^{TM} pET303 (Invitrogen) vector using T4 Ligase. This initial construct was cloned by Mrs Susan Tickner (Glycobiology Institute, University of Oxford).

Following the ligation reaction, the plasmid was transformed into XL-Blue Competent *E. coli* cells (Agilent Technologies) following the manufacturer's instructions and was selected

for by carbenicillin resistance. Positive colonies were grown overnight in Miller LB prior to extraction of plasmid DNA. DNA sequencing was carried out to verify correct insertion of the *ideS* gene.

The codon-optimised serotype M1 gene was synthesised by LifeTechnologies GeneArt into the pMA-T vector to contain both 3′ XbaI and 5′ XhoI sites. The gene was then amplified by PCR where additional 3′ and 5′ sequences were incorporated to render the product compatible with the In-Fusion® cloning system. This product was then cloned into the expression vector Champion™ pET303 (Invitrogen) using the In-Fusion® system (see Section 2.2.2).

IdeS constructs lacking the N-terminal signal sequence were made using the M1 and M59 *ideS*-pET303 vectors as templates. In both cases, PCR was used to amplify the DNA fragment corresponding to protein residues 39–339. The fragments were then cloned back into the pET303 vector using the In-Fusion® protocol (see Section 2.2.2).

2.2.5 Immunoglobulins

Human IgG1 fragments

The constructs for IgG1 Fc, CH2-H and CH2 were cloned for recombinant expression in mammalian cells. The gene for human IgG1 Fc encoding residues 224–446 (UniProt ID: P01857) was cloned into the mammalian expression vector pHLSec by Dr Ben Yu (Glycobiology Institute, University of Oxford).

Using this IgG1 Fc sequence as a template, a CH2-H construct was designed to contain the hinge region and CH2 domain of IgG1 Fc (residues 224–338) and a CH2 construct was made to encompass solely the CH2 domain of IgG1 Fc (residues 231–338). Both the CH2-H and CH2 genes were synthesised by LifeTechnologies GeneArt as linear DNA to contain an additional 15 bp of 5′ and 3′ sequences to allow compatibility with the In-Fusion® cloning system and were cloned as such into the vector pHLSec (see Section 2.2.2).

An additional IgG1 Fc construct was designed to contain no C-terminal hexahistidine tag. The primers detailed in Table 2.4 were used to amplify IgG1 Fc from the pHLSec vector by PCR. The PCR product was designed to contain a 3′ AgeI and a 5′ XhoI as well as

additional 3′ and 5′ sequences included to permit subsequent cloning via the In-Fusion® method. The pHLSec vector was digested with AgeI and XhoI so that upon insertion of the PCR product via the In-Fusion® cloning protocol, the hexahistidine tag was no longer present and was replaced with a stop codon.

Table 2.4: IgG cloning primers

Primer	Sequence (5′ – 3′)
Fc no His-tag F	GTTGCGTAGCTGAAACCGGTCACACATGCCCACCGT
Fc no His-tag R	CGATACTAGTCTCGAGTGATCATTAACCCGGAGAC

Human 2G12 IgG

The expression vectors pγ1HC and pγLC containing the 2G12 IgG sequence were a kind gift from Dr Katie Doores (King's College London, U.K.) and Prof. Dennis Burton (The Scripps Research Institute, La Jolla, U.S.A.).

Murine IgG1 Fc and IgG2a Fc

Murine 6A6-IgG1 and 6A6-IgG2a Fc domains were cloned into the pHLSec expression vector by Dr Ben Yu (Glycobiology Institute, University of Oxford). Vectors encoding the template IgG heavy chain for each isotype were kindly provided by Prof. Falk Nimmerjahn (University of Erlangen-Nuremberg, Germany) [201]. Both murine 6A6-IgG1 Fc (*Mm*IgG1Fc) and 6A6-IgG2a Fc (*Mm*IgG2aFc) were cloned to include residues 224–446.

2.2.6 Control proteins and enzymes

Several constructs for the production of control proteins and glycosyltransferases were required for this project. Their cloning strategies are detailed below.

Maltose binding protein

To generate a C-terminally hexahistidine-tagged maltose binding protein (MBP) construct, the MBP sequence from pETM-44 (EMBL Bacterial Expression Vectors) was amplified

by PCR. The PCR product was designed to be compatible with the In-Fusion® cloning system and was cloned as such into Champion™ pET303 (Invitrogen).

Table 2.5: MBP cloning primers

Primer	Sequence (5′ – 3′)
pET303 MBP F	AAGAAGGAGGTCTAGAATGAAAATCGAAGAAGGTAAAC
pET303 MBP R	GGTGGTGGTGCTCGAGAGTCTGCGCGTCTTTCAG

Glycosyltransferases

Mus musculus β-1,4-galactosyltransferase (B4GalT1) (NCBI Reference Sequence: NP_071641) and *Rattus norvegicus* α-2,6-sialyltransferase (ST6Gal1) (NCBI Reference Sequence: NP_001106815) were cloned by Dr Ben Yu (Glycobiology Institute, University of Oxford) into the mammalian expression vector pHLSec [70]. Both were cloned to contain the soluble region of each protein where residues 127–399 and 89–403 were included for B4GalT1 and ST6Gal1, respectively.

2.2.7 Mutagenesis

Mutagenesis reactions were carried out using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) as per the manufacturer's instructions. DNA primers used for mutagenesis of EndoS and IdeS are listed in Tables 2.6 and 2.7, respectively.

2.3 Protein expression

This section describes the protocols used to recombinantly express proteins required during this thesis. A summary of the construct boundaries for proteins and protein fragments are described in Table 2.8.

2.3.1 EndoS and EndoS fragments

pGEX-4T-1 containing *ndoS* was transformed into BL21 (DE3) SOLO™ (Lucigen) heat shock competent cells following the manufacturer's instructions. Using baffled 2 L conical

Table 2.6: EndoS mutagenesis primers

Primer	Sequence ($5' - 3'$)
D231L F	CAAATACAATCTGGACGGCCTGTTAGTGGACGTTGAACATGATTCAA
D231L R	TTGAATCATGTTCAACGTCCACTAACAGGCCGTCCAGATTGTATTTG
D233L F	AATCTGGACGGCCTGGATGTGTTAGTTGAACATGATTCAATTCCG
D233L R	CGGAATTGAATCATGTTCAACTAACACATCCAGGCCGTCCAGATT
E235L F	GACGGCCTGGATGTGGACGTTTTACATGATTCAATTCCGAA
E235L R	TTCGGAATTGAATCATGTAAAACGTCCACATCCAGGCCGTC
E235Q F	CGGCCTGGATGTGGACGTTCAGCATGATTCAATTCCGAAAG
E235Q R	CTTTCGGAATTGAATCATGCTGAACGTCCACATCCAGGCCG
E235L F	GACGGCCTGGATGTGGACGTTTTACATGATTCAATTCCGAA
E235L R	TTCGGAATTGAATCATGTAAAACGTCCACATCCAGGCCGTC
E235A F	CCTGGATGTGGACGTTGCACATGATTCAATTCCGA
E235A R	TCGGAATTGAATCATGTGCAACGTCCACATCCAGG
E235Q F	CGGCCTGGATGTGGACGTTCAGCATGATTCAATTCCGAAAG
E235Q R	CTTTCGGAATTGAATCATGCTGAACGTCCACATCCAGGCCG
E235A F	CCTGGATGTGGACGTTGCACATGATTCAATTCCGA
E235A R	TCGGAATTGAATCATGTGCAACGTCCACATCCAGG
C986S F	GCTGGTGTGCTGAAAGACAGTATTGAAAAACGCCAAC
C986S R	GTTGGCGTTTTTCAATACTGTCTTTCAGCACACCAGC
D231L D233L F	TACGTGTACAAATACAATCTGGACGGCCTGTTAGTGTTAGTTGAACA TGATTCAATTCCGAAAGTGGAT
D231L D233L R	ATCCACTTTCGGAATTGAATCATGTTCAACTAACACTAACAGGCCGT CCAGATTGTATTTGTACACGTA
D233L E235L F	GTACAAATACAATCTGGACGGCCTGGATGTGTTAGTTTTACATGATT CAATTCCGAAAGTGGATAAAAAA
D233L E235L R	TTTTTTATCCACTTTCGGAATTGAATCATGTAAAACTAACACATCCA GGCCGTCCAGATTGTATTTGTAC

Table 2.7: IdeS mutagenesis primers

Primer	Sequence ($5' - 3'$)
IdeSM59 C94A F	AACGGCAAGGATGACCTGCTGGCCGGTGCAGCAACCGCA
IdeSM59 C94A R	ATTACCTGCGGTTGCTGCACCGGCCAGCAGGTCATCCTT
IdeSM59 C94S F	AACGGCAAGGATGACCTGCTGAGCGGTGCAGCAACCGCA
IdeSM59 C94S R	ATTACCTGCGGTTGCTGCACCGCTCAGGTCATCCTT
IdeSM59 S94A F	CAAAGATGATCTGCTGGCCGGTGCAGCAACCGCA
IdeSM59 S94A R	TGCGGTTGCTGCACCGGCCAGCAGATCATCTTTG
IdeSM1 S94C F	AAAGATGATCTGCTGTGCGGTGCAGCAACCG
IdeSM1 S94C R	CGGTTGCTGCACCGCACAGCAGATCATCTTT

Table 2.8: Summary of protein constructs.

Uniprot ID	Fragment name	Residues
Q9APG4	EndoS	1–995
	ChitLRR	1–760
	CBMKO	1–760, 925–995
	CBMCT	761–995
	CBM	761–924
	CT	924–995
Q7DAM2	IdeSM1FL	1–339
	IdeSM1-SS	39–339
H8HA37	IdeSM59FL	1–339
	IdeSM59-SS	39–339
P01857	Fc	224–446
	FcnoH	224–446
	CH2-H	224–338
	CH2	231–338
See [201]	*Mm*IgG1Fc	224–446
	*Mm*IgG2aFc	224–446
P15535-1	B4GalT1	127–399
P13721-1	ST6Gal1	89–403
ADO79537.1*	MBP	20–385

*GenBank accession number

flasks, a 500 mL culture of cells was grown in Miller LB containing carbenicillin (100 μg/mL) at 37 °C with shaking (250 rotations per minute (rpm)). Upon reaching an optical density at a wavelength of 600 nm (OD_{600}) of 0.6–0.7, cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) before incubating for a further 4 hours at 37 °C, 250 rpm. Cells were harvested by centrifugation (6000 × g for 15 minutes) and stored at –80 °C.

All remaining EndoS constructs were transformed into *E. coli* BL21 (DE3) cells (NEB) following the manufacturer's instructions. Expression protocols for each of the fragments are described below.

ChitLRR

ChitLRR was expressed as an N-terminal GST-fusion using an identical protocol to that described above for expression of full-length EndoS.

CBM-KO, CBM, CBM-CT and CT

CBM-KO, CBM, CBM-CT and CT were all expressed with a C-terminal hexa-histidine tag. For each construct, a 500 mL culture of cells was grown in a 2 L baffled conical flask using Magic Media^{TM} (Invitrogen) containing carbenicillin (100 μg/mL). Cells were grown at 37 °C with shaking (250 rpm) to an OD_{600} of 0.7–0.8 at which point the temperature was reduced to 25 °C for a further 24 hours of growth. All cells were harvested by centrifugation at 6000 × g for 15 minutes. Cell pellets were frozen and stored at –80 °C prior to lysis and purification.

Isotopically labelled CBM

Expression of isotopically labelled CBM was achieved using a modified recipe for C-75052 media [202]. Here, the recipe was amended to contain no α-lactose or glycerol and was instead supplemented with 0.5% D-glucose. A 500 mL culture was grown in modified C-75052 containing carbenicillin (100 μg/mL) at 37 °C with shaking (250 rpm). The cells were cultured to an OD_{600} 0.7–0.8 before induction with 1 mM IPTG. The cells were then incubated at 25 °C with shaking (250 rpm) for a further 20 hours post induction. For ^{13}C labelled protein, ^{13}C-D-glucose was used in the place of D-glucose and for deuterated samples, all cultures were grown using 100% D_2O.

CBM selenomethionine derivitisation

Selenomethionine (SeMet) derivitisaion of recombinantly expressed CBM was achieved using PASM-5052 media as detailed in [202]. PASM-5052 is an auto-induction media which, when used with *E. coli* BL21 cells, gives over 90% replacement of methionine residues with SeMet.

2.3.2 IdeS

All IdeS constructs were transformed into *E. coli* BL21 (DE3) cells (NEB) as per the manufacturer's instructions. All IdeS proteins were expressed as C-terminal hexa-histidine tagged fusions using the following protocol. Using 2 L baffled conical flasks, a 500 mL

culture of cells was grown at 30 °C with shaking at 250 rpm to an OD_{600} of 0.7-0.8. The temperature of growth was then reduced to 25 °C and cells were grown for a further 24 hours. Cells were then harvested by centrifugation (6000 × g) for 15 minutes. Pelleted cells were stored at –80 °C before subsequent lysis and purification.

2.3.3 MBP

The MBP plasmid was transformed into *E. coli* BL21 (DE3) cells (NEB) as per the manufacturer's instructions. MBP was expressed as a C-terminal hexa-histidine tagged fusion using an identical protocol to that used for the IdeS constructs as detailed in Section 2.3.2.

2.3.4 Immunoglobulins

Human IgG1 fragments

Human serum IgG was purchased from Sigma. The IgG1 fragments Fc, CH2-H and CH2 were transiently expressed in human embryonic kidney (HEK) 293T cells (American Type Culture Collection (ATCC) number CRL-1573) as described previously [203] where DNA for transfections was prepared using the EndoFree® Plasmid Mega Kit (Qiagen).

Briefly, cells were grown in standard T225 flasks (Corning) at 37 °C in a humidified incubator containing 5% v/v CO_2. Cells were cultured in Dulbecco's Modified Eagles Medium (DMEM; Gibco) supplemented with 10% v/v fetal bovine serum (FBS) and 1% v/v penicillin/streptomycin (P/S). For transient expression, endotoxin free plasmid DNA containing the relevant construct was mixed with polyethyleneimine (PEI) at a mass ratio of 1:1.5 in DMEM containing 1% v/v P/S. Cells were cultured to 90% confluence before being transfected with the DNA:PEI mixture. The cells were grown for a further 4 days in DMEM, 1% v/v FBS and 1% v/v P/S at 37 °C, 5% v/v CO_2.

Human 2G12 IgG

Human 2G12 IgG was transiently expressed using HEK 293T cells by Dr Camille Bonomelli (Glycobiology Institute, University of Oxford). Cells were cultured as in Section 2.3.4 prior

to transfection. For transient expression of 2G12, endotoxin free plasmids containing the light chain and heavy chain of 2G12 were mixed in a 3:1 mass ratio. The DNA mixture was combined with PEI at a mass ratio of 1:1.5 in DMEM containing 1% v/v P/S, this mixture was then used to transfect cells cultured to 90% confluence. The cells were then grown undisturbed for 4 days in DMEM, 1% v/v FBS and 1% v/v P/S at 37 °C, 5% v/v CO_2.

Murine IgG1 Fc and IgG2a Fc

*Mm*IgG1Fc and *Mm*IgG2aFc were transiently expressed using HEK 293T cells following the protocol used for human IgG1 fragments described above. Expression of these constructs was carried out by Dr Ben Yu.

2.3.5 Glycosyltransferases

B4GalT1 and ST6Gal1 were transiently expressed in HEK 293T cells using the protocol for expression of human IgG1 fragments in Section 2.4.6.

2.3.6 *In vivo* glycoform modulation

For $Man_5GlcNAc_2$ glycoforms, proteins were expressed using GlcNAc transferase I (GnTI) deficient HEK 293S cells [82].

$Man_9GlcNAc_2$ glycoforms were produced in HEK 293T cells by addition of 20 μM kifunensine (Toronto Research Chemicals) to the cell culture medium at the time of transfection. Hybrid glycoforms were produced in HEK 293T cells where 20 μM swainsonine (Sigma) was added to the cell culture medium at the time of transfection [204].

2.3.7 *In vitro* glycoform modulation

Desialylation of proteins was carried out using neuraminidase from *Clostridium perfringens* (NEB) which hydrolyses α2-3, α2-6 and α2-8 linkages. Recombinant IgG1 Fc was incubated with neuraminidase for 48 hours at 37 °C.

Hyper-β1,4-galactosylation was achieved by treatment of the desialylated Fc with B4GalT1. Fc was incubated with B4GalT1 in 50 mM 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES) pH 7.5, 150 mM NaCl, 10 mM $MnCl_2$ supplemented with 80 μM uridine 5′-diphosphogalactose (UDP-galactose). The hyper–β1,4-galactosylated Fc was treated with ST6Gal1 to achieve hyper-α-2,6-sialylation. Fc was incubated with ST6Gal1 in 50 mM HEPES pH 6.5, 150 mM NaCl, 10 mM $MnCl_2$ containing cytidine-5′-monophospho-*N*-acetylneuraminic acid (CMP-sialic acid) for 48 hours at 37 °C.

2.4 Protein purification

All fast protein liquid chromatography (FPLC) methods were performed using an ÄKTA Pure system (GE Healthcare). Chromatography columns used for protein purification were purchased from GE Healthcare. In all cases, the presence of protein was monitored by absorbance at a wavelength of 280 nm (A_{280}) and the purity of each protein was monitored by sodium-dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Unless otherwise stated, pelleted cells were resuspended in phosphate buffered saline (PBS) containing lysozyme (Sigma), Deoxyribonuclease I (DNaseI) (Roche) and protease inhibitors (c*O*mplete EDTA free; Roche). Cells were then mechanically lysed using a single pass through a cell disruptor at 200 MPa (Constant Systems). Unbroken cells and cell debris were removed by centrifugation at 30,000 × g for 30 minutes. The cell lysate supernatant was passed through a 0.22 μm filter before further purification as detailed below for individual constructs.

2.4.1 EndoS

GST-EndoS was purified from cell lysate using a GSTPrep^{TM} FF 16/10 column as previously described [143, 150] or Glutathione Sepharose 4B beads (GE Healthcare). For Glutathione Sepharose 4B beads, the cell lysate was incubated with the beads for four hours at room temperature, the bead-lysate slurry was then applied to a gravity column where cell lysate was allowed to flow through. The beads were then washed before bound protein was eluted following an identical protocol to that used for the GSTPrep^{TM} FF 16/10 column. Fractions containing putative GST-EndoS were pooled and total protein

content was measured by bicinchoninic acid assay [205] (Pierce). Thrombin was then added at 1 U per 0.1 mg protein and incubated at room temperature for 24 hours to remove the GST tag. Cleavage was carried out as the sample was dialysed into 20 mM Tris pH 8.0, 10 mM NaCl in preparation for anion exchange chromatography (AEC).

AEC was carried out using a HiPrep Q HP column. Protein was separated using buffer A (20 mM Tris pH 8.0) and a gradient of buffer B (20 mM Tris pH 8.0, 1 M NaCl) increasing from 1-15% over 15 column volumes (CV). Fractions containing putative EndoS were pooled and concentrated using a Vivaspin20 concentrator (10 kDa molecular weight cut-off (MWCO)). The sample was then purified by size exclusion chromatography (SEC) using a HiPrep Sephacryl S-200 16/60 column equilibrated with 10 mM HEPES pH 8.0, 150 mM NaCl.

2.4.2 ChitLRR

GST-ChitLRR was purified following the protocol described for GST-EndoS with amendments to the AEC step as follows: the thrombin cleaved sample was dialysed into 20 mM Tris pH 8.0, 20 mM NaCl. The sample was then purified using a gradient of buffer B increasing from 2-50% over 10 CV. Fractions containing putative ChitLRR were pooled and concentrated using a Vivaspin20 concentrator (10 kDa MWCO). The sample was then purified by SEC using a Superdex 200 HiPrep 16/60 column equilibrated with 10 mM HEPES pH 8.0, 150 mM NaCl.

2.4.3 CBM-KO

The hexa-histidine tagged CBM-KO was purified from the cell lysate using HisPur^{TM} Ni-nitrilotriacetic acid (Ni-NTA) resin (ThermoScientific). Supernatant was incubated with the resin for 2 hours at room temperature. The protein-resin slurry was then applied to a gravity column where the supernatant was allowed to flow through. The resin was washed with 20 CV of PBS followed by a second washing step using 20 CV PBS with 20 mM imidazole. Protein was eluted using 5 CV PBS containing 300 mM imidazole.

The eluted protein was concentrated using a Vivaspin20 concentrator (10 kDa MWCO) for further purification. The putative CBM-KO sample was then dialysed into 20 mM Tris pH

8.0, 20 mM NaCl in preparation for an AEC step using a HiPrep Q HP column. Dialysed sample was separated using buffer A (20 mM Tris pH 8.0) and a gradient of buffer B (20 mM Tris pH 8.0, 1M NaCl) increasing from 2-50% over 10 CV. Finally, putative CBM-KO containing fractions were purified by SEC. The sample was applied to a Superdex 200 HiPrep 16/60 column equilibrated with 10 mM HEPES pH 8.0, 150 mM NaCl.

2.4.4 CBM, CBM-CT, CT and MBP

C-terminally hexa-histidine tagged CBM, CBM-CT, CT and MBP were purified from the cell lysate supernatant using HisPur^{TM} Ni-NTA resin (ThermoScientific) as described for the CBM-KO construct.

The eluted fractions were concentrated using Vivaspin20 concentrators (10 kDa MWCO) before a SEC step. The samples were purified by using a Superdex 75 HiPrep 16/60 column equilibrated with 20 mM Tris pH 7.5, 150 mM NaCl.

2.4.5 IdeS

Both full-length IdeS (IdeSFL) and IdeS without the signal sequence (IdeS-SS) were purified using an identical protocol. All steps were carried out on ice or at 4 °C. The harvested cell pellet from a 250 mL *E. coli* culture was resuspended in 80 mL PBS containing lysozyme (Sigma) and DNaseI (Roche). The cells were lysed by sonication then cell debris was pelleted by centrifugation at 30,000 × g for 25 minutes. The supernatant was then decanted and diluted with an equivalent volume of binding buffer (100 mM HEPES pH 7.5, 150 mM NaCl, 5 mM imidazole). The clarified cell lysate was then filtered through a 0.22 μm pore filter.

The cell lysate was applied to a 5 mL HisTrap^{TM} FF column which was then washed with 10 CV of binding buffer. Using binding buffer as buffer A, a stepwise gradient of buffer B (100 mM HEPES pH 7.5, 150 mM NaCl, 300 mM imidazole) was carried out to give six 25 mL steps with imidazole concentrations of 20, 40, 60, 80, 100 and 300 mM. The fractions from the final three imidazole concentrations were pooled and concentrated using a Vivaspin20 concentrator (10 kDa MWCO). The concentrated sample was further purified using a Superdex 75 HiPrep 16/60 column equilibrated with 10 mM HEPES pH

8.0, 150 mM NaCl.

2.4.6 Immunoglobulins

Human IgG1 fragments

For the IgG Fc, CH2-H and CH2 constructs, cell culture supernatant was decanted from culture flasks 4 days post-transfection and centrifuged at 2000 × g for 5 minutes to pellet any residual cells. The supernatant was then passed through a 0.22 μm filter and diluted 1:1 with binding buffer (PBS, 10 mM HEPES pH 8.0). The protein of interest was then purified by immobilised metal affinity chromatography using HisPur^{TM} Ni-NTA resin (ThermoScientific). Here, protein was bound to resin for 2 hours at room temperature. The protein-resin slurry was then applied to a gravity column. The supernatant was allowed to flow through before the resin was washed with 20 CV of 10 mM HEPES pH 8.0. The resin was washed further with 20 CV 10 mM HEPES pH 8.0, 20 mM imidazole prior to elution of the bound protein using 5 CV 10 mM HEPES pH 8.0, 300 mM imidazole.

Eluted fractions containing the target protein were concentrated using a Vivaspin20 concentrator (10 kDa MWCO for Fc, 3 kDa MWCO for CH2-H and CH2) for further purification by SEC. Samples were applied to a pre-equilibrated Superdex 200 HiPrep 16/60 (Fc) or Superdex 75 HiPrep 16/60 (CH2-H, CH2) column. Protein was eluted with 10 mM HEPES pH 8.0, 150 mM NaCl.

For the FcnoH construct the filtered cell culture supernatant was diluted 1:1 with 20 mM sodium phosphate, pH 7.0 before being loaded onto a 1 mL HiTrap rProtein A FF column. To remove unbound protein, the column was washed with 20 CV of 20 mM sodium phosphate, pH 7.0. FcnoH was eluted using 100 mM citrate pH 3.5 and immediately neutralised using 1 M Tris pH 9.0. Eluted protein was then concentrated using a Vivaspin20 concentrator (10 kDa MWCO) before SEC and analysis following an identical protocol to the Fc construct as above.

Human 2G12 IgG

The supernatant from cultured HEK 293T cells was decanted from culture flasks 4 days post-transfection. Residual cells were pelleted by centrifugation at 2000 × g for 5 minutes. The supernatant was decanted from the cells and diluted 1:1 with 20 mM sodium phosphate pH 7.0. The solution was filtered using a 0.22 μm filter and loaded onto a 1 mL HiTrap rProtein A FF column. The column was washed with 20 CV of 20 mM sodium phosphate pH 7.0 to remove unbound protein. Bound 2G12 was eluted using 100 mM citrate pH 3.5 and immediately neutralised using 1 M Tris pH 9.0. Eluted protein was then concentrated using a Vivaspin20 concentrator (10 kDa MWCO) before being loaded onto a Superdex 200 HiPrep 16/60 column. Protein was eluted with 10 mM HEPES pH 8.0, 150 mM NaCl.

Murine IgG1 Fc and IgG2a Fc

For the murine IgG Fc constructs, cell culture supernatant was decanted from culture flasks 4 days post-transfection and centrifuged at 2000 × g for 5 minutes to pellet any residual cells. The supernatant was then passed through a 0.22 μm filter and diluted 1:1 with PBS. *Mm*IgG1Fc and *Mm*IgG2aFc were bound to HisPur^{TM} Ni-NTA resin (ThermoScientific) for 2 hours at room temperature. Using a gravity column, the beads were washed with 20 CV PBS, 20 mM imidazole. Proteins were then eluted with 5 CV PBS, 500 mM imdazole. Eluted fractions containing the target protein were concentrated using a Vivaspin20 concentrator (10 kDa MWCO).

Both *Mm*IgG1Fc and *Mm*IgG2aFc were subjected to EndoH (NEB) treatment to remove high-mannose glycans. EndoH deglycosylation was carried out under native conditions for 16 hours at 37 °C. The EndoH treated samples were subsequently treated with EndoS as described in Section 2.9.1. After deglycosylation samples were applied to a pre-equilibrated Superdex 200 HiPrep 16/60. Protein was eluted with 10 mM HEPES pH 8.0, 150 mM NaCl.

An additional SEC step using a Superdex 75 HiPrep 16/60 column was used in an attempt to purify *Mm*IgG1Fc and *Mm*IgG2aFc from residual EndoH. Protein was eluted with 10 mM HEPES pH 8.0, 150 mM NaCl.

2.4.7 Glycosyltransferases

For both B4GalT1 and ST6Gal1, cell culture supernatant was decanted from culture flasks 5 days post-transfection and centrifuged at 2000 × g for 5 minutes to pellet residual cells. Supernatant was diluted 1:1 with PBS, 20 mM imidazole before being passed through a 0.22 μm filter. Filtrate was incubated with HisPur^{TM} Ni-NTA resin (ThermoScientific) for 2 hours at room temperature. The protein-resin slurry was then applied to a gravity column where the supernatant was allowed to flow through. The resin was washed with 20 CV of PBS, 20 mM imidazole before bound protein was eluted using 2.5 CV PBS, 500 mM imidazole.

Eluted protein was concentrated using a Vivaspin20 concentrator (10 kDa MWCO) and was then applied to a Superdex 200 16/60 column equilibrated with PBS.

2.5 Enzyme substrate complex formation

For the complexes used in subsequent biophysical experiments, catalytically inactive mutants E235L (EndoS) and C94A (IdeS) were used to produce complexes of these enzymes with Fc. EndoS was also complexed to the full-length antibody 2G12.

2.5.1 EndoS

EndoS was mixed with IgG1 Fc to give approximately a 1:1.5 molar ratio (*i.e.* a molar excess of IgG1 Fc). The sample was mixed by pipette action and was immediately applied to a Sephacryl S-200 column (GE Healthcare) equilibrated with 10 mM HEPES pH 8.0, 150 mM NaCl. Combinations of EndoS and IgG1 Fc used for complex formation are discussed in Section 3.4. For EndoS:2G12 complexes, an identical protocol was followed.

2.5.2 IdeS

IdeS constructs containing a C94A mutation were mixed with IgG1 Fc in approximately a 1:1 molar ratio. The mixed sample was immediately applied to a Superdex 200 column (GE-Healthcare) equilibrated with 10 mM HEPES pH 8.0, 150 mM NaCl.

2.6 Biophysical methods

2.6.1 Determining protein concentration

Protein concentration in solution was calculated by measuring the A_{280} in conjunction with the predicted extinction coefficients as listed in Table 2.9. The A_{280} was measured using a NanoDrop 1000 (ThermoFischer). Extinction coefficients (ε) were calculated using Protparam from the ExPASY server [206]. This server employs an Edelhoch based method to determine ε based on primary sequence [207, 208, 209].

For serum glycoproteins, values used to calculate protein concentration based on the A_{280} were recommended by the suppliers. These values are referenced in Table 2.10. Approximate values for molecular weight (M_w) are due to differential glycosylation as well as sequence variation in human derived samples.

Table 2.9: Extinction coefficients used for determining protein concentration. The values quoted for IgG1 Fc refer to the monomer.

Protein	M_w (g mol^{-1})	ε (M^{-1} cm^{-1})	Abs (1% w/v)
EndoS	111,864.5	121,130	10.83
CBM-KO	94,133.7	90,315	9.58
ChitLRR	85,300.4	90,190	10.57
CBM	20,010.3	30,940	15.46
CBM-CT	27,778.4	30,940	11.14
IdeSM1-FL	39,069.8	50,880	13.02
IdeSM59-FL	39,389.3	53,860	13.67
IdeSM1-SS	35,241.4	49,390	14.01
IdeSM59-SS	35,429.7	52,370	14.78
IgG1 Fc*	26,965.5	35,785	13.27
IgG1 FcnoH*	25,856.3	35,785	13.84
EndoS Fc	166,864.5	191,950	11.50
IdeSM1-ss Fc	90,241.4	120,210	13.32
IdeSM59-ss Fc	90,429.7	123,190	13.62
IdeSM59FL Fc	94,389.3	124,680	13.21

* Values for IgG1 Fc and FcnoH refer to the monomer

2.6.2 Gel electrophoresis and western blot methods

All gels and buffers used for analysis of proteins by electrophoresis were purchased from Life Technologies. Details of each procedure are described below.

Table 2.10: Parameters for determining serum glycoprotein protein concentration. All values are approximate and were determined by the supplier.

Protein	M_w (g mol^{-1})	Abs (1% w/v)
Serum IgG	150,000	14.0
Serum IgA	160,000	13.0
Transferrin	76,000-81,000	11.0
α-1-antitrypsin	52,000	4.5
α-2-macroglobulin	725,000	9.0

SDS-PAGE

SDS-PAGE analysis of protein was carried out using Novex^{TM} Bis-Tris 4-12% gels. NuPAGE® LDS Sample Buffer was added to samples which were then incubated at 95 °C for 2 minutes. For reducing conditions, dithiothreitol (DTT) was added to a final concentration of 10 mM prior to incubation at 95 °C. SDS-PAGE gels were run with NuPAGE® MES SDS Running Buffer (50 mM MES, 50 mM Tris, 0.1% SDS w/v, 1 mM EDTA, pH 7.3) and stained with InstantBlue (Expedeon). Gels were destained with water. The Novex® Pre-stained Protein Marker was used unless otherwise stated.

For SDS-PAGE gels used for glycan analysis (Section 2.9.1), gels were stained with a colloidal Coomassie stain containing 0.3% w/v Coomassie R-250 in 45% v/v methanol, 10% v/v acetic acid. Gels were destained extensively in a solution of 30% v/v methanol, 7% v/v acetic acid and were then washed in water to remove residual destaining solution.

Western blot

For western blot analysis, protein from SDS-PAGE gels was transferred to a nitrocellulose Hybond^{TM}-ECL membrane using an XCell II^{TM} Blot Module. The transfer was run at 30 V, 145 mA for 2 hours using NuPAGE® MES SDS Running Buffer spiked with 20% v/v methanol.

Blotted membranes were blocked overnight at 4 °C with blocking buffer (5% w/v milk solution in PBS containing 0.05% Tween 20 (PBS-T)) . Blocking buffer was removed and then a 1:10000 dilution of antibody (Anti-His(C-term)-Horseradish peroxidase (HRP) Antibody, Life Technologies) in blocking buffer was incubated with the membrane for 1 hour at room temperature. To remove residual antibody and blocking buffer, membranes

were washed three times with PBS-T for 10 minutes at room temperature.

Western blots were developed using ECL Plus Western Blotting Detection Reagents (GE Healthcare) as per the manufacturer's instructions. Membranes were visualised using a FUJIFILM LAS-1000 camera in concert with a FUJIFILM IDX2 Intelligent Dark Box II (Fuji).

Iso-electric focusing

For iso-electric focusing (IEF) gel electrophoresis, proteins were analysed using Novex® pH 3-10 IEF Protein Gels. Samples were prepared by addition of 5 μL IEF Sample Buffer to 5 μL of the protein sample.

Both IEF Anode Buffer (7 mM phosphoric acid) and IEF Cathode Buffer (2 mM lysine, 2 mM arginine) were cooled to 4 °C before use. Gels were run using an XCell SureLock™ Mini-Cell Electrophoresis System (Life Technologies). IEF gels were run at 100 V for 1 hour, 200 V for 1 hour and finally 500 V for 30 minutes.

Gels were fixed in 12% trichloroacetic acid for 30 minutes at room temperature with gentle agitation prior to staining with InstantBlue (Expedeon).

2.6.3 Circular dichroism

Circular dichroism (CD) spectra were collected using a Jasco J-815 spectrometer (Jasco). Protein concentration was determined as in Section 2.6.1. Spectra were acquired using a quartz cuvette with a 0.1 cm path length. The spectra were collected between 260–180 nm at 50 nm/min at 0.5 nm increments with a response time of 0.5 s and a data pitch of 0.5 nm. Baselines were collected in the same manner using buffer only.

Spectra were averaged over 4 scans then corrected for solvent background and smoothed using a Savitzky-Golay method [210]. To determine secondary structure components, various deconvolution methods were applied to the data, these were accessed using the DICHROWEB server [211].

2.6.4 Small-angle X-ray scattering

The principles of biomolecular small angle X-ray scattering (SAXS) are described in a recent review by Jacques and Trewhella [212]. A more detailed explanation of SAXS can be found in Feigin and Svergun (1987) [213]. A brief explanation of SAXS is given in this section where several equations are included to ease explanation of both methods and results in later chapters.

SAXS provides a method for determining low resolution structural information about protein molecules in solution, including protein flexibility and conformations [214]. For SAXS, the protein is treated as an object which uniformly scatters X-rays. To account for scattering of the solvent, the mean scattering of the protein (ρ_r) minus the mean scattering of the solvent (ρ_s) is used to give the mean contrast scattering ($\Delta\rho$) (Equation 2.1). The X-ray scattering intensity at a given scattering vector ($I(q)$) is proportional to square of this contrast (Equation 2.2). The intensity scattering profile of a protein over a range of q can be used to determine information about the size, shape and behaviour of the protein in solution.

$$\Delta\rho = \bar{\rho}_r - \bar{\rho}_s \tag{2.1}$$

$$I(q) \propto \Delta\rho^2 \tag{2.2}$$

Sample preparation

All protein samples were purified as per their respective purification protocols (Section 2.4) and were shipped on wet ice and stored at 4 °C prior to SAXS experiments. For buffer subtraction measurements, buffer eluted from the relevant column used in the final SEC step of each protein purification was collected. This was done after the column was equilibrated and before sample was applied.

Immediately prior to SAXS measurements, protein solutions were filtered using Ultrafree-MC microcentrifuge filters with 0.22 μm pore size (Millipore) to remove aggregated protein. Protein concentrations were measured as detailed in Section 2.6.1 and a two-fold dilution

series was made for each protein. Buffer blanks were prepared in an identical manner. Buffer measurements were taken before and after each protein measurement. Details of data collection at beamlines P12 (Deutsches Elektronen-Synchrotron (DESY), Hamburg) and BM29 (European Synchrotron Radiation Facility (ESRF), Grenoble) are given below.

DESY P12

Scattering data were collected using a Pilatus 2M detector (Dectris). Beamline P12 was set to a wavelength of 1.23955 Å and the detector was set at a distance of 3.1 m from the sample. The range of the modulus of the scattering vector collected was 0.012 $Å^{-1} < q >$ 0.45 $Å^{-1}$. Equation 2.3 describes the scattering vector q where λ is the X-ray wavelength and θ is half the angle between the incident and scattered radiation.

$$q = \frac{4sin(\theta)}{\lambda} \tag{2.3}$$

For data collection, a total of 50 μL was continuously flowed through the sample capillary for each protein dilution series. Using a 0.05 s exposure time, 20 frames were collected during the sample flow using 100% transmission. The flow cell was kept at 20 °C for all experiments.

ESRF BM29

Scattering data were collected using a Pilatus 1M detector (Dectris). BM29 was set to a wavelength of 0.9919 Å and the detector was set at a distance of 2.867 m. The range of the modulus of the scattering vector collected was 0.014 $Å^{-1} < q > 0.45$ $Å^{-1}$.

Volumes of either 40 or 50 μL were continuously flowed through the sample capillary for each protein dilution series. Using a 1.0 s exposure time, 10 frames were collected during the sample flow using 100% transmission. The flow cell was kept at 20 °C for all experiments.

Data processing

Data reduction was carried out automatically at the beamline using the ATSAS package [215]. The curves of the scattered intensities I (represented as $\ln(I)$ versus q) were calculated for all frames of each protein concentration. The frames were averaged to give a single curve. Frames containing radiation damage (as judged by increasing intensities at low q values) were discarded. The same averaging procedure was carried out for buffer measurements. To account for solvent scattering, the average curve for buffer measurements taken both before and after the relevant protein sample was subtracted from the protein curve.

Data were further processed using SCÅTTER [216]. Here the extrapolated zero angle scattering intensity $I(0)$ and the radii of gyration (R_g) were both calculated using the Guinier approximation [217]. This assumes that for small values of q (where, $q \cdot R_\mathrm{g} < 1.3$), that the $I(0)$ and R_g can be approximated by Equation 2.4.

$$lnI(q) = lnI(0) \cdot (1 - \frac{q^2 \cdot {R_g}^2}{3}) \tag{2.4}$$

SCÅTTER [216] was also used to calculate the paired-distance distribution function (*P(r)*), the maximum dimension (D_{max}) and real space values for R_g and $I(0)$. The Kratky [218] and normalised Kratky plots were calculated as described by Rambo and Tainer [214] using SCÅTTER [216] and plotted using GraphPad Prism (GraphPad Software Inc., La Jolla, U.S.A.).

The $I(0)$ measured for a sample of BSA was used to calculate the approximate molecular weight of the protein or protein complex present in each sample. This was calculated using Equation 2.5 and the measured $I(0)$ for BSA samples shown in Table 2.11, where 66 kDa is the molecular weight of BSA. The $I(0)$ from the ESRF was calculated using a single solvent corrected curve of BSA collected from a 5 mg mL^{-1} sample. The $I(0)$ from the DESY was calculated using solvent corrected curves from a 2-fold dilution series of BSA with a top concentration of 4 mg mL^{-1}.

$$M_w = \frac{I(0)_{sample} \times 66kDa}{I(0)_{BSA}} \tag{2.5}$$

Table 2.11: $I(0)$ values of BSA standards

	ESRF	DESY
$I(0)$	62.9	13800

Further processing of SAXS data was carried out using programmes available through the ATSAS suite [219] where details are given in the relevant results sections. Figures were generated using PyMOL [220].

2.6.5 Nuclear magnetic resonance

All nuclear magnetic resonance (NMR) spectra were collected on a Bruker Avance III 600 MHz spectrometer fitted with a TCI cryoprobe. Experiments were performed at 37 °C. Spectra were processed with NMRpipe [221] and visualised using CcpNmr Analysis [222].

Monosaccharide titrations

Purified ^{15}N-CBM (0.15 mM) was titrated with increasing concentrations of D-galactose, D-glucose, or L-fructose (62.5 mM to 2 M) and *N*-acetylglucosamine (62.5 mM to 250 mM). All experiments were performed in 10 mM BisTris pH 6.5, 15 mM NaCl. ^{15}N-^{1}H heteronuclear single-quantum coherence (HSQC) spectra were acquired with 16 scans. Acquisition time and spectral width were 107 ms and 9579 Hz in the direct dimension (^{1}H) and 66 ms and 1929 Hz in the indirect dimension (^{15}N), respectively.

2.6.6 Enzyme-linked immunosorbent assays

Enzyme-linked immunosorbent assay (ELISA) experiments were carried out using 96-well flat bottom high-binding plates (Costar). Plates were coated with recombinantly expressed FcnoH by incubation of a 5 μg/mL solution overnight at 4 °C.

Three successive washing steps of PBS-T were used to remove excess protein. Plates were then blocked for 1 hour at room temperature with gentle agitation using using a solution of 3% w/v bovine serum albumin (BSA) in PBS-T.

The BSA solution was removed before proteins of interest were added at various concentrations to assess IgG Fc binding capability. Proteins were incubated for 2 hours at room

temperature with gentle agitation to allow binding interactions to occur. After incubation, five successive washing steps using PBS-T were used to remove excess or unbound protein.

Bound His-tagged protein was detected using HisProbe-HRP (ThermoScientific). HisProbe-HRP was added at a final concentration of 2.5 μg/mL in PBS-T and was incubated for 1 hour at room temperature. Bound HisProbe-HRP was detected by addition of 1-Step^{TM} Ultra TMB-ELISA (ThermoScientific) which was incubated for 30 mins in the dark. The colorimetric reaction was stopped using 2 M H_2SO_4.

Analyses of ELISA experiments were carried out using GraphPad Prism (GraphPad Software, Inc., La Jolla, U.S.A.). Here, Boltzmann sigmoidal curves were fitted to triplicate measurements of each ligand protein concentration.

2.6.7 Isothermal titration calorimetry

Isothermal calorimetry titrations were carried out using a MicroCal iTC200 (GE Healthcare) instrument. All measurements were carried out at 20 °C with a stirring speed of 1000 rpm. Purified CBM was dialysed extensively into PBS and a stock 1 M D-galactose solution was made by dissolving the appropriate mass of D-galactose in the dialysate produced during the dialysis of the CBM. The CBM (1.4 mM) was titrated with repeated 2 μL injections of 42 mM D-galactose. For the heat of dilution controls, CBM was titrated with PBS and PBS was titrated with D-galactose. Control experiments utilised identical concentrations to those used in the binding experiment.

2.7 Protein crystallisation

To ascertain appropriate protein concentrations for crystallisation trials, the PCT^{TM} Pre-Crystallization Test (Hampton Research) was used following the manufacturer's instructions.

All high-throughput (HT) crystallisation trials were carried out in 96-well MRC 2 Well Crystallisation Plates (Swissci). A Mosquito Crystal robot (TTP Labtech) was used to pipette the nanolitre volumes required in these trials. Crystallisation trials were conducted using the sitting-drop vapour diffusion method [223] where drops typically comprised 100

nL protein solution and 100 nL mother liquor. The volume of mother liquor in the reservoir was 50 μL. Sitting drop vapour diffusion experiments that were set up in a 24-well format utilised Cryschem M Plates (Hampton Research). Drops consisted of 1 μL protein solution and 1 μL mother liquor with a reservoir volume of 500 μL.

Crystallisation trials used the following crystallisation screens from Molecular Dimensions: PACT, JCSG+, ProPlex, Morpheus, MIDAS^{TM} and the JCSG Core Suite. PEG/Ion HT and Index HT screens were purchased from Hampton Research. Crystallisation trials were set up at both 4 °C and 18 °C. Plates incubated at 4 °C were stored in a Rigaku Gallery HT and imaged automatically with a Minstrel^{TM} HT (Rigaku) at 0, 1, 3, 5, 7 and 12 days post setup and then weekly for a further 6 weeks. Plates incubated at 18 °C were inspected by eye using a Leica MZ6 light microscope. From May 2014, plates incubated at 18 °C were stored in a Rigaku Gallery HT and were imaged in an identical manner to those incubated at 4 °C.

Unless otherwise stated, crystals were harvested from drops using Mounted CryoLoop^{TM} nylon loops (Hampton Research). Cryoprotected crystals were cryocooled by immersing the harvested crystals in liquid nitrogen. Crystals were stored in liquid nitrogen until X-ray diffraction studies could be conducted.

2.7.1 CBM

Crystallisation trials for CBM were set up at concentrations of 12.8–13 mg mL $^{-1}$. Of the initial conditions which yielded crystals, several were chosen to create optimisation screens for both 96-well and 24-well sitting drop vapour diffusion experiments. The conditions which were used for optimisation are listed in Table 2.12. SeMet derivatised protein was set up for crystallisation in the optimisation screens used for crystallisation of the native protein.

The crystallisation condition used to grow diffraction quality crystals was modified from Morpheus [224], condition F7. The optimised condition contained freshly purified 12.8 mg mL^{-1} SeMet CBM, 0.1 M HEPES:MOPS (3-(*N*-morpholino)propanesulfonic acid) pH 6.5, 120 mM Monosaccharides (20 mM D-galactose, 20 mM D-glucose, 20 mM D-mannose, 20 mM *N*-acetylglucosamine, 20 mM xylose and 20 mM L-fucose), 9.7% PEG 8000 and

Table 2.12: Conditions used for CBM crystal optimisation.

Screen	Well number	Condition components
Morpheus	F6	0.12 M Monosaccharides[†] 0.1 M HEPES:MOPS pH 7.5 10 % w/v PEG 8000 20 % w/v ethylene glycol
Morpheus	F7	0.12 M Monosaccharides[†] 0.1 M HEPES:MOPS pH 7.5 10 % w/v PEG 4000 20 % w/v glycerol
PEG/Ion HT	D10	0.2 M sodium citrate tribasic dihydrate pH 8.3 20 % w/v PEG 3350

[†]Monosaccharides solution contained 20 mM D-galactose, 20 mM D-glucose, 20 mM D-mannose, 20 mM *N*-acetylglucosamine, 20 mM xylose and 20 mM L-fucose.

19.4% ethylene glycol. The buffer and monosaccharide components of this screen were purchased as stock solutions from Molecular Dimensions.

Crystals were harvested from the drops and cryoprotected using mother liquor with an increased ethylene glycol concentration totalling 25%. Crystals were routinely fished from the drops using Elliptical LithoLoops (Molecular Dimensions).

For heavy atom soaks, solutions of Tb, Lu, Sm, and Gd were prepared in mother liquor to a final concentration of 1 mM. Crystals were soaked in the heavy atom solutions for intervals ranging from a few seconds to several minutes. Crystals were then soaked in the cryoprotectant solution described above before cryocooling.

2.7.2 Deglycosylated *Mm*IgG1Fc

Crystallisation trials were set up using protein at 5 mg mL^{-1}, 7.5 mg mL^{-1} and 10 mg mL^{-1}. Crystals used for diffraction studies were grown in Index HT condition E7 with protein at 5 mg mL^{-1}. The mother liquor contained 50 mM $MgCl_2 \cdot 6H_2O$, 100 mM HEPES pH 7 and 30% w/v PEG 550 monomethyl ether (MME). For cryoprotection, crystals were soaked in mother liquor containing 10% ethylene glycol.

2.7.3 Deglycosylated *Mm*IgG2aFc

Initial crystallisation trials of *Mm*IgG2aFc used protein concentrations of 7.5 mg mL^{-1} and 10 mg mL^{-1}. Crystals used for X-ray diffraction studies were grown from PACT condition A10. Here, mother liquor contained 200 mM $MgCl_2$, 100 mM sodium acetate pH 5.5 and 20% w/v PEG 6000. Crystals were soaked in mother liquor containing 25% ethylene glycol for cryoprotection.

2.8 X-ray crystallography

For structural studies, protein crystals grown and harvested as described in Section 2.7 were irradiated with X-rays to produce X-ray diffraction data. A brief description of data collection and processing is given below. Further details on both the theory and results of these studies can be found in subsequent chapters.

2.8.1 Data collection

All X-ray diffraction datasets used for protein structure determination were collected at Diamond Light Source (DLS), Harwell, U.K.. The parameters used for data collection from protein crystals of EndoS, *Mm*IgG1Fc, *Mm*IgG2aFc and IdeSFc are described in Table 2.13. Data collection parameters for protein crystals of CBM are described in Table 2.14. Crystals were kept under liquid nitrogen prior to and after irradiation. During data collection crystals were kept in a cryo-stream of gaseous nitrogen at 100 K.

Table 2.13: X-ray diffraction data collection parameters

Parameter	Protein			
	EndoS	*Mm*IgG1Fc	*Mm*IgG2aFc	IdeSFc
DLS beamline	I03	I02	I04-1	I03
Wavelength (Å)	0.9763	0.9786	0.9200	0.9763
Transmission	10% (9.995)	50% (49.97)	100%	50% (50.16)
Beam size (μm)	20 × 20	82 × 21	60 × 50	50 × 20
Oscillation angle	0.5°	0.2°	0.1°	0.2°
Time per image (s)	1.0	0.2	0.1	0.2
Number of images	3	900	900	1800

Table 2.14: X-ray diffraction data collection parameters for CBM crystals

Parameter	Protein			
	SeMet CBM	CBMgal	CBMglc	CBMman
DLS beamline	I02	I03	I03	I03
Wavelength (Å)	0.9786	0.9763	0.9763	0.9763
Transmission	50% (49.97)	50% (49.97)	50% (49.97)	50% (50.16)
Beam size (μm)	82 × 21	80 × 20	50 × 20	50 × 20
Oscillation angle	0.2°	0.2°	0.2°	0.15°
Time per image (s)	0.2	0.2	0.2	0.6
Number of images	1800	1800	1800	600

2.8.2 Data processing

Processing of X-ray diffraction data, phasing, model building and refinement are discussed in detail in the relevant results chapters. Briefly, the autoPROC suite [225] and xia2 [226] were used for initial data reduction. Molecular replacement was carried out using Phaser [227]. Iterative rounds of model building and refinement used Coot [228] and BUSTER [229], respectively. Figures were generated using PyMOL [220] and VMD [230].

2.9 Glycobiology

2.9.1 Enzymatic release of *N*-linked glycans

Peptide *N*-glycosidase F

Glycerol free peptide *N*-glycosidase F (PNGaseF) (NEB) was used to release *N*-linked glycans from proteins. Digests were carried out following the manufacturer's instructions with the exception that the incubation time was increased to 24 hours.

For in-gel release of *N*-linked glycans, the protein of interest was subjected to SDS-PAGE. The resultant gel was stained briefly in colloidal Coomassie stain and destained thoroughly with destaining solution (Section 2.6.2). The gel was then washed extensively with water prior to excision of the protein band of interest.

Excised gel bands were washed with buffer containing 200 mM ammonium bicarbonate, 40% v/v acetonitrile followed by two successive washes of 100% v/v acetonitrile and water. The band was washed again with 100% v/v acetonitrile before the wash solution was

decanted and the residual acetonitrile evaporated using a vacuum centrifuge.

The dried gel piece was rehydrated in 30 μL 20 mM sodium bicarbonate pH 7.0 before addition of 1 μL PNGaseF. The reaction was then incubated at 37 °C for 16 hours. The reaction solution (containing released glycans) was transferred to a new vessel before the gel piece was washed with three successive rounds of water to collect residual released glycans. The pooled solution of glycans was dried using a vacuum centrifuge and was then stored at –20 °C prior to further analysis.

EndoS

For release of *N*-linked glycans by EndoS, reactions were performed in PBS at a molar ratio of 1:10 EndoS:substrate. The reaction was allowed to proceed for 24 hours at 37 °C unless otherwise stated. Where substrate proteins needed to be purified from EndoS for subsequent experiments, the EndoS:substrate mixture was applied to an Sephacryl S-200 16/60 column and eluted with 10 mM HEPES pH 8.0, 150 mM NaCl unless otherwise stated.

2.9.2 Exoglycosidase digests of *N*-linked glycans

Exoglycosidase digests were carried out using *N*-linked glycans which had been enzymatically released from their protein substrates. Prior to digestion reactions, glycans were fluorescently labelled as in Section 2.9.4. All exoglycosidase reactions were carried out at 37 °C for 16 hours. β-*N*-acetylglucosaminidase, β-1,4-galactosidase and α-(1-2, 3, 6) mannosidase, were purchased from QA-Bio. Neuraminidase was purchased from NEB and α- L-Fucoside was purchased from Calbiochem. All digests were performed as per the manufacturers' instructions.

2.9.3 Purification

Enzymatically released glycans and products of exoglycosidase digests were purified using a Ludger Clean^{TM} Post-Exoglycosidase Clean-up Plate (Ludger). These plates contain a nitrocellulose membrane to remove residual glycoprotein and enzymes from the reaction.

Once purified, glycans were dried by vacuum centrifugation and were then stored at –20 °C prior to fluorescent labelling or analysis by mass spectrometry.

2.9.4 Fluorescent labelling

For HPLC analysis, glycans were fluorescently labelled using either LudgerTag^{TM} 2AA (2-aminobenzoic acid) glycan labelling kit (Ludger) or LudgerTag^{TM} 2AB (2-aminobenzamide) glycan labelling kit. To remove excess dye, labeled glycans were purified with LudgerClean^{TM} T1 Glycan Cleanup Cartridges using a vacuum manifold (Ludger). Glycans were dried by vacuum centrifugation and stored at –20 °C prior to analysis.

2.9.5 Mass spectrometry

For matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry (MS), dried glycans were first resuspended in water. To remove residual contamination, the resuspended glycan solution was cleaned using a Nafion 117 membrane (Sigma) for at least 1 hour as previously described [231].

A 1 μL aliquot of the resultant glycan solution was pipetted onto the stainless steel MALDI target. Next, 0.5 μL of a the matrix solution (saturated 2,5-dihydroxybenzoic acid (DHB) solution in 70% v/v acetonitrile) and 0.3 μL aqueous NaCl solution (1 mg mL^{-1}) were added to the drop. The drop was then allowed to dry under ambient conditions. Once dry, the sample/matrix mixture was recrystallised by addition of 0.2 μL 100% ethanol with an aim to produce a homogeneous target surface.

Positive ion MALDI-TOF mass spectra were collected using a Shimadzu-Kratos (Kratos Analytical) AXIMA TOF^{2TM} instrument with a nitrogen laser. Data were recorded in Reflection-DHB mode as previously described [232].

Nano-electrospray ionisation mass spectrometry (ESI MS) was performed with a Waters Synapt G2 ion mobility quadrupole-time-of-flight (Q-TOF) instrument. Samples in 1:1 (v/v) methanol:water containing 0.5 mM ammonium phosphate were infused through Waters thin-wall nanospray capillaries. The ion source conditions were: temperature, 80 °C; infusion needle potential, 1.2 kV; cone voltage 100 V. The travelling-wave ion-mobility cell (nitrogen) was operated with a wave velocity of 450 m s^{-1} and a wave height of 40 V.

For MS/MS data acquisition, the parent ion was selected at low resolution (approximately 4 *m/z* mass window) and fragmented with argon. The voltage on the collision cell was adjusted with mass to give an even distribution of fragment ions across the mass scale. Typical values were 80–120 V. Spectra recorded from 2 s scans were acquired with a digitisation rate of 4 GHz. Other operating voltages were as recommended by the manufacturer. Instrument control, data acquisition and processing were performed with MassLynx software Version 4.1. Waters DriftScope software was used to extract singly charged glycan ions from the total profile and to reject MS/MS fragment ions that were not associated with the target glycan [233]. Glycan fragments were labelled according to the scheme proposed by Domon and Costello [234].

2.9.6 High performance liquid chromatography

HPLC methods for glycan analysis were performed using a Waters Alliance® HPLC System. Glycans were analysed using either a LudgerSep™ N2 High Resolution Amide column (Ludger) or an XBridge Amide column (4.6 x 250 mm, Waters). These columns were run at 0.7 mL min^{-1} and 1.0 mL min^{-1}, respectively. Fluorescently labelled glycan samples were solvated in 35% v/v water, 65% v/v acetonitrile and loaded onto the column. Glycans were then separated using a gradient of solvent A (1.25 mM ammonium formate, pH 4.4) and solvent B (100% v/v acetonitrile). Details of the gradients used with each column can be found in Table 2.15. Fluorescence was detected using an excitation wavelength of 360 nm and a detection wavelength of 425 nm. In the relevant results chapters, fluorescence is reported as relative fluorescence (RF) and chromatograms are labelled using Oxford glycan nomenclature [83] with the colour scheme of the Consortium for Functional Glycomics as previously implemented [91, 87].

All spectra were recorded and analysed using the Empower Pro software package (Waters). To determine glucose-units (GU) of eluted peaks, 2AA or 2AB standards (Ludger) were applied to the column and eluted in an identical manner to the samples. Peak identification was confirmed by a combination of MALDI-MS, ESI-MS, exoglycosidase digests and comparison to previously published spectra.

Table 2.15: HPLC gradients

Method	Time (min)	A%	B%
LudgerSep^{TM} N2 A	0	35	65
	6	35	65
	22	46	54
	24	50	50
	26	100	0
	28	35	65
	46	35	65
LudgerSep^{TM} N2 B	0	35	65
	22	46	54
	22.5	100	0
	24.5	100	0
	26	35	65
	30	35	65
	46	35	65
XBridge	0	35	65
	6	35	65
	46	46	54
	48	80	20
	50	80	20
	52	35	65
	75	35	65

2.9.7 Serum glycoproteins

Serum glycoproteins used to assess EndoS specificity were purchased as purified proteins. Serum IgG, serum IgA and transferrin were purchased from Sigma. Both IgG and IgA were supplied as solutions in PBS.

Transferrin was purchased as lyophilised protein which was subsequently saturated with iron to form holo-transferrin following the manufacturer's instructions. Briefly, lyophilised apo-transferrin was resuspended in 100 mM sodium carbonate pH 5.9 containing ferrous ammonium sulfate hexahydrate (2% mass equivalent of transferrin). The protein was incubated at room temperature for 1.5 hours before the pH was increased to 8.5 with 100 mM sodium carbonate. The protein was then incubated for a further 2 hours at room temperature before dialysis against water overnight at 4 °C.

Both α-2-macroglobulin and α-1-antitrypsin were purchased from Calbiochem as lyophilised protein. Both were resuspended in water prior to glycan analysis and subsequent enzyme assays.

Preparation of IgG Fab and Fc fragments from serum IgG

Human serum IgG (Sigma) was digested into Fab and Fc fragments using agarose-immobilised papain (Thermo Scientific). Prior to the digestion reaction, IgG was dialysed into 20 mM sodium phosphate, 10 mM EDTA, pH 7.0 overnight at 4 °C. Dialysed IgG was added to immobilised papain and the reaction was carried out in 20 mM sodium phosphate, 10 mM EDTA, 10 mM cysteine, pH 7.0. The reaction was allowed to proceed at 37 °C in a ThermoMixer® C shaking incubator (Eppendorf) for 18 hours at 1500 rpm. The reaction was centrifuged (1 min, 1000 × g) to pellet the immobilised resin so that the digested IgG in the supernatant could be removed. The resin was then washed with 10 mM Tris pH 7.5 and centrifuged again in the same manner. The supernatant, containing residual reaction products, was collected. The completeness of the reaction was assessed using reducing SDS-PAGE.

The supernatant was then applied to a 1 mL HiTrap rProtein A FF column. The unbound Fab fragments were washed from the column using 20 CV of 20 mM sodium phosphate pH 7.0. Bound Fc fragments and undigested IgG were eluted using 100 mM citrate pH 3.5 and were immediately neutralised using 1 M Tris pH 9.0. The mixture of Fc fragments and intact IgG was then subjected to a SEC step where the Fc fragments were separated from the mixture with a Superdex 200 10/300 GL column equilibrated with PBS.

3

Structural characterisation of EndoS

EndoS is a glycoside hydrolase secreted by *S. pyogenes* during infection. In addition to its apparent contribution to immune evasion by *S. pyogenes*, it is of interest as a tool for treatment of immunological diseases (reviewed in [180]), as a transglycosidase [143, 235] and as a tool for enhancing specific antibody-receptor interactions through deglycosylation of competing serum IgG [142]. The interaction between EndoS and IgG was further investigated in order to maximise the potential of EndoS in these applications.

At the outset of this project, little information on the atomic structure of EndoS was available. Homology modelling using the amino-acid sequence of EndoS had revealed that the N-terminal region comprised a glycoside hydrolase domain responsible for glycolytic activity followed by a LRR domain with suggested involvement in protein–protein interactions [147]. The C-terminus of EndoS remained largely uncharacterised and the arrangement of the aforementioned domains in relation to each other was not known.

Here, both bioinformatic and biophysical methods were employed with the goal of determ-

ining an atomic structure of EndoS and its complex with IgG1 Fc. The following chapter describes work towards this goal. The work described here is then put into context of a crystal structure of EndoS, which was published during this project [149].

3.1 Bioinformatic analysis of EndoS

As an initial step towards structural characterisation of EndoS, bioinformatic analysis of the EndoS sequence was carried out. This was done with an aim to identify regions of EndoS with sequence or structural homology to other well-characterised proteins. This approach was pursued with an aim to understand which parts of EndoS influence its enzymatic activity or its interactions with IgG. In addition, bioinformatics requires sequence information only, making it an attractive first step towards structural characterisation. Although the chitinase and LRR domains of EndoS had been identified by previous studies, the C-terminal domain remained largely uncharacterised [147].

Considering the size of EndoS (110 kDa), bioinformatic analyses were used to help determine domain boundaries for the design of constructs for the expression and purification of smaller EndoS fragments. Expressing individual domains or truncations of a larger protein can be advantageous for crystallography. Alleviating flexible linkers or isolating compact or distinct domains can potentially minimise the energy barrier which must be overcome in order for the protein to form an ordered crystal lattice. This is exemplified by the structural characterisation of the epithelial growth factor receptor, for which a number of structures of separate domains were determined first, then interpreted to form a model of the entire receptor [236].

3.1.1 RONN and PONDR

Analysis of the full-length EndoS sequence using the disorder prediction servers RONN and PONDR suggested that EndoS did not contain any large (>100 residues) regions of disorder, but rather it contained many shorter regions of potentially disordered residues (Fig. 3.1). Furthermore, regions of predicted order were not consistent across both methods, for instance, the first 100 residues of the N-terminus were predicted to be largely disordered when assessed using RONN, but the opposite when assessed using PONDR.

Overall, analysis with these prediction servers did not convincingly identify any likely domain boundaries or regions which should be truncated due to potential disorder.

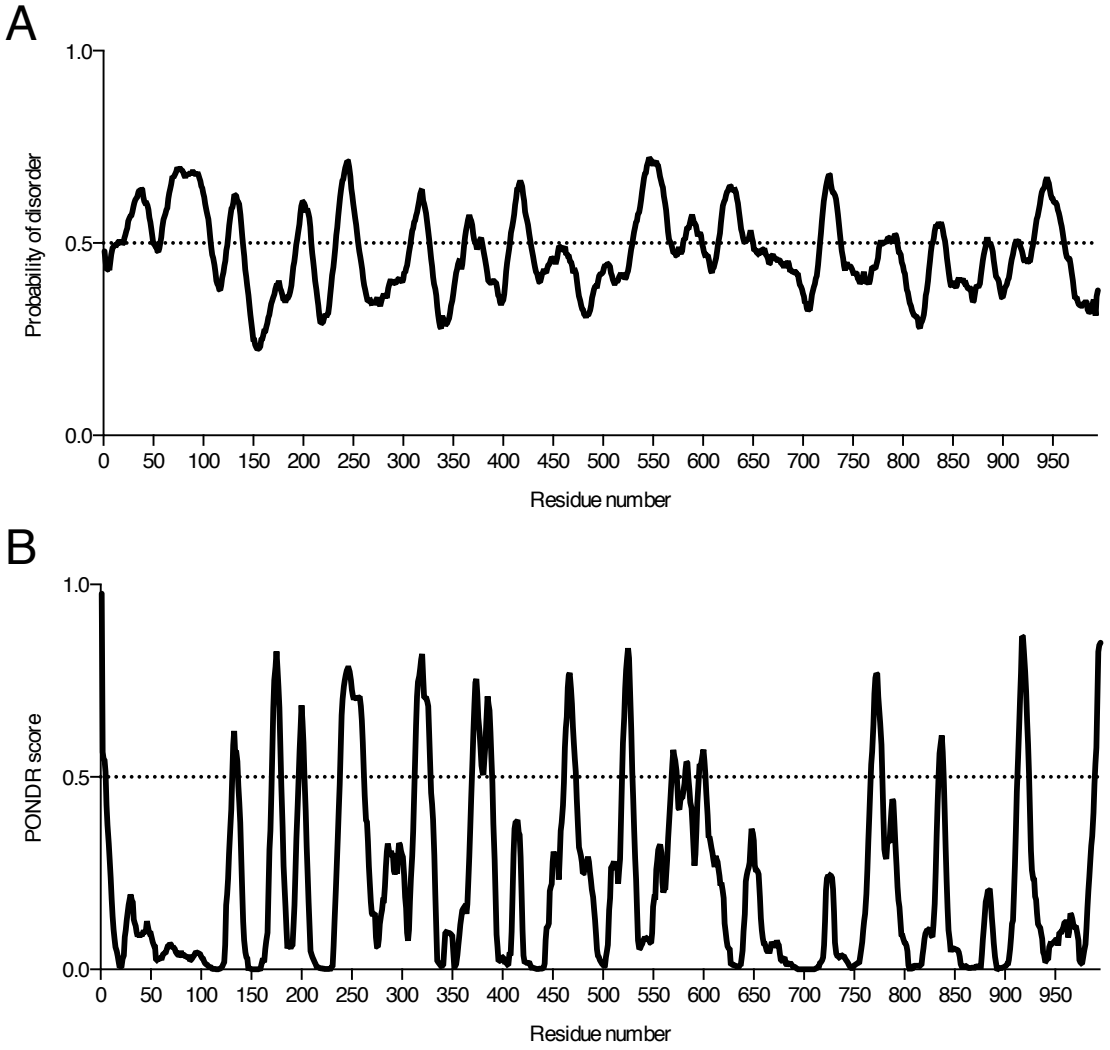


Figure 3.1: Disorder prediction using RONN and PONDR. (A) The probability of disorder for each residue was calculated using RONN [196] and was plotted against residue number for EndoS. (B) The PONDR score, which indicates the likelihood of disorder for a given residue, was calculated for each residue using the PONDR server [237, 238, 239]. These values were then plotted against residue number for EndoS. For both RONN and PONDR, scores > 0.5 (horizontal dotted line) indicate predicted regions of disorder.

3.1.2 FUGUE

Given that the initial disorder prediction methods did not provide any strong indications of potential domain boundaries or regions of disorder, homology based methods were employed. Using the full-length amino acid sequence of EndoS as well as smaller sections of the sequence (as detailed in Section 2.1), the programme FUGUE [195] revealed fur-

ther insight into the potential domain architecture of EndoS over and above a previously described homology model [147]. The results of these analyses are detailed in Table 3.2. FUGUE was able to identify homologous structures for the majority of the amino acid sequence where the top scoring solutions (judged by highest Z-score) were investigated. The top scoring solutions for the N-terminal chitinase domain and the LRR were identical to the PDB entries used for the previously published homology model of the N-terminal of EndoS [147].

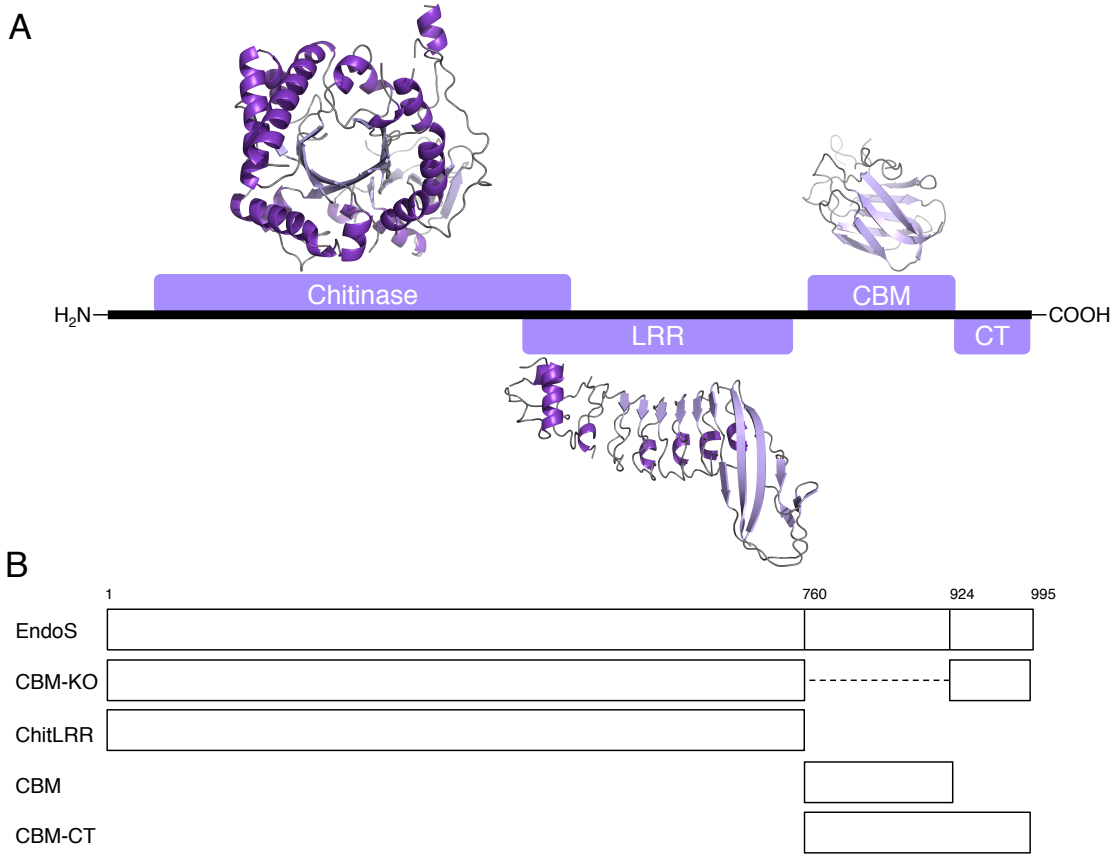


Figure 3.2: Predicted EndoS domain architecture and construct boundaries. (A) The 995 amino acid sequence of EndoS is depicted as a black line with domains identified by FUGUE analysis annotated as purple rectangles. The model of the chitinase domain spans residues 97–443 and is based on an alignment to PDB ID: 1CTN with a Z-score of 17.75. The LRR model includes residues 447–745 and is based on an alignment against PDB ID: 1M9S with a Z-score of 14.40. Finally, the CBM spans residues 771–924 where the model is based on an alignment to PDB ID: 1EUU and 1GOF and has a Z-score of 10.24. FUGUE was unable to detect any homologous structure to the CT region of EndoS from residues 925–995. (B) A schematic diagram details the constructs of EndoS designed for use in structural and functional studies.

Table 3.1: Bioinformatic analysis of EndoS and EndoS fragments using FUGUE.

Input EndoS sequence	PDB ID	Z-score	Protein name	Source	Aligned residues of EndoS
1–995	1CTN [240]	17.75	Chitinase A	*Serratia marcescens*	97–443
	1D2K [241]	17.70	Chitinase 1	*Coccidioides immitus*	58–565
	1E15 [242]	17.70	Chitinase B	*Serratia marcescens*	58–565
	1E9L [243]	17.70	Macrophage secretory lectin (Ym1)	*Mus musculus*	50–547
	3BXW [244]	15.22	Stabilin-1 Interacting Chitinase Like Protein	*Homo sapiens*	66–464
446–995	1M9S [245]	14.40	Internalin B	*Listeria monocytogenes*	447–995
	2Y5P [246]	13.43	Internalin B B-repeat	*Listeria monocytogenes*	752–828
	1M0Z [247]	10.20	Glycoprotein Ibα	*Homo sapiens*	446–693
	1A9N [248]	8.52	U2 small nuclear ribonucleoprotein A	*Homo sapiens*	456–644
	3O53 [249]	8.25	Internalin J	*Anopheles gambiae*	446–726
730–995	1EUU [250]	10.24	Neuraminidase	*Micromonospora viridifaciens*	760–924
	1GOF [251]	10.24	Galactose oxidase	*Dactylium dendroides*	751–923
	2XQX [252]	9.92	Endo-β-1,4-*N*-acetylglucosaminidase D CBM	*Streptococcus pneumoniae*	765–923
	4A3Z [253]	9.25	α-*N*-acetylglucosaminidase	*Clostridium perfringens*	765-925
	2Y8M [254]	8.48	*Ct*CBM62	*Clostridium thermocellum*	764–925
925–995	2DE3 [255]	4.64	Dibenzothiophene desulfurization enzyme B	*Rhodococcus erythropolis*	925–995
	3U4V [256]	3.03	Telomerase processivity factor Teb1 OB-A	*Tetrahymena thermophila*	925–995
	2ZSH [257]	2.94	Gibberellin receptor GID1	*Arabidopsis thaliana*	934–995
	1AUV [258]	2.90	Synapsin IA synC domain	*Bos taurus*	925–995
	3BTP [259]	2.69	Single-strand DNA-binding protein VirE2	*Agrobacterium tumefaciens*	940–967

In addition, FUGUE identified a region in the C-terminus of EndoS that has homology to a family of proteins called carbohydrate binding modules (CBMs) (for a review see [260]). Based on Z-score, the final 100 residues of EndoS were unable to be confidently assigned to any homologous structures using FUGUE. Despite the lack of significant hits for this region, all of the proteins that showed some homology comprised several short alpha helical stretches connected by flexible linkers.

During analysis, FUGUE uses both the input protein sequence and the corresponding homologous proteins it identifies to produce structural models of the query protein. Models generated from the top hits for the chitinase domain, the LRR domain and the putative CBM are depicted in Figure 3.2A.

3.2 Production of EndoS and EndoS fragments

Having identified regions of homology and potential domain boundaries using FUGUE, various constructs were designed for production of proteins which encompassed or excluded domains predicted by this bioinformatic analysis (Fig. 3.2B). The focus of these constructs was the previously undescribed C-terminal region of EndoS. Constructs which separated the predicted chitinase and LRR domains were not pursued due to previous reports of poor soluble expression of these domains in isolation [147]. Overall, the aim of producing these constructs was to isolate their role in glycoside hydrolase activity against IgG and to use them for structural analysis of EndoS.

3.2.1 Cloning

An existing construct for the recombinant expression of EndoS as an N-terminal GST-fusion was used for experiments regarding full-length EndoS [143]. To further investigate the putative CBM from EndoS, the results of the FUGUE analysis were used to guide the position of domain boundaries for the design of an isolated CBM construct. Further constructs were made to contain both the predicted CBM and the C-terminal region of EndoS (CBM-CT) and just the N-terminal chitinase and LRR portions of EndoS (ChitLRR). Finally, a CBM deletion mutant was cloned to ascertain the effect of removing this apparent domain from EndoS (CBM-KO). Details of exact domain boundaries can be found in 2.2.3

and are shown schematically in Figure 3.2B.

The constructs described were then cloned into bacterial expression vectors compatible with an *E. coli* expression system. To increase the options for success in downstream expression and purification protocols, all constructs were cloned as N-terminal GST-fusion proteins as well as C-terminally hexahistidine tagged constructs.

3.2.2 Expression

EndoS was expressed as an N-terminal GST fusion as previously described [143]. *E. coli* cells containing the pGEX-4T-1-*ndoS* plasmid were grown to an OD_{600} of 0.6–0.7 and were then induced with 1 mM IPTG for 4 hours. To check for expression, the cell lysate from *E. coli* cells before and after induction was analysed using SDS-PAGE and by western blot (Fig. 3.3). The cell lysate from the induced *E. coli* culture revealed a band at the correct molecular weight for GST-EndoS which was not present in the uninduced sample. Further analysis by western blot confirmed the presence of a GST-tag in this band, supportive of the fact this was the over-expressed GST-EndoS fusion. Two further bands, one at approximately 60 kDa and another at 25 kDa were also detected in the western blot.

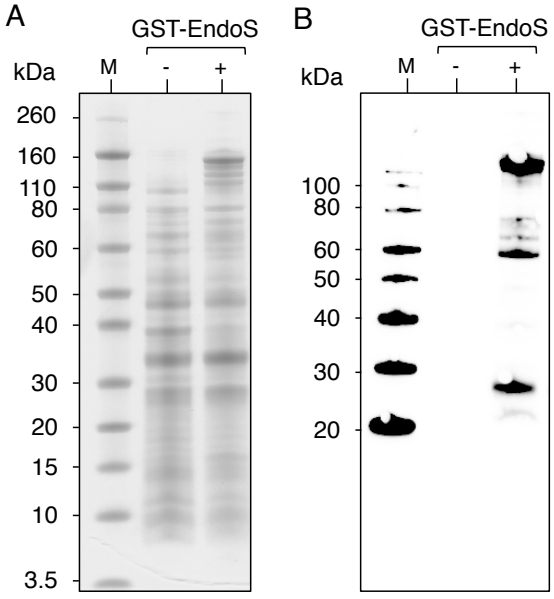


Figure 3.3: *E. coli* over-expression of GST-EndoS. (A) An SDS-PAGE gel and (B) an anti-GST western blot are shown for cell lysate fractions from an expression trial of GST-EndoS using *E. coli* cells. 'M', molecular weight marker; '–' lysate prior to induction with IPTG; '+', lysate 4 hours after induction with IPTG.

For the newly designed EndoS fragments, expression of the C-terminally hexahistidine tagged constructs was pursued first. These constructs were tested first because in contrast to GST-tagged proteins, these constructs do not, or are unlikely to require cleavage of the affinity tag for downstream characterisation, shortening the purification protocol by one step. A standard expression protocol for the relevant vector was followed where all EndoS fragments were able to be expressed in the soluble fraction. Expression was confirmed by SDS-PAGE and western blot analysis. However, due to a cloning error in the ChitLRR hexahistidine construct where a stop codon had been inserted before the hexahistidine tag, the GST-tagged construct was used to produce the ChitLRR protein. This N-terminal GST-fusion of ChitLRR was expressed in an identical manner to full-length EndoS. The SDS-PAGE and western blot analyses of the *E. coli* expression of these constructs is shown in Figure 3.4.

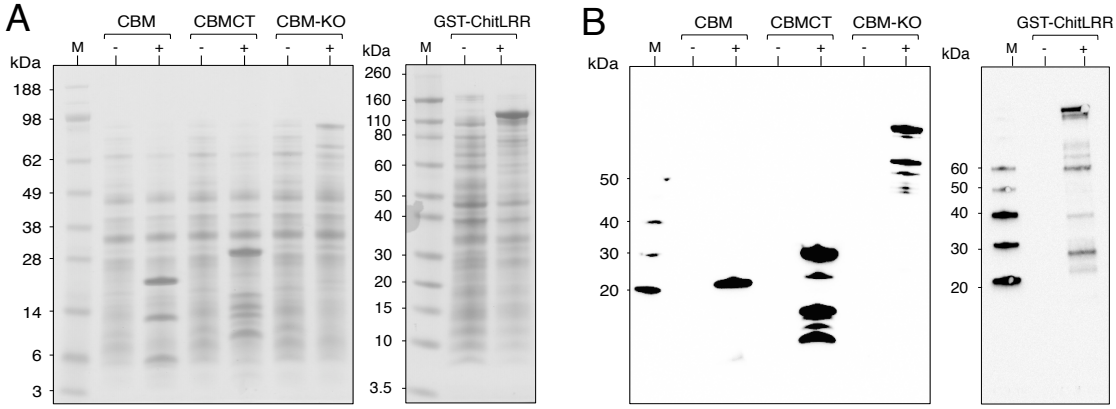


Figure 3.4: *E. coli* over-expression of EndoS fragments.(A) SDS-PAGE gels and (B) western blots are shown for cell lysate fractions from expression trials of these constructs using *E. coli* cells. The CBM, CBMCT and CBM-KO constructs were expressed as C-terminal hexahistidine tagged fusions where '–' indicates a sample taken immediately prior to the 24 hour incubation at 25 °C and '+' indicates a sample taken after 24 hours incubation at 25 °C. ChitLRR was expressed as an N-terminal GST fusion (GST-ChitLRR), '–' indicates lysate sampled prior to induction with IPTG; '+', indicates lysate sampled 4 hours after induction with IPTG. The western blots were probed with anti-His for CBM, CBMCT and CBM-KO and anti-GST for GST-ChitLRR. 'M', molecular weight marker.

3.2.3 Purification

Purification strategies for each of the EndoS constructs were developed so that each of the isolated domains could be purified to > 95% purity as judged by SDS-PAGE. Details of the purification strategy for each can be found below. In all cases, cells were treated with lysozyme before being mechanically lysed with a cell disruptor. Cell debris was removed

by centrifugation and further purification steps were carried out using this clarified cell lysate.

EndoS

Affinity purification of GST-EndoS was carried out using a GSTPrep^{TM} FF 16/10 column. Figure 3.5 shows the SDS-PAGE analysis of fractions from this affinity purification step. The glutathione-eluted protein fraction contained a band at a molecular weight consistent with that of GST-EndoS. This band was also consistent with the molecular weight of the over-expressed GST-tagged protein present in the the cell lysate as identified by western blot (Fig. 3.3B). There was an additional band in the eluted fraction at approximately 25 kDa, which was assumed to be free GST. The eluted fraction was incubated with thrombin in an effort to cleave the GST tag from the putative GST-EndoS fusion. SDS-PAGE analysis of the thrombin treated sample showed a reduction in molecular weight of both major bands in the sample. The reduction in the molecular weight of the putative GST-EndoS band upon treatment with thrombin was consistent with the loss of a GST tag.

The thrombin-cleaved protein was further purified using a SEC step, however, a number of contaminating bands remained present in the sample (data not shown). Although this previously established protocol of affinity purification followed by SEC was sufficient for enzyme activity studies [143], further purification was desired in order to obtain a sample of sufficient purity for structural and biophysical characterisation. Additional purification attempts involved a reverse affinity column approach, where the thrombin cleaved sample was again submitted to affinity purification over a GSTPrep^{TM} FF 16/10 column. This was done in an effort to bind contaminating proteins which originally coeluted with the putative GST-EndoS meanwhile allowing the now untagged construct to flow through the column. Although some contaminating proteins were removed, the yield from this method was poor and a large proportion of the putative EndoS protein was lost during this step (data not shown).

Ultimately, $> 95\%$ pure protein for subsequent structural and biophysical characterisation was obtained by including an ion exchange chromatography step. Analysis of the thrombin cleaved fraction using IEF revealed a number of bands over a range of isoelectric points

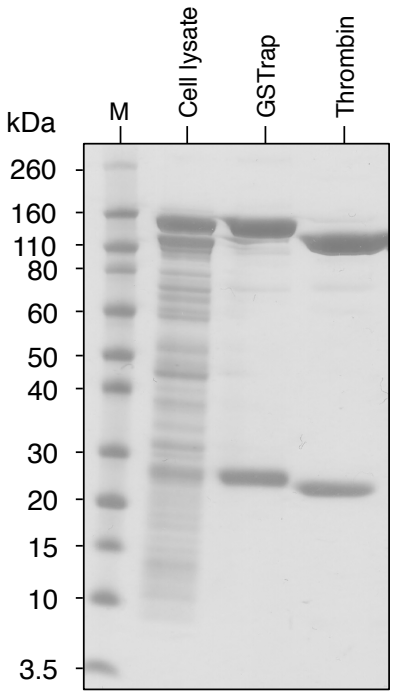


Figure 3.5: Affinity purification of GST-EndoS. SDS-PAGE analysis shows: 'Cell lysate', the total cell lysate taken from GST-EndoS expressing *E. coli* cells; 'GSTrap', the protein eluted from a GSTPrep^{TM} FF 16/10 column loaded with the cell lysate; 'Thrombin', the eluted protein sample from a GSTPrep^{TM} FF 16/10 column after a 24 hour incubation with bovine thrombin.

(Fig. 3.6B, 'Load'). At this stage it was unclear which band corresponded to EndoS which, based on protein sequence, had a predicted pI of 6.56 [206]. The thrombin-cleaved sample was loaded onto a HiPrep Q-HP 16/60 anion exchange column and was eluted using a gradient of increasing NaCl at pH 8.0. The AEC chromatogram along with IEF and SDS-PAGE analysis of the eluted fractions is shown in Figure 3.6. Putative EndoS protein eluted early in the gradient where it appeared as several isoforms. These isoforms all had basic pIs as judged by IEF (Fig. 3.6B). Fractions containing the prominent isoform (peak 3, Fig. 3.6A) were pooled and subjected to further purification by SEC (Fig. 3.7).

SEC of EndoS revealed that a small proportion of the protein formed dimers. Dimersation was disrupted by treatment with DTT but not by heat-denaturation suggesting dimerisation was due to disulfide bond formation between cysteine residues. The EndoS amino acid sequence contains three cysteine residues at positions 13, 15 and 986. Nevertheless, the majority of EndoS was eluted from the size exclusion column as a monomer and was shown to be $> 95\%$ pure by SDS-PAGE (Fig. 3.7B).

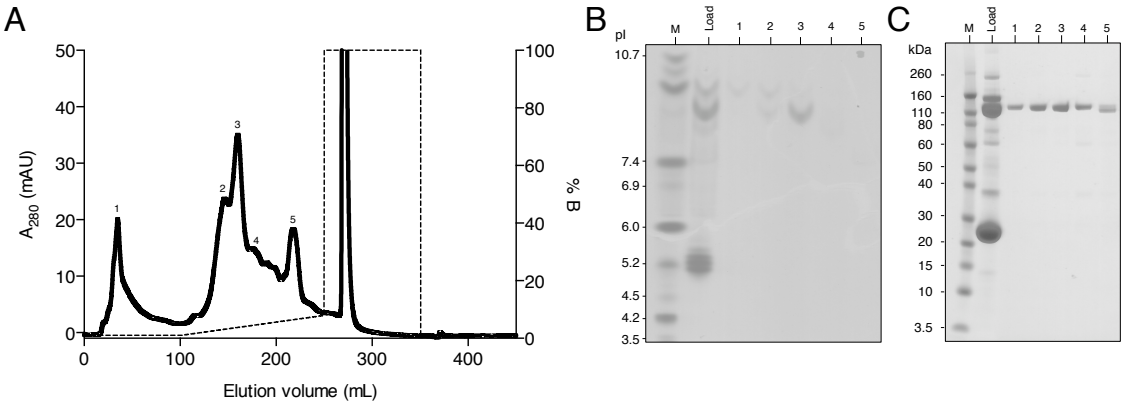


Figure 3.6: Purification of EndoS using AEC. (A) The chromatogram from an AEC purification step using a HiPrep Q-HP 16/60 column shows the UV absorbance (solid black line) and the concentration of buffer B (dotted black line) throughout the experiment. Several peaks are numbered which correspond to the fractions that were analysed by both (B) IEF and (C) SDS-PAGE. For both electrophoresis gels, 'Load' indicates the sample taken prior to AEC purification and 'M' denotes either the approximate pI marker (IEF) or the molecular weight marker (SDS-PAGE).

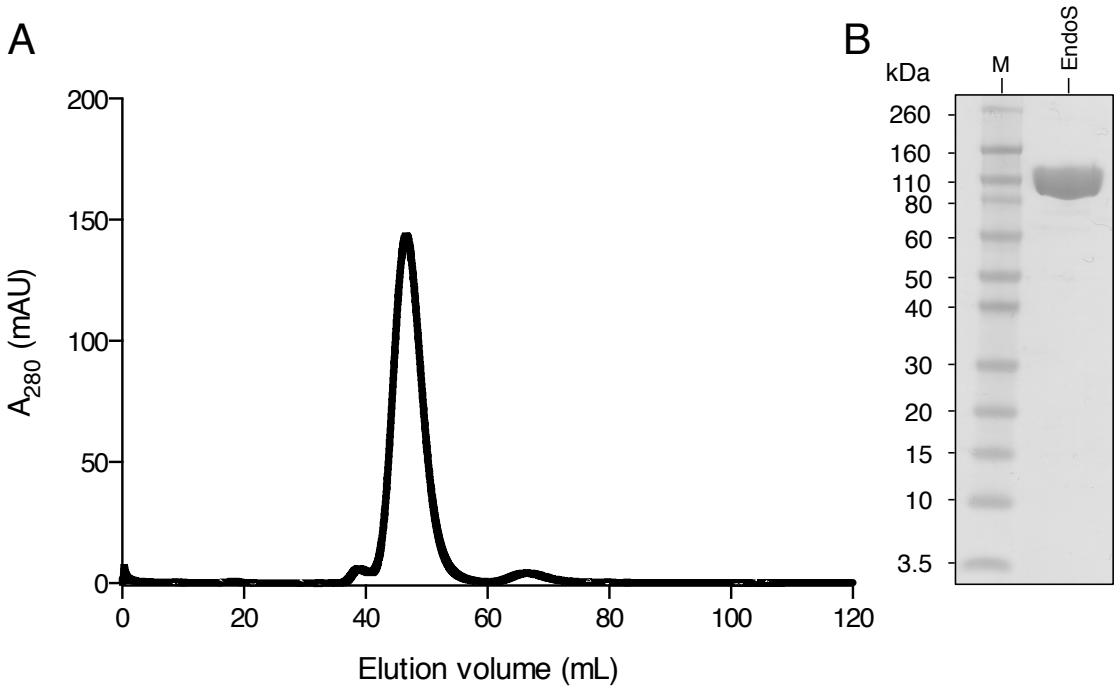


Figure 3.7: Purification of EndoS using SEC. (A) A chromatogram is shown for an SEC purification using a Superdex 75 16/60 column. EndoS elutes at approximately 50 mL. (B) SDS-PAGE analysis of the pooled peak EndoS fractions indicates the molecular weight and purity of the sample.

ChitLRR

The ChitLRR construct was designed as an N-terminal fragment of EndoS which encompasses residues 1–760 of EndoS. ChitLRR was expressed as an N-terminal GST fusion and was purified following a protocol nearly identical to that used for EndoS. The fractions from the affinity purification were analysed by SDS-PAGE and are shown in Figure 3.8. The glutathione-eluted fractions contained two bands, the heavier band corresponded to the correct molecular weight of GST-ChitLRR. Upon treatment with thrombin, the molecular weight of this band was reduced. The reduction in molecular weight of this band was consistent with the loss of the GST tag.

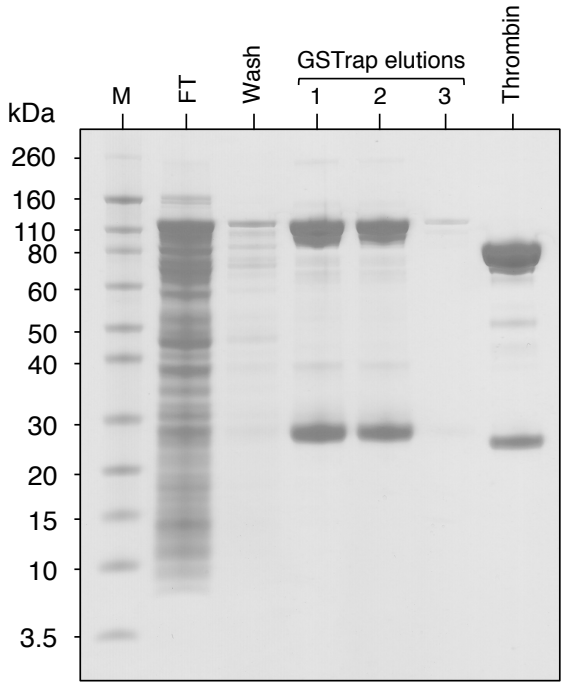


Figure 3.8: Affinity purification of GST-ChitLRR. SDS-PAGE analysis shows: 'FT', the flow-through of a GSTPrep^{TM} FF 16/10 column loaded with cell lysate from *E. coli* cells expressing GST-ChitLRR; 'Wash', the PBS wash fraction from a GSTrap column loaded with cell lysate; 'GSTrap elutions', three sequential 10 mL elutions (1, 2 and 3) from a GSTrap column using 10 mM glutathione, 50 mM Tris pH 8.0; 'Thrombin', the eluted protein sample from a GSTPrep^{TM} FF 16/10 column after a 24 hour incubation with bovine thrombin.

Given the success of AEC in the purification of full-length EndoS, AEC was used to further purify ChitLRR from contaminating proteins. ChitLRR was purified from the thrombin-cleaved sample using a HiPrep Q-HP 16/60 column and an increasing gradient of NaCl at pH 8.0 (Fig. 3.9A). SDS-PAGE analysis of the peak fractions from the AEC step indicated that a number of the contaminating protein species were removed by this method (Fig.

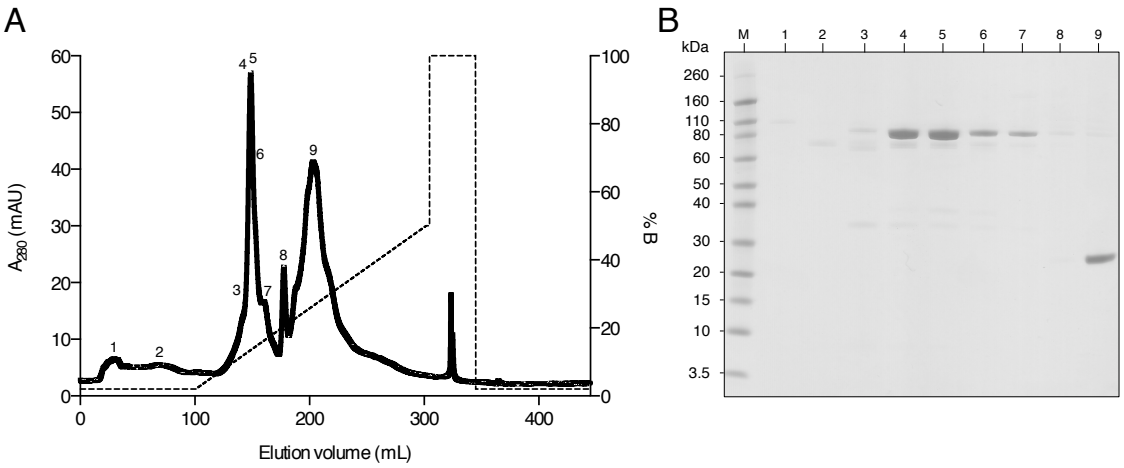


Figure 3.9: Purification of ChitLRR using AEC. (A) The chromatogram from an AEC purification step of ChitLRR using a HiPrep Q-HP 16/60 column shows the UV absorbance (solid black line) and the concentration of buffer B (dotted black line) throughout the experiment. Several peaks are numbered which correspond to the fractions that were analysed by (B) SDS-PAGE.

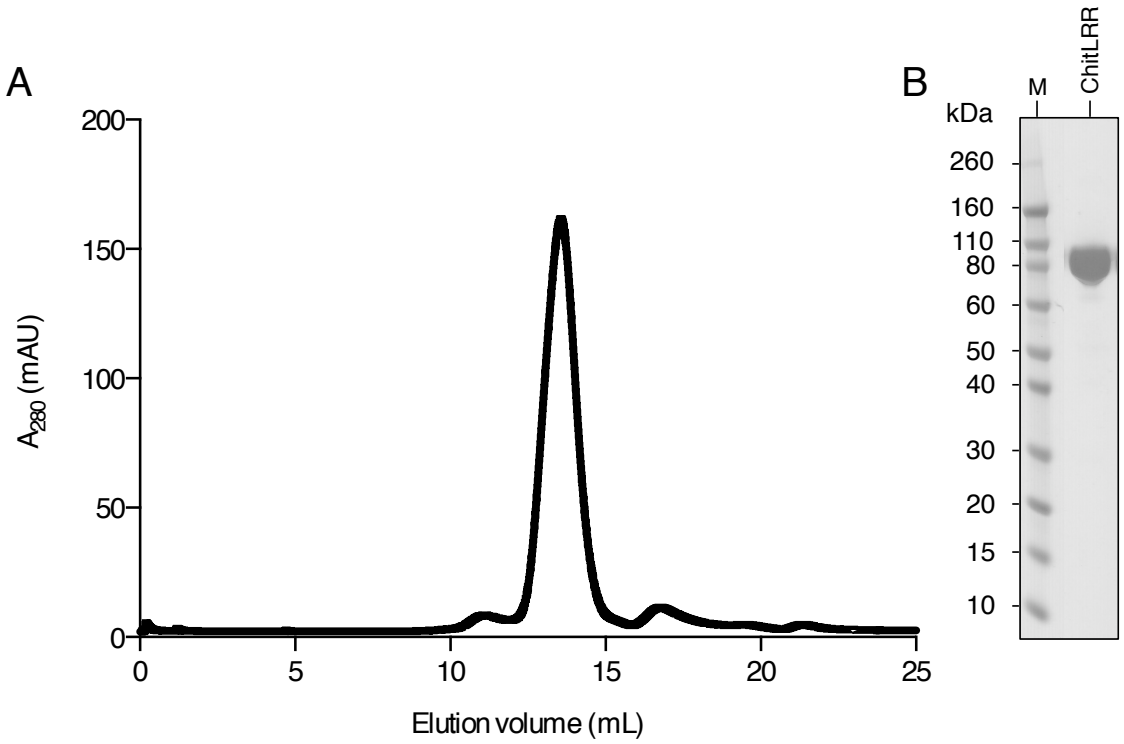


Figure 3.10: Purification of ChitLRR using SEC. (A) A chromatogram is shown for an SEC purification using a Superdex 200 10/300 GL column. ChitLRR elutes at approximately 14 mL. (B) SDS-PAGE analysis of the pooled peak ChitLRR fractions indicates the molecular weight and purity of the sample.

3.9B). After AEC, a SEC purification step was carried out. The SEC chromatogram and corresponding SDS-PAGE analysis of purified fractions is shown in Figure 3.10. The SDS-PAGE gel indicates sample purity of $> 95\%$.

CBM-KO

CBM-KO was expressed as a C-terminal hexahistidine tagged fusion and was purified by IMAC. SDS-PAGE analysis of the fractions collected during IMAC purification are shown in Figure 3.11. SDS-PAGE analysis of the eluted IMAC fractions revealed a dominant band corresponding to the correct molecular weight for CBM-KO as well as a number of other bands corresponding to contaminating proteins. To remove these contaminants, an AEC purification step was carried out following an identical protocol to that used for ChitLRR. CBM-KO did not bind the column at all and eluted during the washing stages of the AEC purification (Fig. 3.12). The CBM-KO containing fractions (1–5) from the AEC purification were pooled and applied to a SEC column (Fig. 3.13A). Although some of the contaminating bands were removed by AEC and SEC, the CBM-KO protein still had minor contaminants as shown by SDS-PAGE (Fig. 3.13B). In addition, it was apparent that CBM-KO formed dimers much like full-length EndoS. Again, these dimers could be disrupted with DTT indicating that dimerisation was due to formation of disulfide bonds involving some or all of the cysteine residues present at positions 13, 15 and 986. As this protein was not going to be used for crystallisation but simply for activity assays, further purification was not pursued.

CBM

The putative CBM of EndoS was expressed as a C-terminal hexahistidine tagged fusion. CBM was first purified by IMAC followed by SEC. An SDS-PAGE gel of the IMAC fractions is shown in Figure 3.14A. Notably, from the cell lysate of a 250 mL culture of *E. coli* cells, approximately 250 mg of protein was eluted from this initial affinity step. The eluted protein was analysed by SDS-PAGE and revealed a band at 20 kDa, consistent with the predicted molecular weight of CBM. The sample was of sufficient purity that the eluted fractions were applied directly to a SEC column without additional purification. An example SEC chromatogram for CBM is given in Figure 3.14B and the corresponding

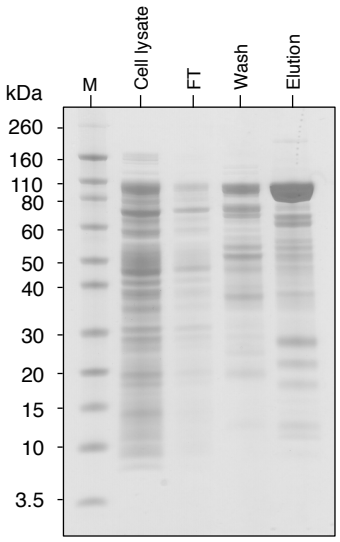


Figure 3.11: Purification of CBM-KO using IMAC. SDS-PAGE analysis of fractions taken throughout IMAC purification show: 'Cell lysate', the total cell lysate from *E. coli* cells expressing a C-terminal hexahistidine tagged CBM-KO fusion; 'FT', the flow-through from the HisPur^{TM} Ni-NTA column; 'Wash', the wash fraction from the HisPur^{TM} Ni-NTA column; 'Elution', the protein eluted using PBS with 300 mM imidazole.

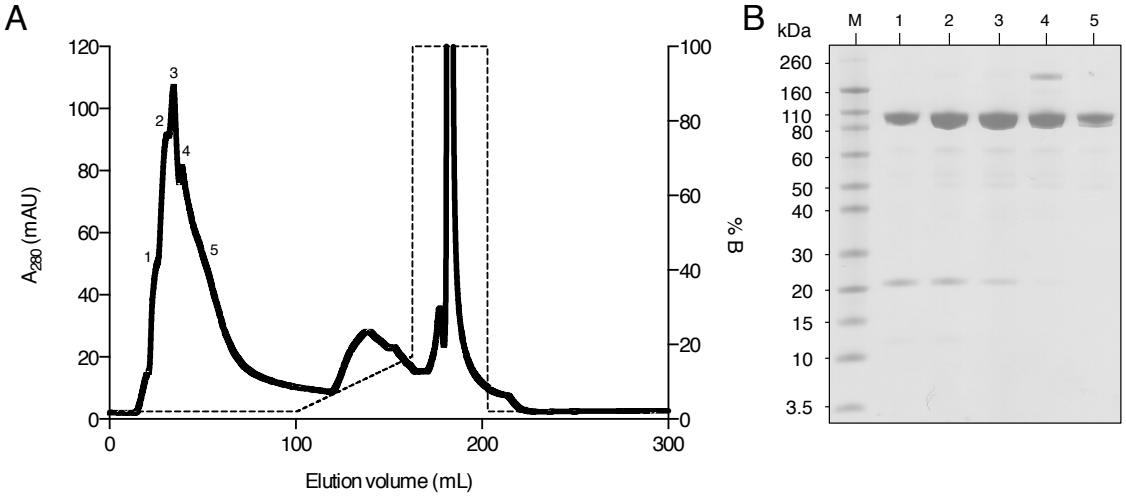


Figure 3.12: Purification of CBM-KO using AEC. (A) The chromatogram from an AEC purification step of CBM-KO using a HiPrep Q-HP 16/60 column shows the UV absorbance (solid black line) and the concentration of buffer B (dotted black line) throughout the experiment. Several peaks are numbered which correspond to the fractions that were analysed by (B) SDS-PAGE.

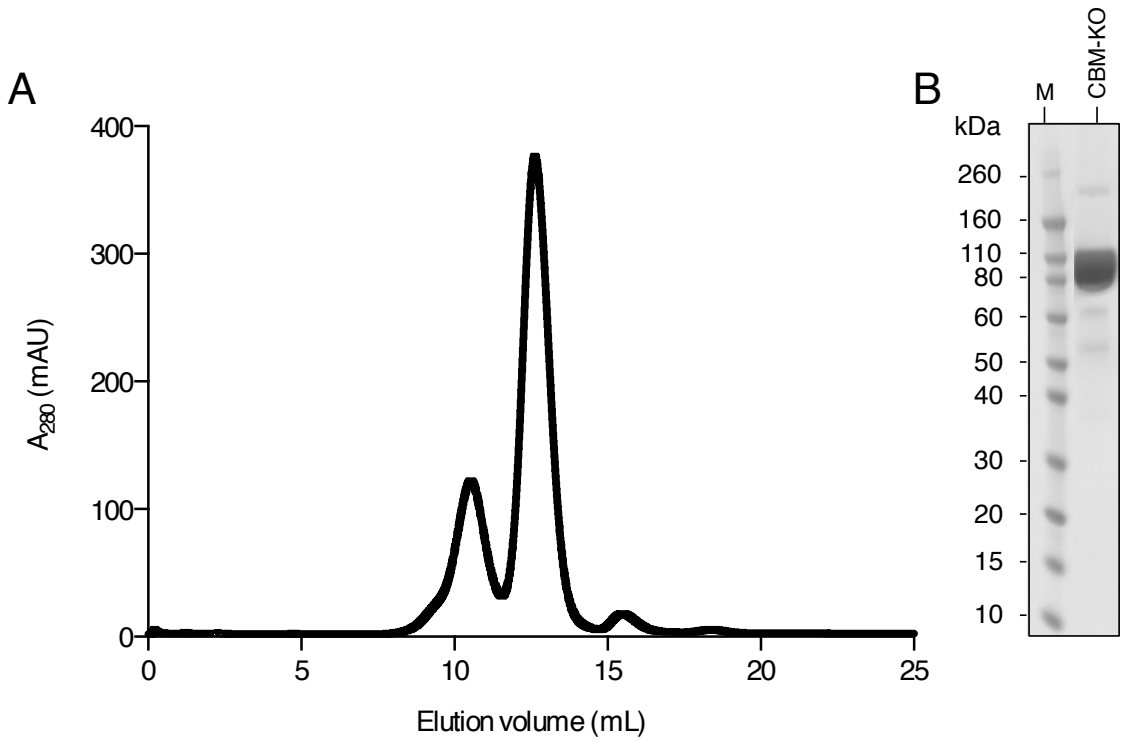


Figure 3.13: Purification of CBM-KO using SEC. (A) A chromatogram is shown for an SEC purification using a Superdex 200 10/300 GL column where CBM-KO elutes at approximately 13 mL. (B) SDS-PAGE analysis of the pooled fractions from the peak at 13 mL indicates the molecular weight and purity of the CBM-KO sample.

SDS-PAGE gel of the pooled peak fractions is shown in Figure 3.14C. SEC was successful at removing contaminating proteins from the sample as well as higher molecular weight aggregates of CBM protein.

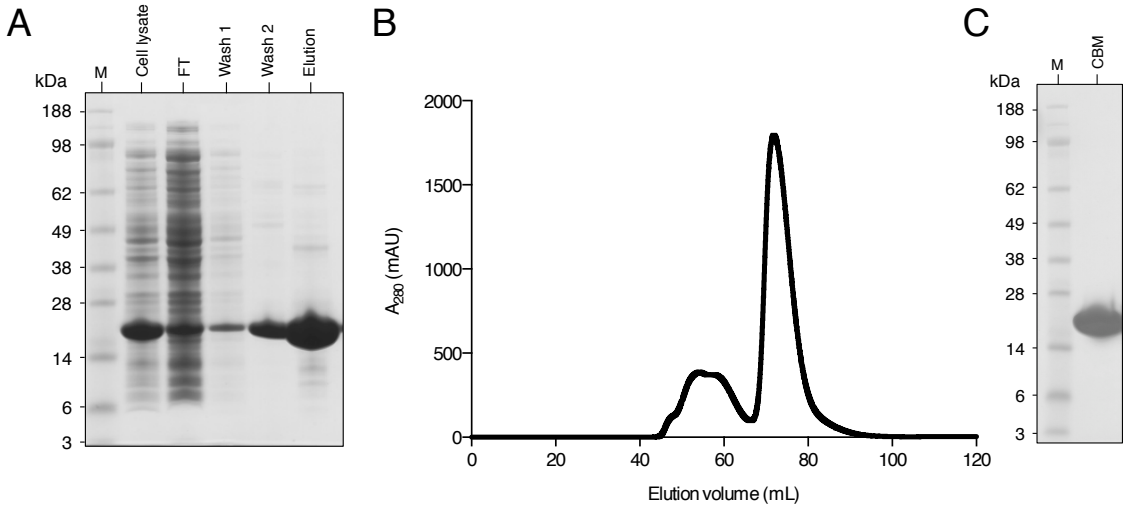


Figure 3.14: Purification of CBM using IMAC and SEC. (A) Purification of CBM using IMAC. SDS-PAGE analysis of fractions taken throughout IMAC purification show: 'Cell lysate', the total cell lysate from *E. coli* cells expressing a C-terminal hexahistidine tagged CBM fusion; 'FT', the flow-through from the HisPur^{TM} Ni-NTA column; 'Wash', the wash fraction from the HisPur^{TM} Ni-NTA column; 'Elution', the protein eluted using PBS with 300 mM imidazole. (B) SEC was carried out using a HiPrep Superdex 75 16/60 column. The peak at 75 mL corresponds to CBM protein. (C) SDS-PAGE analysis of the pooled fractions from the peak at 75 mL indicates the molecular weight and purity of the CBM sample.

CBM-CT

The CBMCT construct was expressed and purified in an identical fashion to CBM. SDS-PAGE analysis of the IMAC fractions, the SEC chromatogram and an SDS-PAGE gel showing the pooled purified protein are shown in Figure 3.15. The protein eluted from Ni-NTA beads using imidazole contained a 28 kDa band when analysed by SDS-PAGE. This molecular weight was in accordance with the predicted size of CBMCT. Similar to CBM, CBMCT protein gave high yields of protein (approximately 100 mg per 250 mL of cell culture). The SEC chromatogram shows CBMCT running as a monomer although dimerisation of CBMCT was observed. Much like ChitLRR, CBM-KO and the full-length EndoS protein, dimerisation was caused by disulfide bond formation as CBMCT contains a single cysteine residue at the position corresponding to residue 986 of the full-length protein.

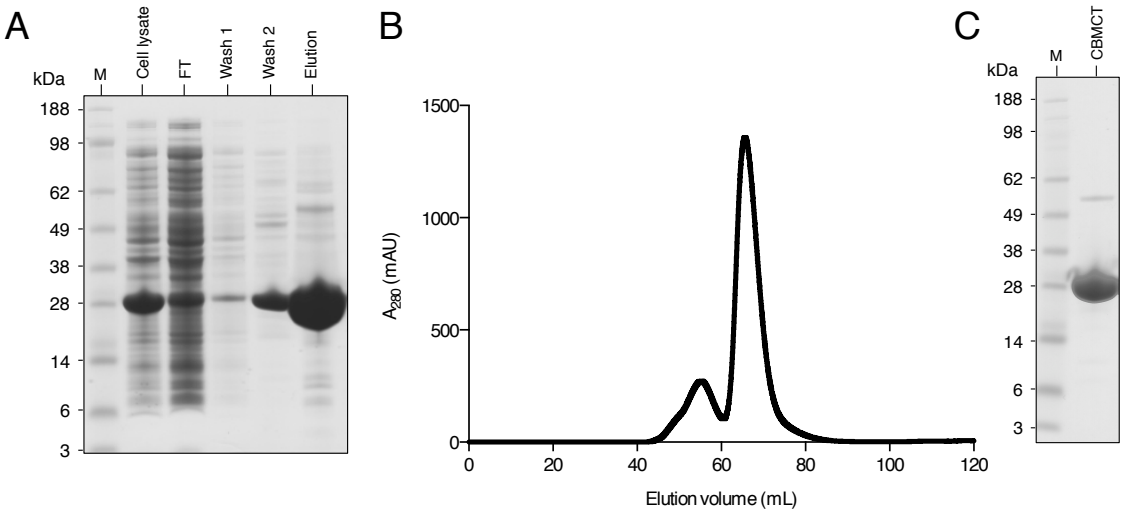


Figure 3.15: Purification of CBMCT using IMAC and SEC. Purification of CBMCT using IMAC. SDS-PAGE analysis of fractions taken throughout IMAC purification show: 'Cell lysate', the total cell lysate from *E. coli* cells expressing a C-terminal hexahistidine tagged CBMCT fusion; 'FT', the flow-through from the HisPur^{TM} Ni-NTA column; 'Wash', the wash fraction from the HisPur^{TM} Ni-NTA column; 'Elution', the protein eluted using PBS with 300 mM imidazole. (B) SEC was carried out using a HiPrep Superdex 200 16/60 column. The peak at 65 mL corresponds to CBMCT protein. (C) SDS-PAGE analysis of the pooled fractions from the peak at 65 mL indicates the molecular weight and purity of the CBMCT sample.

3.3 Production of EndoS glycoprotein substrates

In order to carry out structural studies of EndoS in complex with its substrate IgG, production of substrate proteins was required. Two substrates were produced, IgG1 Fc and 2G12 IgG.

Considering the flexibile multidomain nature of full-length IgG molecules, the smaller, readily crystallisable isolated IgG1 Fc domain was chosen as the primary substrate for structural work with EndoS. Fc was expressed as a C-terminal hexahistidine tagged protein using HEK 293T cells. Fc was purified from the supernatant of these cells by IMAC and SEC. A SEC chromatogram, along with an SDS-PAGE gel of the pooled peak SEC fractions, is shown in Figure 3.16A.

The full-length 2G12 IgG antibody was also prepared to be used as a substrate. 2G12 is an antibody which targets the human immunodeficiency virus type 1 (HIV-1) glycan shield [261, 262]. 2G12 shows a unique conformation of the Fab domains of IgG where the so-called 'V_H domain-swapping' of the Fab arms results in an 'I' shaped molecule instead of the canonical 'Y' shape of IgG [263]. This antibody was prepared with the rationale

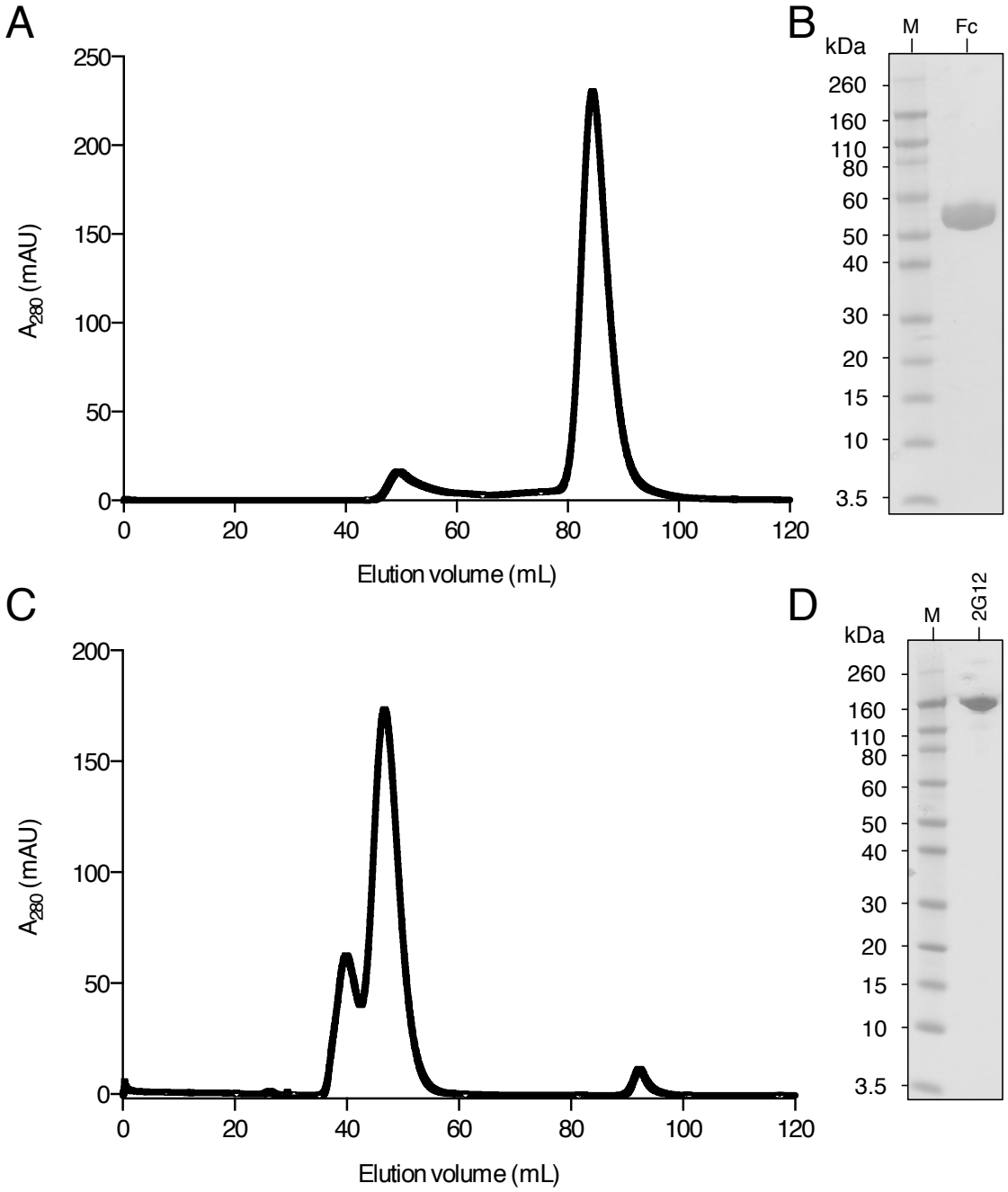


Figure 3.16: Purification of IgG1 Fc and 2G12 IgG using SEC. The SEC chromatograms and corresponding non-reducing SDS-PAGE gels of peak fractions for (A, B) Fc and (C, D) 2G12 are shown. Fc was purified using a HiPrep Superdex 200 16/60 column. The homodimeric Fc protein eluted at approximately 85 mL. The SDS-PAGE gel shows the glycosylated dimer of Fc at approximately 55 kDa. 2G12 was also purified using a Superdex 200 HiPrep 16/60 column. The peak at 40 mL corresponds to aggregates of protein in the 2G12 sample. The 2G12 homodimer of heterodimers eluted at approximately 47 mL. The SDS-PAGE gel shows glycosylated 2G12 at a molecular weight of approximately 150-160 kDa with some minor contaminants.

that the Fab arms are likely to be rigidified by this domain-swapping. This meant that a full-length IgG could be used in EndoS-substrate cocrystallisation experiments without introducing the amount of flexibility normally associated with IgG.

2G12 was expressed by cotransfecting HEK 293T cells with two plasmids, one encoding the heavy chain and one encoding the light chain. The expressed 2G12 protein was then purified from cell culture supernatant using a Protein A column. Both the expression and Protein A purification were carried out by Dr Camille Bonomelli (Glycobiology Insitute, University of Oxford). The resultant protein was purified by SEC. A proportion of the 2G12 sample formed aggregates eluting at approximately 40 mL. The remainder of the sample eluted at approximately 47 mL corresponding to the anticipated molecular weight of 2G12 (150 kDa). This peak was pooled and concentrated for use in further experiments. The chromatogram for this final SEC step is shown in Figure 3.16C accompanied by the SDS-PAGE analysis of the pooled 2G12 peak in Figure 3.16D.

3.4 Production of EndoS:substrate complexes

To obtain a stable enzyme:substrate complex, two approaches were taken, both with the common goal of capturing the interaction of these two components before hydrolysis of the IgG Fc glycan. The first approach involved using a catalytically inactive EndoS construct where the critical glutamate residue at position 235 was mutated to a glutamine. This reportedly inactive mutant [147] was cloned and expressed in an identical manner to the wild-type enzyme and its ability to form a stable complex was first tested with Fc. The second approach involved using catalytically active EndoS in complex with the reportedly EndoS resistant Fc glycoform, $Man_5GlcNAc_2$.

Neither of these combinations formed stable complexes as judged by SEC. This was attributed to the catalytic activity of EndoS observed in both circumstances. Further investigation of EndoS glycan specificity revealed that EndoS can cleave $Man_5GlcNAc_2$ as is discussed later in Chapter 5. As for the E235Q mutant, this was convincingly shown to have catalytic activity by MS, HPLC and SDS-PAGE gel analysis of the reaction products (data not shown). Personal communication with Dr Mattias Collin (Lund University, Sweden) confirmed that catalytic activity of this mutant has been observed based on the pH of

the reaction by their laboratory also, to which the original publication can be attributed. Considering that crystallisation trials often utilise a vast array of different pHs and that a stable complex is necessary for structural characterisation, further catalytic mutants were designed and tested in order to find one suitable for use in studies of the EndoS:substrate complex.

3.4.1 EndoS catalytic mutants

The chitinase domain of EndoS shows significant homology to glycoside hydrolase family 18 (GH18) domains based on bioinformatic analysis carried out here and in previous publications [147]. GH18 chitinase domains, including EndoS, contain the common active site motif 'DXDXE', where X can be substituted for any amino acid residue [264]. To identify a suitable catalytically inactive mutant, these acidic active site residues were mutated to the hydrophobic non-polar amino acid leucine. Leucine was chosen as opposed to alanine in an effort to maintain the size of the amino acid side chain and to minimise potential steric effects of mutagenesis whilst still abolishing the charge associated with the wild-type side chains. Combinations of both single and double mutations of Asp231, Asp233 and Glu235 were made using site-directed mutagenesis. All of the mutants were expressed as N-terminal GST fusions in an identical fashion to wild-type EndoS. SDS-PAGE analyses of the fractions collected during affinity purification of each of the mutants are shown in Figure 3.17.

The acitivty of the mutant GST-EndoS fusions was tested against recombinantly expressed IgG1 Fc protein. EndoS mutants were incubated with Fc for 16 hours at 37 °C and then the products of the reaction were analysed by SDS-PAGE. The molecular weight of the Fc protein band was used to indicate whether Fc maintained its glycan or not in the presence of the EndoS mutants. In this assay, a decrease in molecular weight of Fc by approximately 3 kDa corresponds to the loss of the *N*-linked glycan. SDS-PAGE gel shift analysis showed that neither the EndoS^{E235L} mutant nor either of the double mutants (EndoS$^{D231L/D233L}$ or EndoS$^{D233L/E235L}$) were able to cause a decrease in the molecular weight of Fc. This is consistent with these mutants having no glycoside hydrolase activity against IgG Fc. The reaction of EndoS^{D233L} with Fc gave two bands on the SDS-PAGE gel, corresponding to both the deglycosylated and glycosylated forms of Fc. The EndoS^{D231L} mutant produces

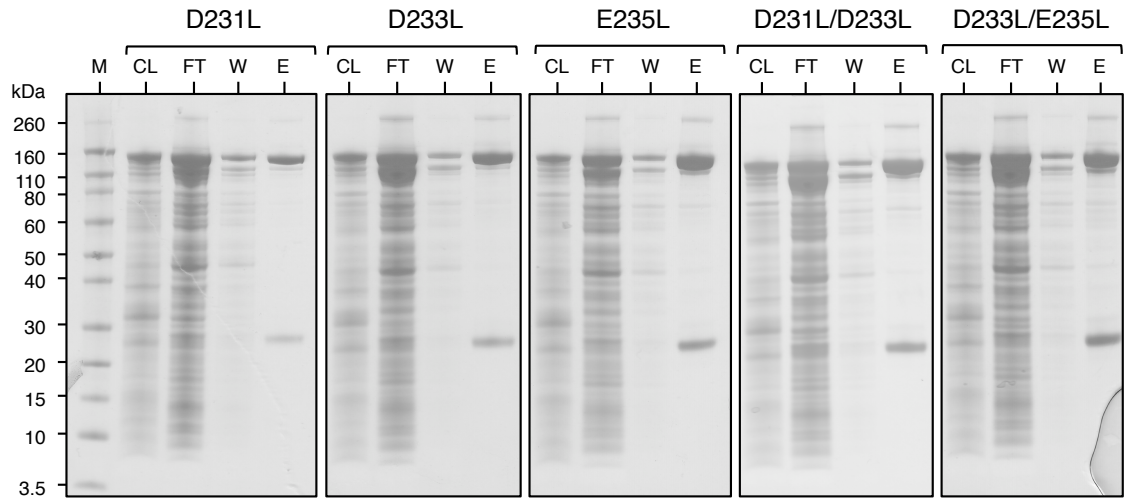


Figure 3.17: Affinity purification of EndoS catalytic mutants. The EndoS active site mutants D231L, D233L, E235L, D231L/D233L and D233L/E235L were expressed as N-terminal GST fusions in *E. coli* cells. Cells expressing these proteins were lysed before cell lysate was centrifuged to remove cell debris. The supernatants were then subjected to affinity purification using Glutathione Sepharose 4B beads. SDS-PAGE analysis of the fractions from this affinity purification are shown for each mutant. 'CL', cell lysate; 'FT', flow through, 'W', wash fraction; 'E', glutathione elution. All samples were reduced and denatured prior to SDS-PAGE.

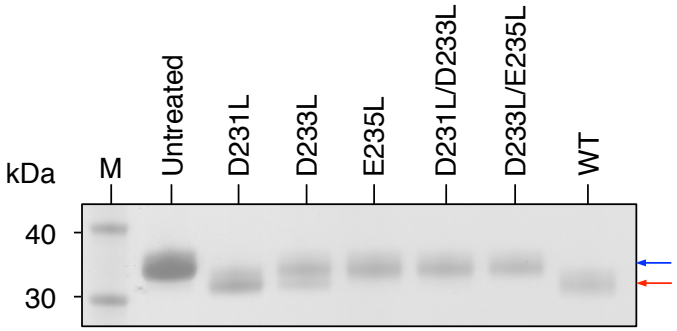


Figure 3.18: Testing activity of EndoS catalytic mutants against IgG1 Fc. EndoS catalytic mutants were added to Fc in a 1:10 molar ratio. The reactions were incubated for 16 hours at 37 °C and then subjected to reducing SDS-PAGE. The bands correspond to monomeric Fc fragments. The molecular weight shift caused by deglycosylation is indicated by the arrows at the right of the gel. The blue arrow indicates the larger molecular weight bands which correspond to glycosylated Fc, the red arrow indicates the smaller molecular weight bands which correspond to deglycosylated Fc.

only the lower molecular weight band of Fc, consistent with this mutant maintaining catalytic activity against IgG1 Fc glycans. Given these results, the EndoS^{E235L} mutant was used for further work towards producing EndoS:substrate complexes. This single mutant was chosen over the double mutants in an effort to minimise the potential structural effects of the inactivating mutations on the protein.

3.4.2 EndoS-Fc complex

EndoS^{E235L} was purified using an identical protocol to that used to purify wild-type EndoS (Section 3.2.3). Using analytical FPLC, the ability of EndoS^{E235L} to form a stable complex with Fc was assessed (Fig. 3.19). A sample of EndoS^{E235L} was applied to a Superdex 200 10/300 GL at room temperature and was eluted with 10 mM HEPES pH 8.0, 150 mM NaCl. The same was repeated for a sample of Fc. The elution volume corresponding to the inflection point of the A_{280} peak was noted for each protein. EndoS^{E235L} eluted at 12.79 mL and Fc eluted at 15.1 mL. EndoS^{E235L} and Fc were combined at a 1:1 molar ratio. This mixture was then loaded directly onto the Superdex 200 10/300 GL column. The protein mixture eluted predominantly as one species where the inflection point of the major peak was seen at an elution volume of 12.53 mL. This peak was symmetrical with two smaller peaks eluting immediately prior and after. SDS-PAGE analysis of the major peak revealed that the peak contained both EndoS^{E235L} and Fc. This peak was consistent with the formation of an EndoS^{E235L}Fc complex with 1:1 stoichiometry. Large scale preparation of this apparent EndoS^{E235L}Fc complex was carried out to provide samples suitable for crystallisation and SAXS experiments.

3.4.3 EndoS-2G12 complex

The same analytical SEC experiment detailed in Section 3.4.3 was carried out using 2G12 in the place of Fc (Fig. 3.20). The 2G12 eluted at 13.00 mL. The mixture of EndoS^{E235L} and 2G12 eluted gave a large single peak at 12.23 mL. However, this peak was not completely symmetrical and had a shoulder on its trailing edge. SDS-PAGE analysis of the peak fractions revealed that the peak contained both EndoS^{E235L} and 2G12. The trailing shoulder of the peak contained EndoS^{E235L} only.

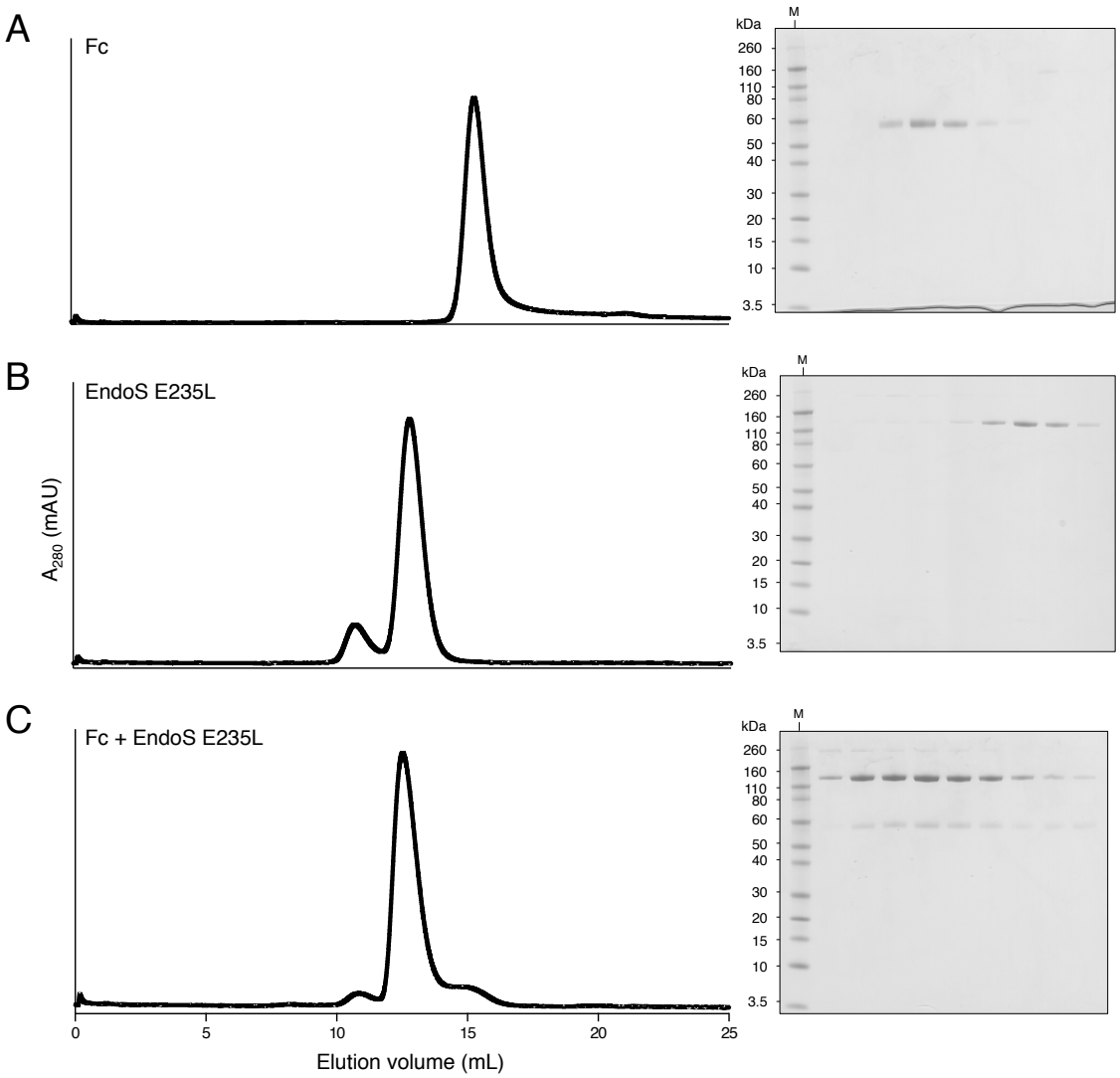


Figure 3.19: Analytical SEC of EndoS^{E235L} and Fc. The analytical SEC chromatograms and corresponding SDS-PAGE gels are shown for (A) Fc, (B) EndoS^{E235L} and (C) a 1:1 molar ratio of EndoS^{E235L} and Fc. Each protein was subjected to SEC using a Superdex 200 10/300 GL column and was eluted using 10 mM HEPES pH 8.0, 150 mM NaCl. SDS-PAGE gels were run under non-reducing conditions to preserve the integrity of the Fc homodimer.

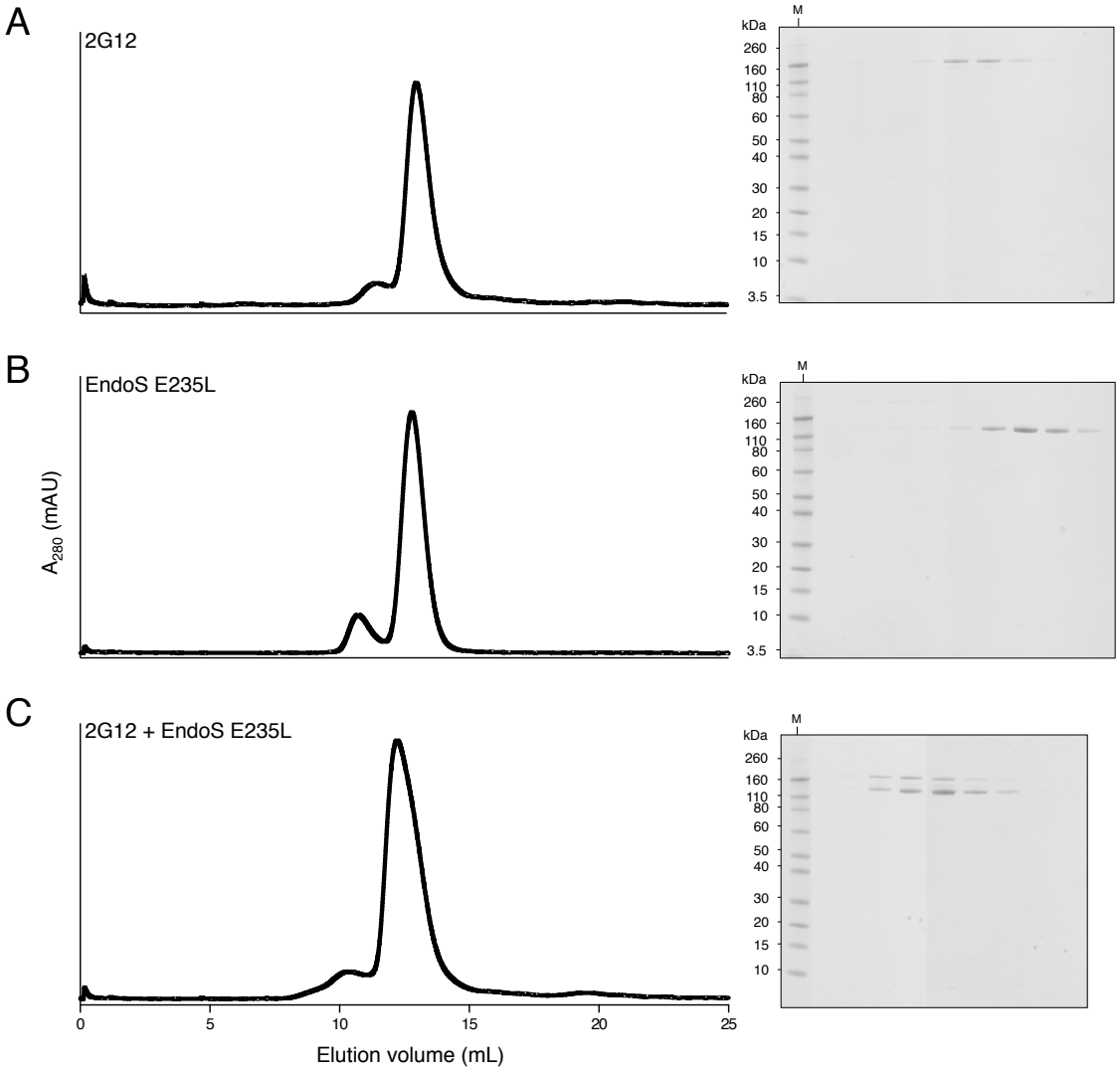


Figure 3.20: Analytical SEC of EndoS^{E235L} and 2G12. The analytical SEC chromatograms and corresponding SDS-PAGE gels are shown for (A) 2G12, (B) EndoS^{E235L} and (C) a 1:1 molar ratio of EndoS^{E235L} and 2G12. Each protein was subjected to SEC using a Superdex 200 10/300 GL column and was eluted using 10 mM HEPES pH 8.0, 150 mM NaCl. SDS-PAGE gels were run under non-reducing conditions.

3.5 Circular dichroism

Having expressed and purified various EndoS fragments as well as potential EndoS-Fc and EndoS-2G12 complexes, initial biophysical experiments were carried out in an effort to experimentally validate the bioinformatics results and to gain further insight into the structure of EndoS.

As an initial approach to determining structural information, CD was used to assess the foldedness of the full-length EndoS protein and various EndoS fragments. CD was used as it can provide a useful estimate of the secondary structure of proteins and often requires only small amounts of protein in solution [265].

The CD spectra recorded for both full-length EndoS, EndoS^{E235L} and the CBM and CB-MCT fragments of EndoS, are described below. Here, the programme CDSSTR [266, 267, 268] was used to deconvolute the spectra in order to predict the proportions of various secondary structure components present in each protein.

3.5.1 EndoS and EndoS catalytic mutants

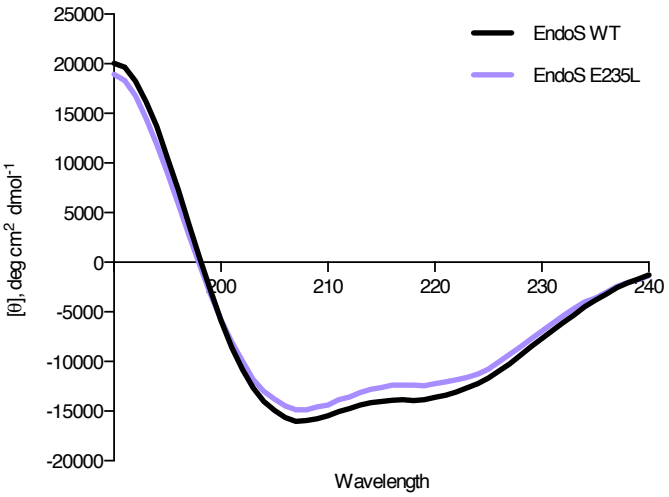


Figure 3.21: CD spectra of EndoS and EndoS^{E235L}. Data are shown for CD measurements taken for EndoS^{E235L} (mauve) and EndoS (black) between 190-240 nm.

To gain an overall picture of the foldedness of EndoS in solution and to ascertain whether the E235L mutation causes any major changes to secondary structure, CD analysis was carried out for both wild-type EndoS and EndoS^{E235L}. The spectra for both proteins are shown in Figure 3.21. Table 3.2 details the values for the secondary structure components

Table 3.2: Analysis of CD spectra from EndoS and EndoS^{E235L} using CDSSTR*.

Sample	Helix	Strand	Turns	Unordered	NRMSD†
EndoS	40	11	14	35	0.014
EndoS^{E235L}	37	14	15	35	0.010

*Deconvoluted using CDSSTR [266] accessed via the DICHROWEB server [211] and represented as a percentage of total protein secondary structure.
†Normalised root mean square deviation between experimental spectrum and spectrum calculated using CDSSTR parameters.

of EndoS and EndoS^{E235L} obtained after deconvolution of the spectra using CDSSTR.

Deconvolution of the CD spectrum of EndoS suggests a large proportion of the protein is unstructured ($\sim$ 35%), however, there is still a significant proportion of the protein that appears to form both α-helical and β-sheet structures. The EndoS^{E235L} spectrum is not significantly different to that of the wild-type EndoS protein where the calculated proportions of secondary structure elements differ by a maximum of 3%.

3.5.2 EndoS fragments

Investigation of the secondary structure of CBM and CBMCT using CD supported the bioinformatic prediction of a primarily β-sheet CBM in the C-terminal of EndoS (Fig. 3.22 and Table 3.3). The CD analysis of CBMCT fusion indicated the presence of α-helical secondary structure in addition to the predicted β-sheet conformation seen for the CBM construct alone.

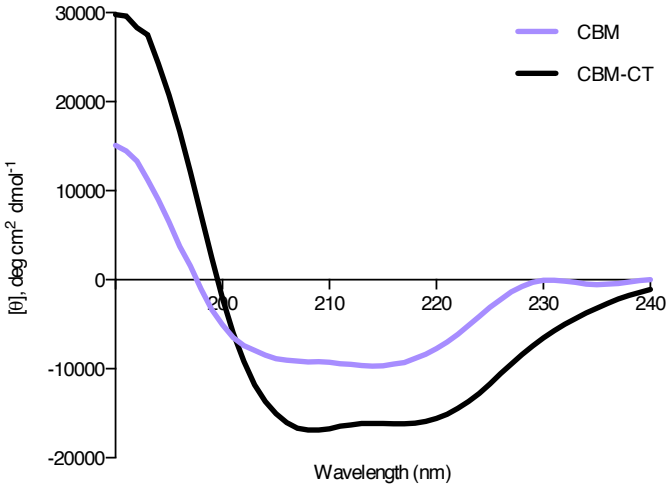


Figure 3.22: CD spectra of CBM and CBMCT. Data are shown for CD measurements taken for CBM (mauve) and CBMCT (black) between 190-240 nm.

Table 3.3: Analysis of CD spectra from CBM and CBMCT using CDSSTR*.

Sample	Helix	Strand	Turns	Unordered	NRMSD†
CBM	5	44	23	28	0.019
CBMCT	40	20	16	24	0.015

*Deconvoluted using CDSSTR [266] accessed via the DICHROWEB server [211] and represented as a percentage of total protein secondary structure.
†Normalised root mean square deviation between experimental spectrum and spectrum calculated using CDSSTR parameters.

3.6 Crystallisation

X-ray crystallography has the potential to give atomic level details of proteins, however, this method often requires large amounts of pure protein. Therefore, a protocol was developed to purify EndoS to $> 95\%$ homogeneity as judged by SDS-PAGE. This was achieved using a combination of affinity puficiation, AEC and SEC (Section 3.2.3).

Before setting up crystallisation trials, a pre-crystallisation test was carried out to ascertain a suitable protein concentration for these experiments. EndoS was highly soluble at room temperature, and did not precipitate readily in any of the test conditions even at concentrations of 30 mg mL^{-1}. At this point, a pre-crystallisation test was also carried out at 4 °C which revealed that EndoS was significantly less soluble at low temperatures. Crystallisation trials were set up at both 4 °C and 20 °C at concentrations of 5, 10 and 20 mg mL^{-1} in an effort to find suitable crystallisation conditions.

An initial hit was observed in a condition from the JCSG-*plus* HT screen [269] set up at 4 °C with 5 mg mL^{-1} EndoS. Crystalline needles were observed in condition A11 which comprises 0.2 ammonium dihydrogen phosphate, 0.1 M Tris pH 8.5, 50% (v/v) 2-methyl-2,4-pentandiol (Fig. 3.23A). Although these crystals were small (< 1 μm wide), a collection of these needles were harvested to check for protein-like diffraction before optimisation of these putative EndoS crystals was pursued.

A collection of these needles were harvested from the well using a Mounted Mesh Litho-Loop (Molecular Dimensions). The crystallisation solution contained sufficient 2-methyl-2,4-pentandiol such that further cryoprotection was not necessary before crystals were cryocooled using liquid nitrogen. Cryoprotection and cryocooling was carried out in an effort to reduce potential radiation damage induced during irradiation of the crystals with

X-rays (as reviewed in [270]). X-ray diffraction analysis of these crystals gave a low resolution diffraction pattern consistent with that of protein. However, the resolution and the quality of the reflections was insufficient for indexing and determination of the cell dimensions or space group.

Attempts were made to optimise this crystallisation condition, however, the crystals were not reproducible and could not be optimised for further structural studies. At this point, crystallisation trials for both the CBM and CBMCT constructs were pursued given the successful expression and purification of both these EndoS fragments. Despite the purity and yield of CBMCT, crystals were not observed in the crystallisation trials set up for this protein. Conversely, the CBM fragment of EndoS was crystallised successfully and further structural analysis of this domain forms the basis of Chapter 4.

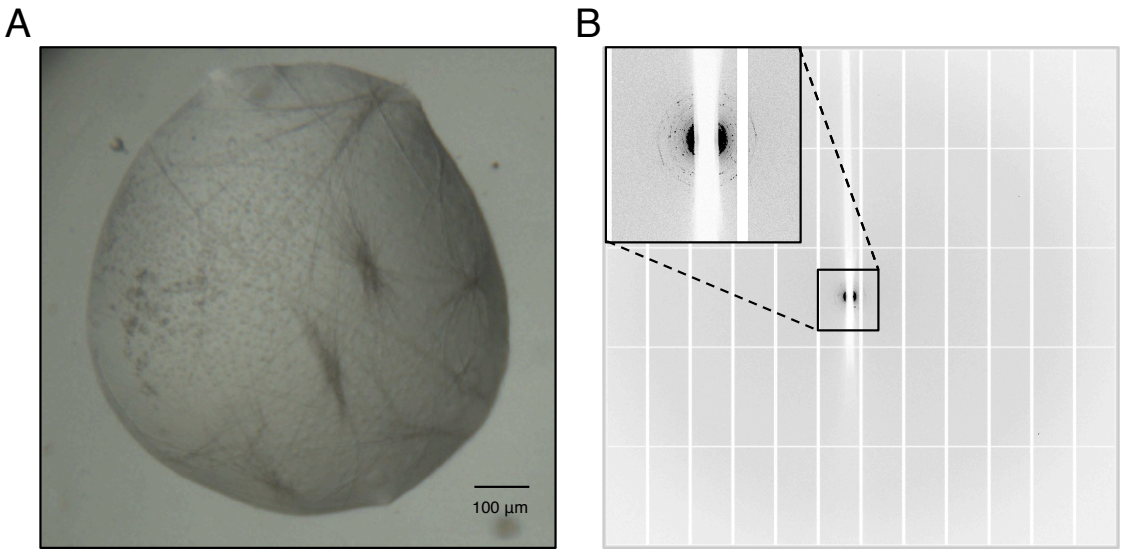


Figure 3.23: Crystallisation of EndoS.(A) The needle-like crystals of EndoS were grown from a condition containing 0.2 ammonium dihydrogen phosphate, 0.1 M Tris pH 8.5, 50% (v/v) 2-methyl-2,4-pentandiol. The crystals were pictured using a Minstrel^{TM} HT (Rigaku). (B) An X-ray diffraction image from EndoS crystals showing low resolution diffraction. The inset panel shows a magnification of the low resolution diffraction about the beamstop.

As for the EndoS^{E235L}Fc and EndoS^{E235L}2G12 complexes, extensive crystallisation trials were set up for both of these enzyme:substrate combinations. Unfortunately, neither combination yielded crystals suitable for X-ray diffraction studies.

3.7 SAXS

Further structural studies of EndoS, EndoS fragments and the EndoS:substrate complexes were pursued using SAXS. SAXS is a method which allows for the structural characterisation of proteins in solution and can provide information about the general shape, size and behaviour of a given protein in solution. SAXS was therefore used to look at EndoS and EndoS fragments in solution in another attempt to validate the bioinformatics studies. Given that little was known about the overall organisation of the predicted domains of EndoS, SAXS was also used to ascertain how these domains may be arranged with respect to one another. Finally, the EndoS^{E235L}Fc complex was analysed using SAXS in an effort to obtain information about how EndoS might interact with Fc in solution.

For the SAXS measurements, all samples were concentrated to 5 mg mL^{-1} except for CBM and CBMCT which were concentrated to 8 mg mL^{-1}. SAXS data were collected on beamline P12, DESY, with the exception of CBM which was collected at beamline BM29, ESRF. All measurements were recorded at 20 °C. Using the concentrated samples, a 2 fold dilution series of each protein was made. SAXS images were then collected for each dilution in the series from 0.05 s (DESY) or 1.0 s (ESRF) exposures as the sample was flowed through a capillary perpendicular to the X-ray beam. A total of 20 exposures were collected for each dilution of the proteins, with 10 exposures collected for CBM at the ESRF. In addition to the protein samples, measurements of buffer only were taken immediately prior and after the protein measurement. These buffer measurements were used for buffer subtraction during data processing.

Preliminary processing of scattering images was carried out automatically at the beamline using the programme PRIMUS [215]. For each dilution in the series, the integrated X-ray scattering data were plotted and averaged for all exposures to produce a curve of the log scattering intensity, $I(q)$ versus q (see Section 2.6.4 for a more detailed description). The same was carried out for buffer measurements. These averaged buffer curves were subtracted from the protein curves to account for the scattering contribution of the solvent.

Further data processing was carried out using the programme SCÅTTER. For all proteins tested, the scattering data from the serially diluted samples was used to extrapolate the curves to the intensity at zero concentration ($I(0)$). SCÅTTER was also used to calculate

the paired-distance distribution function, *P(r)* and the real space values for R_g and $I(0)$. Further processing of the data to produce 3-dimensional (3D) envelopes was carried out using the programmes DAMMIF [271] and DAMMIN [272] where the programme GNOM [273] was used to produce the input file for these methods. The programmes SITUS [274] and SCULPTOR [275] were used to convert the models produced by DAMMIF and DAMMIN into volumetric density maps. The results of the analyses of SAXS data collected from EndoS, EndoS fragments and for the potential complex of EndoS^{E235L}2G12 are presented below.

3.7.1 EndoS

EndoS has a calculated molecular weight of 110 kDa [206] and bioinformatic analyses presented earlier in this Chapter (Section 3.1), suggested the presence of multiple domains within the protein. The R_g of EndoS in solution, calculated using the Guinier approximation, was 44.93 Å. Values for additional parameters calculated using the SAXS data are given in Table 3.4. The Porod volume was consistent with a protein of molecular weight $\sim$ 135 kDa, assuming a protein density of 1.35 g cm^{-3} [276]. The *P(r)* for EndoS gave a bimodal distribution, characteristic of a multi-domain protein, with a D_{max} of 133 Å(Fig. 3.24).

Table 3.4: SAXS analysis of EndoS.

	EndoS
$I(0)^{\dagger}$	22700 (21200)
R_g (Å)†	44.93 (42.70)
M_w (kDa)*	109
D_{max} (Å)	133
Porod volume (Å^3)‡	181503
V_c (Å^2)‡	691.3

†Values were calculated using the Guinier approximation [217].Values in parentheses are real space values calculated using SCÅTTER [277].
*Approximated using the $I(0)$ and the $I(0)_{BSA}$.
‡The Porod volume and Volume-of-Correlation (V_c) were calculated using SCÅTTER as described in [214].

Ab initio models of EndoS in solution were generated using the programmes DAMMIF [271] and DAMMIN [272]. As its name suggests, *ab initio* shape determination from SAXS

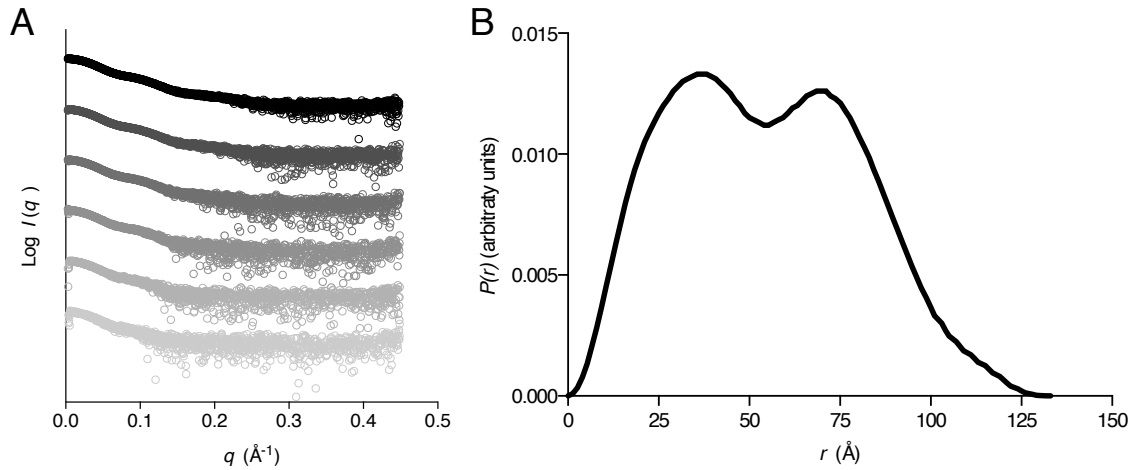


Figure 3.24: SAXS data and the calculated *P(r)* for EndoS. (A) The solvent corrected scattered intensities are shown for the 2-fold dilution series of EndoS. The data collected from the highest concentration of EndoS (5 mg mL^{-1}) are shown in black with the lower concentrations shown in increasingly lighter shades of grey. Each dilution is offset along the y axis by a factor of 10^3 so that all curves can be visualised on the same axes. (B) The *P(r)* of EndoS represents a frequency distribution of interatomic vectors of length r. The curve was calculated using SCÅTTER and was normalised to the area under the curve. $D_{\max} = 133$.

data allows for the prediction of protein shape and conformation in solution without the need for *a priori* structural information. Generally, a series of models are created using the programme DAMMIF, which are then aligned and averaged before being refined to a single model using DAMMIN. DAMMIF starts by declaring a spherical search volume, in real space, with a radius of half the calculated $D_{\max}$ of the particle. This search volume is then packed with many smaller spheres or dummy atoms. Each of these dummy atoms can be declared as either particle or solvent. The initial dummy atom model (DAM) of the particle in solvent is iteratively refined using a simulated annealing procedure. The model is continually refined to minimise differences between the theoretical scattering curve calculated from the model and the observed scattering data. Here, DAMMIF was used to calculate 23 models of EndoS in solution which are shown in Figure 3.25.

The DAMAVER suite of programmes [278] was used to align, compare and average these models in an effort to produce a single model of EndoS in solution. DAMAVER aligns the DAMMIF models by trying to minimise the difference between them, this difference is measured using a metric called the normalised spatial discrepancy (NSD) [278]. There were 2 models which were excluded from further model refinement and averaging based on their NSD values. These are shown in dark grey in Figure 3.25.

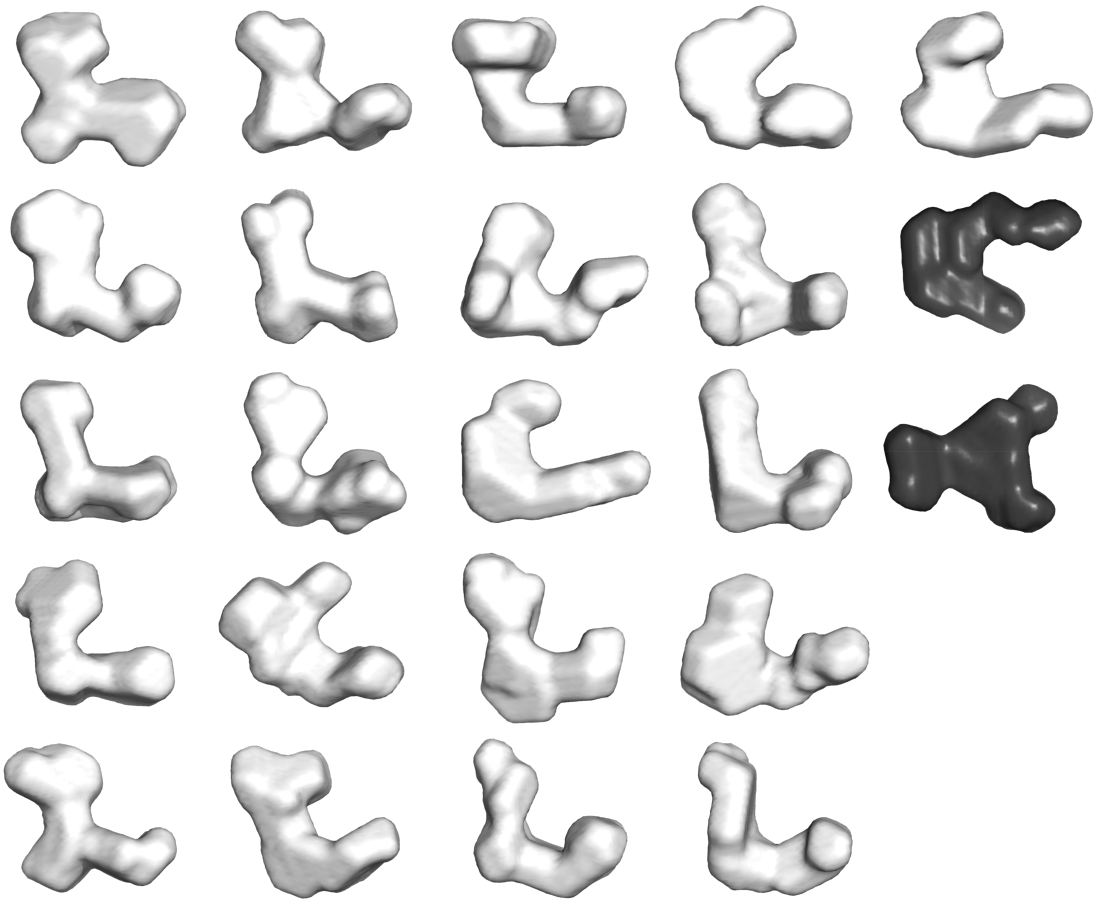


Figure 3.25: DAMMIF models of EndoS in solution. The 23 DAMMIF models were aligned using DAMAVER [278] and are shown in their aligned orientations (mean NSD = 0.884 ± 0.040). The DAMMIF models were converted from DAMs into volumetric density maps using SITUS [274] and SCULPTOR [275]. The 21 models used to produce an average model are shown in white and the 2 models excluded by DAMAVER from averaging are shown in dark grey.

The DAMMIF models all showed an overall L-shaped conformation. There were two arms at near right-angles to one another, which extended from a central point. The main differences between the models arose from the globular features in each of the two arms whereby a fixed conformation of the bulbous features at the extremities of the arms was not observed across the models.

The DAMAVER averaged model was used as a starting point for a final round of model refinement using the programme DAMMIN [272] which, like DAMMIF, uses a simulated annealing procedure to refine towards a model resembling the observed data. The resultant model from DAMMIN is presented in Figure 3.26.

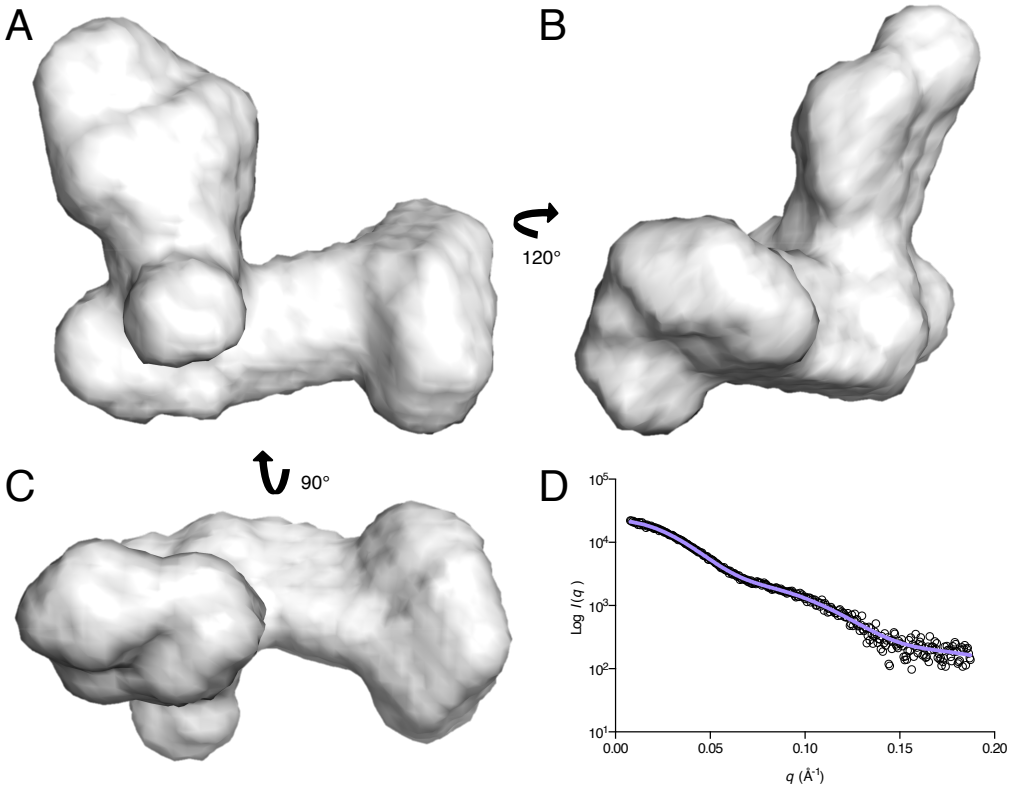


Figure 3.26: DAMMIN refined model of EndoS in solution. The final DAMMIN refined model was converted from a DAM into a volumetric density map using SITUS [274] and SCULPTOR [275]. The model is shown in three orientations, to show (A) the L-shaped conformation of the calculated envelope and views of the model when rotated (B) approximately 120° and (C) 90° for 'top-down' view. (D) A theoretical scattering curve was calculated from the model (mauve) and is plotted against the scattering data (black circles), $\chi^2 = 0.7064$.

The final DAMMIN model, much like the DAMMIF models used to produce it, showed two arms with different features protruding from an underlying L-shape. The arms were similar in length extending out at approximately right-angles to one another. Figure 3.26D shows a fit of the theoretical scattering curve calculated from the final DAMMIN model to the observed scattering data ($\chi^2 = 0.7064$).

3.7.2 EndoS fragments

Having determined an overall envelope for EndoS in solution, SAXS analysis of fragments of EndoS was also carried out. This was done with an aim to identify which subdomains of EndoS contribute to the various features of the overall envelope. By doing this, it was hoped that an overall model of EndoS could potentially be predicted.

The SAXS data collected for the EndoS fragments ChitLRR, CBM and CBMCT were analysed in an identical fashion to those collected for full-length EndoS. Key parameters describing these fragments in solution are listed in Table 3.5. The paired-distance distribution functions for each fragment are given in Figure 3.27 along with the final DAMMIN produced models of these proteins.

Table 3.5: SAXS analysis of EndoS fragments.

	ChitLRR	CBM	CBMCT
$I(0)^{\dagger}$	12500 (12100)	17.9 (16.9)	4910 (4610)
R_g (Å)†	34.49 (35.07)	18.90 (18.59)	22.72 (22.91)
M_w (kDa)*	59.8	18.7	23.5
D_{max} (Å)	115	61	77
Porod volume (Å^3)‡	13200	32800	46300
V_c (Å^2)‡	501	186.6	234.6

†Values were calculated using the Guinier approximation [217].Values in parentheses are real space values calculated by GNOM [273].
*Approximated using the $I(0)$ and the $I(0)_{BSA}$.
‡The Porod volume and Volume-of-Correlation (V_c) were calculated using SCÅTTER as described in [214].

The ChitLRR construct gave an elongated boot-shaped structure with an R_g of 34.49 Å. CBM formed a small featureless globular protein with an R_g of 18.90 Å. Together with the CD data, this globular shape is consistent with the prediction that this region of EndoS does indeed form a CBM-like fold. The calculated *P(r)* function of CBMCT indicates a more elongated protein compared to CBM. Indeed, the DAMMIN envelope supports this

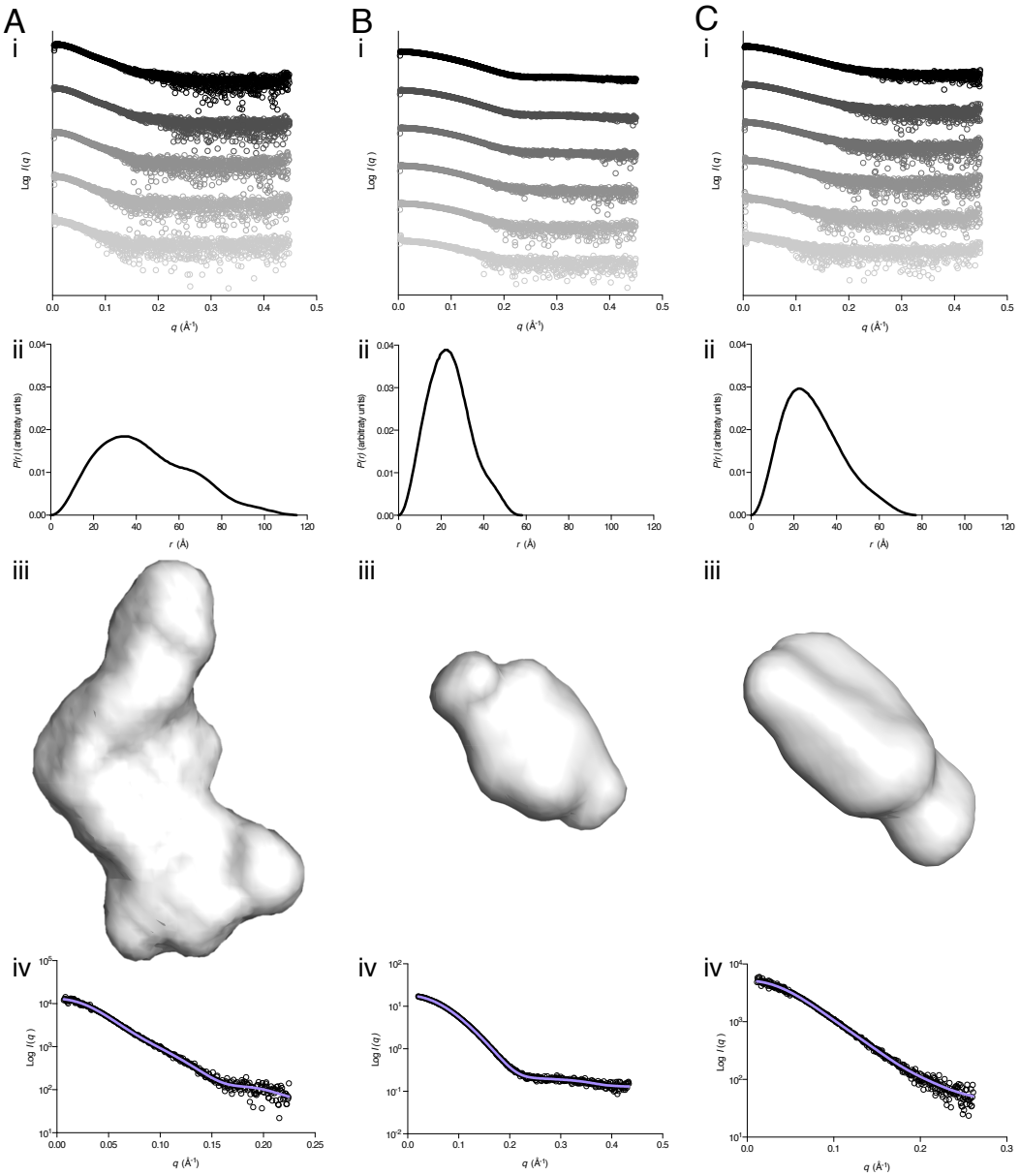


Figure 3.27: SAXS analysis of EndoS fragments. Data are depicted for (A) ChitLRR, (B) CBM and (C) CBMCT. (i) The solvent corrected scattered intensities are shown for the 2-fold dilution series of the EndoS fragments. The data collected from the highest concentrations of ChitLRR (5 mg mL^{-1}), CBM and CBMCT (8 mg mL^{-1}) are shown in black with the lower concentrations shown in increasingly lighter shades of grey. Each dilution is offset along the y axis by a factor of 10^3 so that all curves can be visualised on the same axes. (ii) The *P(r)* for each fragment represents a frequency distribution of interatomic vectors of length r. The curves were calculated using SCÅTTER and were normalised to the area under the curve and are presented on a common scale. (iii) The final DAMMIN refined envelopes for each protein fragment reveal the overall shape of each construct. The 23 DAMMIF envelopes used to produce the DAMMIN model gave mean NSD values of 0.797 ± 0.055, 0.500 ± 0.024 and 0.653 ± 0.049 for ChitLRR, CBM and CBMCT, respectively. (iv) A theoretical scattering curve was calculated for each of the DAMMIN models (mauve) and is plotted against the scattering data (black circles). The fit of the theoretical to the experimental scattering curves gave χ^2 values of 0.5856, 0.6512, and 0.6411 for ChitLRR, CBM and CBMCT, respectively.

observation where CBMCT forms a long, more cylindrical domain in comparison to CBM.

It was not clear from these structures how each of these fragments were arranged into the full-length EndoS envelope, however, the publication of a crystal structure of EndoS, which is described in the following section, led to the ability to model the various fragments of EndoS into these envelopes to assess the quality of these SAXS-based predictions of protein structure.

3.7.3 Models of EndoS in solution using the X-ray crystal structure of EndoS

As mentioned above, a crystal structure of EndoS (residues 101-989) was published during the course of this project [149]. An overview of the structure is given in Figure 3.28 where individual domains of the proteins are labelled accordingly. Using SeMet derived EndoS crystals, Trastoy *et al.* determined the structure of EndoS to 3.2 Å resolution using multiwavelength anomalous dispersion. This partially refined structure was subsequently used as a molecular replacement search model to determine the structure of EndoS using a dataset from native EndoS crystals which diffracted to 1.9 Å resolution. All constructs used in these crystallisation experiments encompassed residues 98–995 and the crystals used for X-ray diffraction studies were grown using a liquid–liquid diffusion method [279].

As predicted by bioinformatic studies, the crystal structure of EndoS confirmed the presence of an N-terminal GH18 chitinase domain with a LRR domain directly downstream. The LRR domain was followed by a hybrid immunoglobulin domain (hIg). This hIg domain was not explicitly annotated in the FUGUE-derived models described in this chapter, but the homologous internalin B protein used to model this section of EndoS in Section 3.1.2 does contains an Ig-like domain in its C-terminus [245] which can be seen in Figure 3.2. The crystal structure of EndoS also confirmed the presence of a CBM between residues 764–924. This matched the bioinformatic prediction obtained using FUGUE as detailed in Section 3.1.2 (Table 3.1) where the domain boundaries defining the position of the putative CBM ranged from 760–925 for the top five FUGUE hits. The α-helical content of the final 70 residues of EndoS (CT) as predicted by CD studies (Section 3.5.2) was consistent with the 3-helix bundle described in the crystal structure of this region.

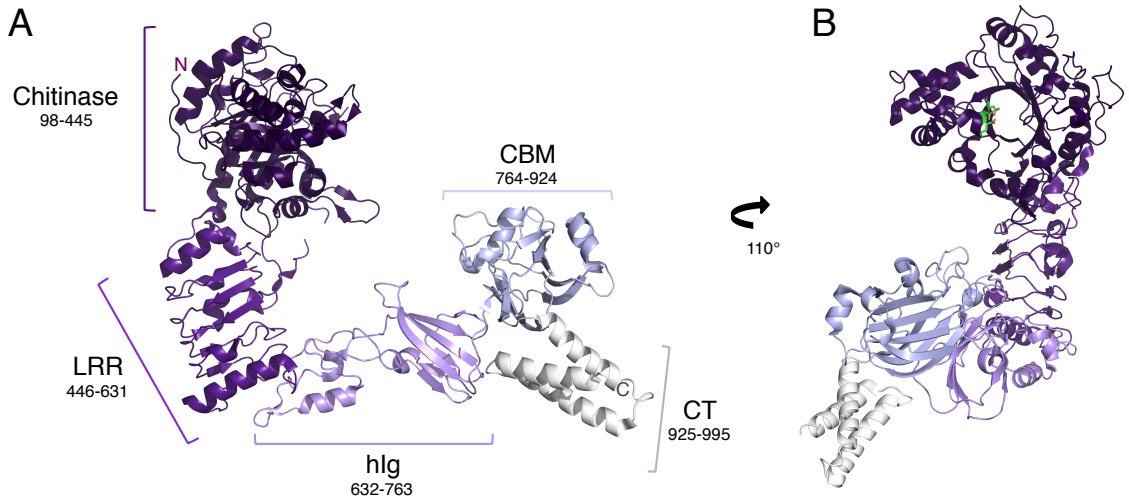


Figure 3.28: The crystal structure of EndoS. (A) The crystal structure of EndoS (residues 101–989) at 1.9 Å resolution was solved by molecular replacement using a recombinantly expressed fragment of EndoS (residues 98–995) [149]. The five domains of EndoS and the residues which they encompass are labelled accordingly: Chitinase domain, 1–445; LRR domain, 446–631; hybrid Ig domain, 632–763; CBM, 764–924; CT, 925–995. The domains are coloured from dark purple to white from the N- to the C-terminus. (B) Rotation of the structure through 110 ° gives a view of the donut-shaped GH18 chitinase domain. The acidic residues of the DXDXE active site motif (Asp231, Asp233 and Glu235) are shown as green sticks with the oxygen atoms coloured red.

Overall, the biophysical characterisation of EndoS described in this chapter was consistent with the X-ray crystal structure published by Trastoy *et al.* [149]. To further illustrate this, the X-ray crystal structure of EndoS was fitted to the SAXS derived envelopes described in Section 3.7.1 and 3.7.2 using the programme SUPCOMB [280]. (Fig. 3.29 and 3.30).

Although the SAXS analysis carried out here was able to produce an overall model of EndoS using DAMMIN (Fig 3.26) the variability in the DAMMIF models suggested that EndoS may show some flexibility in solution. The program EOM (Ensemble optimisation method) [281] was therefore used in an attempt to model this potential flexibility. EOM aims to describe experimental SAXS data with an ensemble of atomic models as opposed to programmes like SASREF or SUPCOMB which use a rigid body approach to produce a single model of the protein in solution.

For this EOM analysis, the domains from the EndoS X-ray crystal structure were used as models (chitinase, LRR, hIg, CBM, CT). In addition, the N-terminal 100 residues of EndoS were included and were modelled as random coil. These residues were included in the model considering these were present in the construct used for SAXS experiments described in this Chapter. Each of the domains was permitted to move independently of

the others and the results from this analysis are shown in Figure 3.31. Modelling with EOM suggested that EndoS largely forms a compact L-shaped conformation similar to that seen in the crystal structure. Some elongated models of EndoS, where the domains are arranged linearly as opposed to in an L-shape, formed part of the ensemble used to describe the behaviour of EndoS in solution as measured by SAXS. However, these models made up $< 10\%$ of the total pool of conformations.

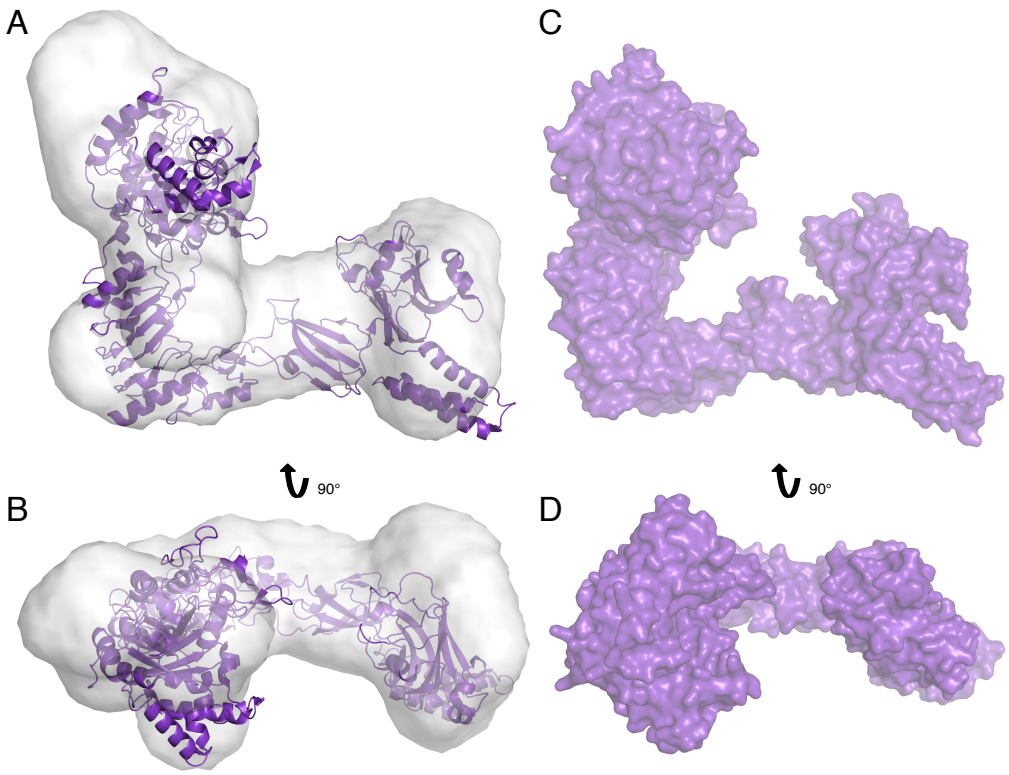


Figure 3.29: SUPCOMB fit of the crystal structure of EndoS to the SAXS envelope.

3.7.4 EndoS-Fc complex

SAXS analysis of the EndoS^{E235L}Fc complex was carried out in an effort to generate a model of this complex in solution. A model of these two proteins in complex with one another would be of use in helping to understand the aspects of EndoS which contribute to the apparent specificity of this interaction. In addition to the complex, SAXS data was collected for the Fc homodimer alone. The analysis of these proteins using SAXS was carried out following an identical protocol to that used for the analysis of EndoS in isolation. Data statistics for the EndoS^{E235L}Fc complex, Fc and EndoS are described in Table 3.6.

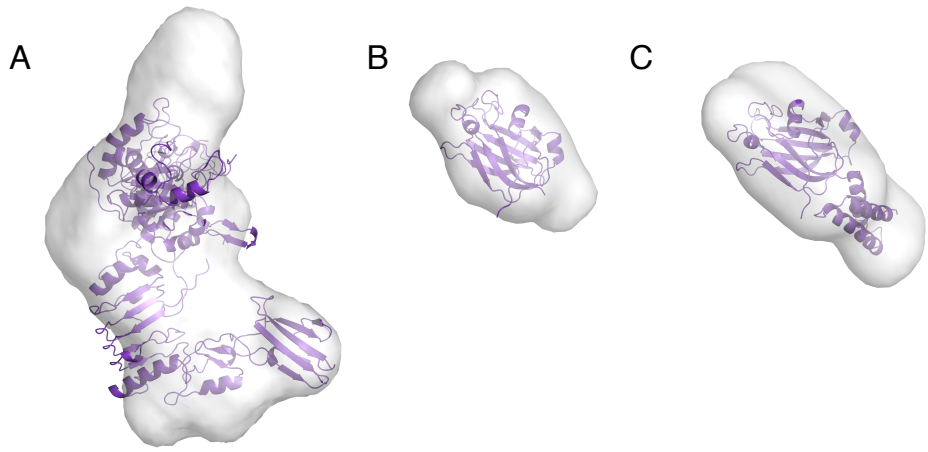
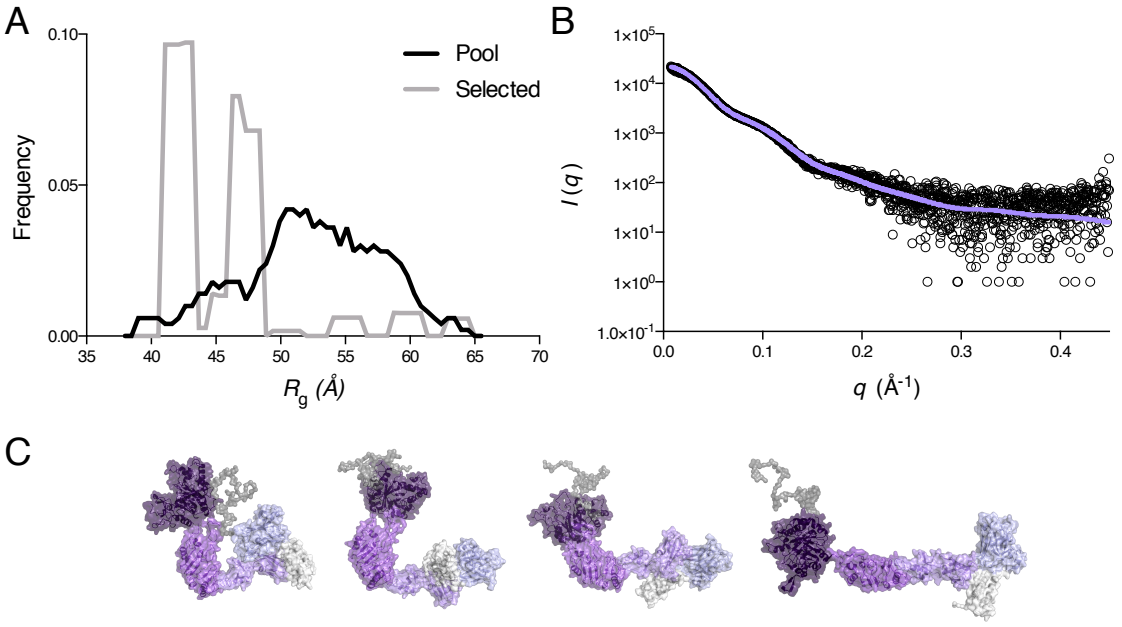


Figure 3.30: SUPCOMB fit of EndoS crystal structure subdomains to SAXS envelopes from EndoS fragments.



Conformation	1	2	3	4	Average
Rg (Å)	42.30	47.36	50.50	64.30	45.94
D_{max}	145.49	151.50	170.12	213.82	152.99
Freq.	9	7	1	1	18

Figure 3.31: Analysis of EndoS SAXS data using EOM. (A) A graph detailing the R_g distribution of the total pool of EndoS models is shown in black. The grey line describes the distribution of the selected subgroup of models which were used to create the final ensemble. (B) A theoretical scattering curve was calculated using the ensemble of models (mauve) and is plotted against the scattering data (black circles). (C) Representative models of EndoS which were selected from the pool to make up the ensemble are shown here. The domains are coloured accordingly: chitinase, dark purple; LRR, purple; hIg, mauve; CBM, light blue; CT, white; Unstructured N-terminal residues, grey. The tabulated values describe the properties of each model in solution and the relative contributions of these models to the ensemble.

A comparison of the data for EndoS^{E235L}Fc to that of EndoS shows there was a modest increase in the Porod volume for the complex, however, both the $I(0)$ and the R_g for EndoS^{E235L}Fc were similar to those recorded for EndoS (Table 3.6).

The DAMMIF models of the EndoS^{E235L}Fc complex are shown in Figure 3.32. These models did not show a consistent conformation of the putative EndoS^{E235L}Fc complex where there was significant heterogeneity across the 23 models.

Table 3.6: SAXS analysis of the EndoS^{E235L}Fc complex.

	EndoS	Fc	EndoS^{E235L}Fc
$I(0)^{\dagger}$	22700 (21200)	55.4 (53.0)	22700 (22800)
R_g (Å)†	44.93 (42.7)	29.17 (28.74)	44.88 (44.91)
M_w (kDa)*	109	58.1	109
D_{max} (Å)	133	101	150
Porod volume (Å^3)‡	181503	86100	202167
V_c (Å^2)‡	691.3	425.0	747.9

†Values were calculated using the Guinier approximation [217].Values in parentheses are real space values calculated by GNOM [273].
*Approximated using the $I(0)$ and the $I(0)_{BSA}$.
‡The Porod volume and the Volume-of-Correlation (V_c) were calculated using SCÅTTER as described in [214].

The *P(r)* functions for both the components and the complex are shown in Figure 3.33 along with the DAMMIN refined envelopes of these proteins. The *P(r)* function for the EndoS^{E235L}Fc complex shows a higher frequency of larger interatomic vectors when compared to the *P(r)* function of EndoS alone. The DAMMIN refined envelope of the EndoS^{E235L}Fc complex provides a potential view of the overall complex conformation in solution. Unfortunately, placement of the EndoS and Fc components could not be unambiguously assigned into this SAXS envelope. This meant that the data could not be used to appreciate the potential interface between EndoS and Fc.

However, to make further use of the SAXS data collected for the complex, the protein docking programme ClusPro [282] was used to make a series of possible models of the EndoS^{E235L}Fc complex. Here, the SAXS data was used to selectively filter out models consistent with experimental data. The crystal structure of EndoS (PDB ID: 4NUY, [149]) and the crystal structure of a glycosylated human IgG1 Fc protein (PDB ID: 3AVE, [46]) were used to make these models. For these complexes, the EndoS structure was declared as the receptor and the Fc structure was treated as the ligand. Briefly, ClusPro

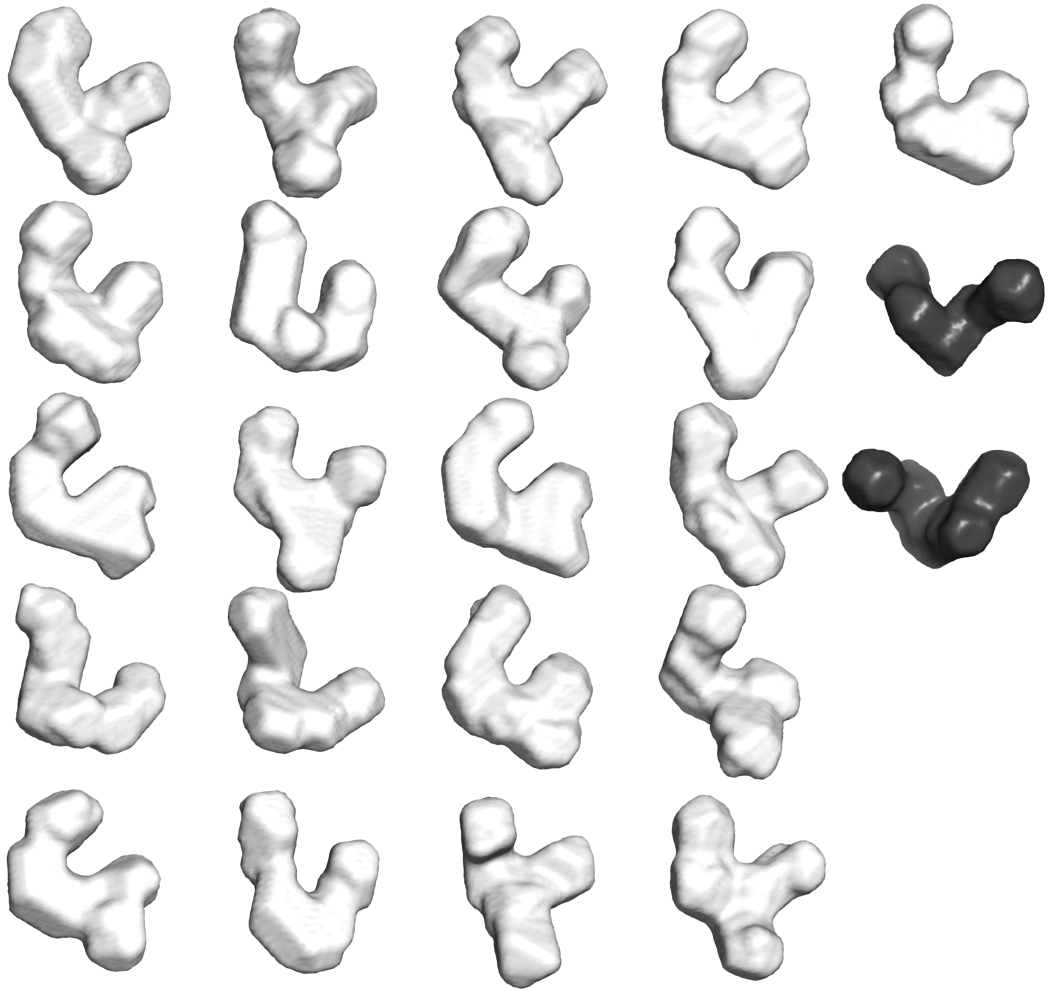


Figure 3.32: DAMMIF models of the EndoS^{E235L}Fc complex. The 23 DAMMIF models were aligned using DAMAVER [278] and are shown in their aligned orientations (mean NSD = 0.884 ± 0.040). The DAMMIF models were converted from DAMs into volumetric density maps using SITUS [274] and SCULPTOR [275]. The 21 models used to produce an average model are shown in white and the 2 models excluded by DAMAVER from averaging are shown in dark grey.

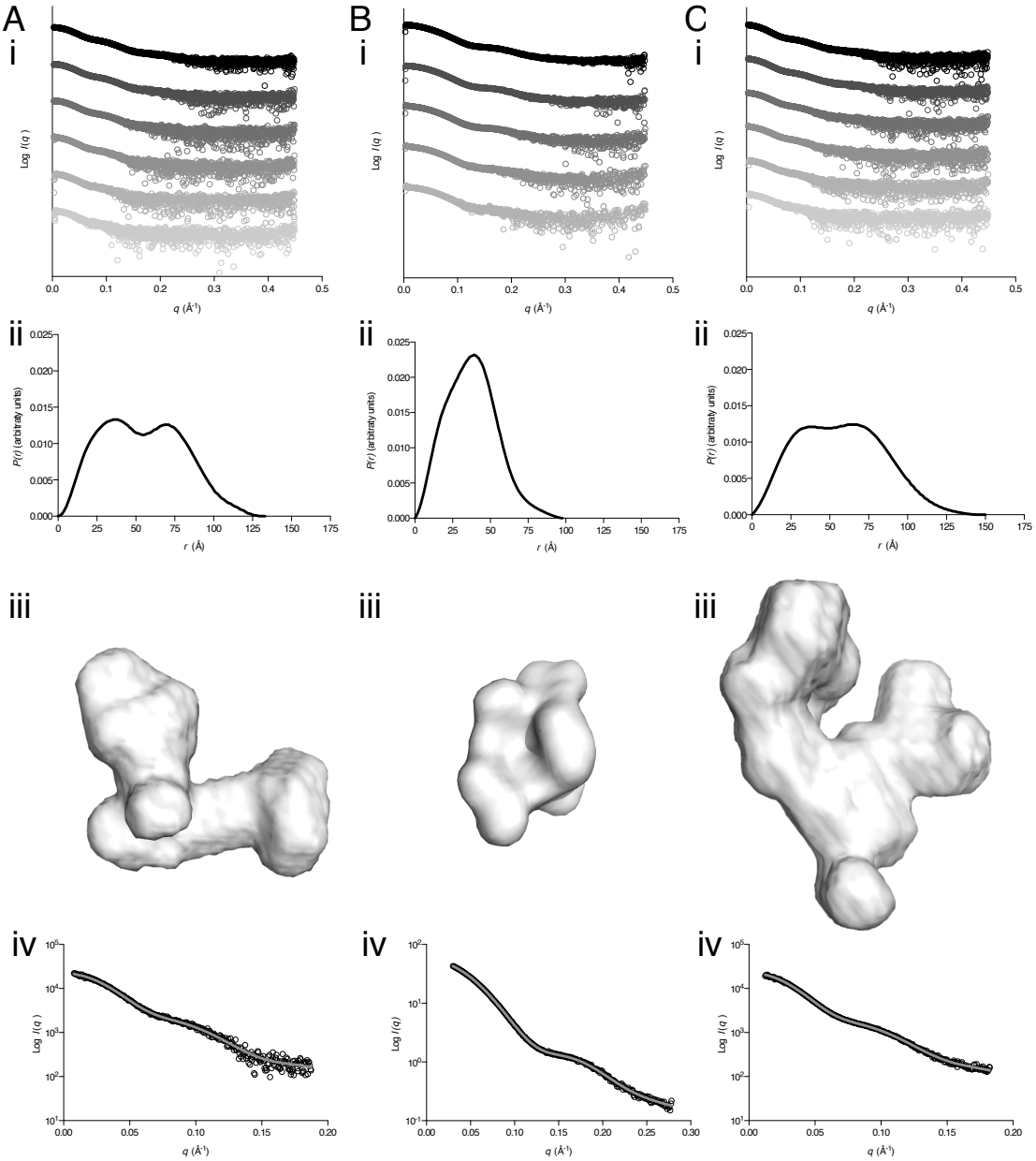


Figure 3.33: DAMMIN models of the EndoS^{E235L}Fc complex. Data are depicted for (A) EndoS, (B) Fc and (C) EndoS^{E235L}Fc. (i) The solvent corrected scattered intensities are shown for the 2-fold dilution series of each sample. The data collected from the highest concentrations of EndoS, Fc and EndoS^{E235L}Fc (5 mg mL^{-1}) are shown in black with the lower concentrations shown in increasingly lighter shades of grey. Each dilution is offset along the y axis by a factor of 10^3 so that all curves can be visualised on the same axes. (ii) The *P(r)* functions for the components and the EndoS-Fc complex represent a probability distribution of interatomic vectors of length r. The curves were calculated using SCÅTTER and were normalised to the area under the curve and are presented on a common scale. (iii) The final DAMMIN refined envelopes reveal the overall shape of the components with a putative model shown for the complex. (iv) A theoretical scattering curve was calculated for each of the models (grey) and is plotted against the scattering data (black circles).

calls the programme PIPER to prepare 70,000 unique rotations of the ligand structure (Fc). Each rotated model is then translated with respect to the rigid receptor structure (EndoS) in a grid across x, y and z dimensions. The translations are scored and the best translation is kept for each rotation of the ligand. From the total pool of the 70,000 best translations (*i.e.* one for each rotation), the 1000 with the best translation scores are kept. At this stage, SAXS intensity curves can be used to filter the results. This was done for the EndoS^{E235L}Fc models presented here which were filtered using the SAXS data for the EndoS^{E235L}Fc complex. After filtering, ClusPro then clusters the remaining ligand positions by searching for neighbours within a 9 Å C^{α} RMSD radius. The ligand with the largest number of neighbours is declared as the cluster centre with the neighbours within the 9 Å radius then becoming members of that cluster. As a cluster is declared, all its members are removed from the pool of ligand positions and the clustering process then moves on to identify the next cluster centre and cluster members. This process is repeated until all potential ligand positions are assigned to a cluster. Clusters are then ranked based on the number of members, where the top clusters are those with the most members. The model structure for each cluster comprises the receptor with the ligand shown in the position of the cluster centre.

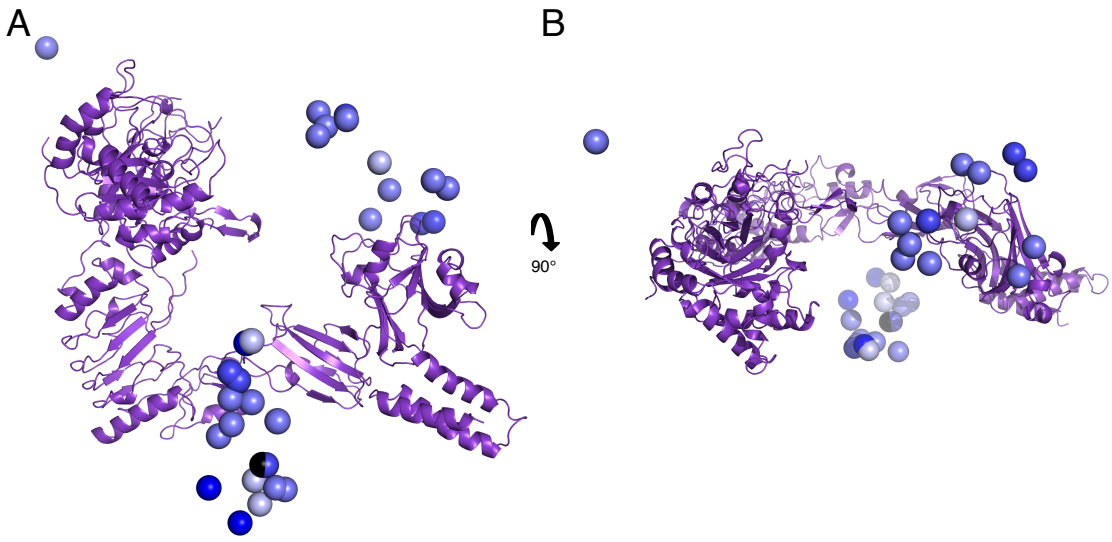


Figure 3.34: Prediction of EndoSFc complex structures using ClusPro. The crystal structure of EndoS used to create these models is shown in purple in cartoon representation. The position of the centre of mass of the Fc dimer from each ClusPro cluster is represented as a sphere. The spheres are coloured from dark blue to light blue based on the size of the cluster they represent. The largest cluster is represented by a black sphere.

Figure 3.34 gives a representation of the results from ClusPro. Here, the structure of EndoS is shown in cartoon representation with the position of the centre of mass of the Fc dimer from each cluster model represented as a sphere. The darker the blue of these spheres, the larger the respective cluster of that model. The clusters localise to two regions, one adjacent to the LRR and hIg domain and a second next to the CBM. The models from all top five clusters show the Fc localised to the LRR and hIg domains. The top cluster was fitted to the DAMMIN refined envelope using SUPCOMB (Fig. 3.35). However, given the poor fit of this structure to the envelope and the distance between the position of the *N*-linked glycan relative to the active site of EndoS, this is not likely to be a correct representation of the structure of the EndoS^{E235L}Fc complex. Therefore, further experiments are required to define how both EndoS and Fc are arranged within the calculated envelope.

3.8 Discussion

Given the strong links between protein structure and function, work in this chapter sought to better understand the structure of EndoS in an effort to appreciate which parts of the protein contribute to the activity and specificity of EndoS for IgG. Prior to these studies, some characterisation of the N-terminal half of EndoS had been reported, but little was known about the domain organisation of the C-terminus or the overall organisation of these predicted domains.

Work in this chapter identified and characterised elements in the C-terminal regions of EndoS and helped generate an overall picture of EndoS in solution. As well as identifying the previously annotated chitinase and LRR domains, bioinformatic analyses described herein revealed the presence of a putative CBM in the C-terminus of EndoS. Various subdomains and fragments of EndoS were recombinantly expressed in *E. coli* and purified to sufficient homogeneity for subsequent biophysical analysis. Characterisation of these EndoS fragments as well as full-length EndoS using CD and SAXS provided an overall view of the structure of EndoS in solution. In addition to these fragments, catalytically inactive EndoS mutants were made so that a stable complex of EndoS and Fc could be generated. SAXS analysis of EndoS^{E235L}Fc was carried out and was used to produce some

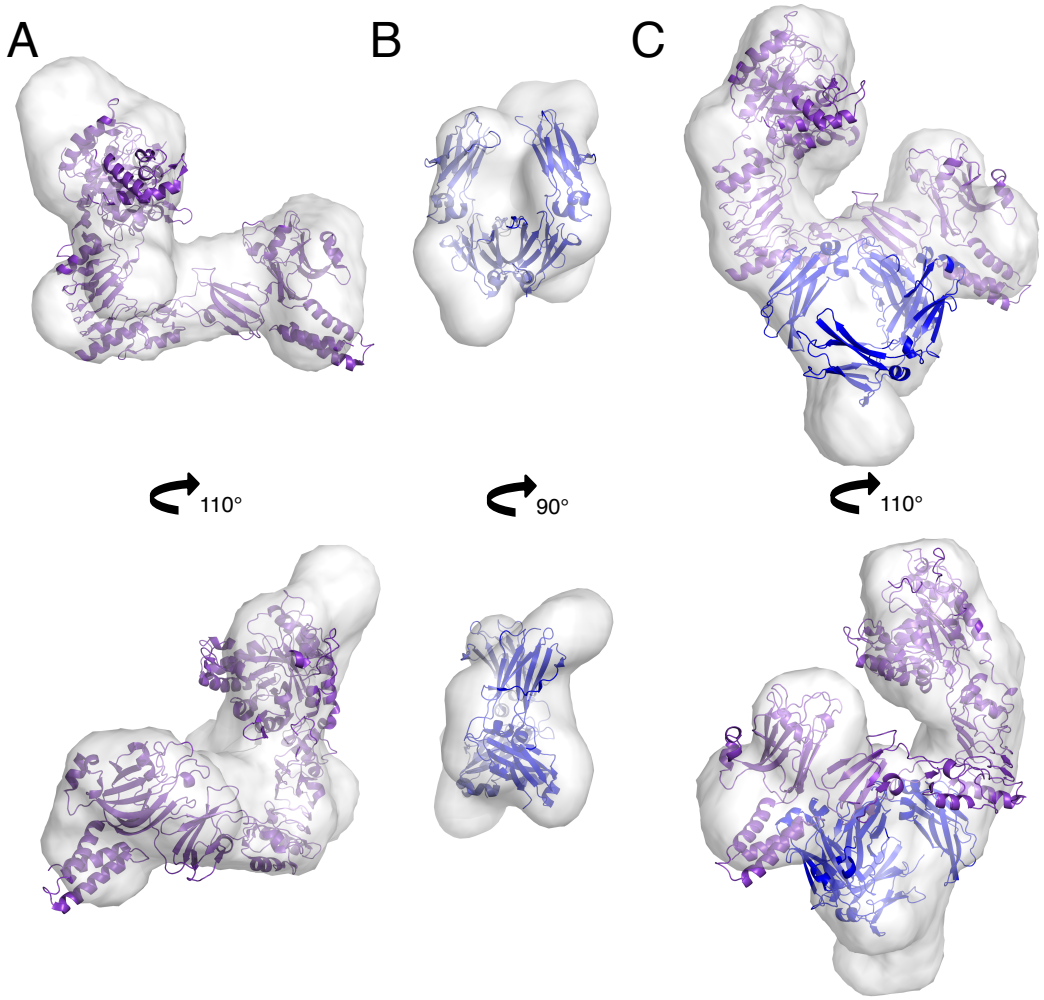


Figure 3.35: Preliminary SAXS model of an EndoSFc complex. Models are shown for (A) EndoS, (B) Fc and (C) EndoS^{E235L}Fc. SUPCOMB was used to fit various protein structures to the DAMMIN refined envelopes. For the EndoS and Fc complex, X-ray crystal structures of these proteins were fitted to the envelopes (EndoS, PDB ID: 4NUY [149]; Fc PDB ID: 3AVE [46]). For the EndoS^{E235L}Fc complex, the top ClusPro prediction was modelled into the envelope.

preliminary solution state models of EndoS in complex with Fc.

As mentioned briefly above, a C-terminal fragment of EndoS was determined to have homology to CBM family 32. Having identified this putative domain, a soluble form of the proposed CBM was successfully cloned, expressed and purified as an isolated domain. Analysis by CD revealed the domain as containing primarily β-sheet structure with a large proportion of loops and turns, supportive of the bioinformatic assignment of this region as a CBM.

The CBM32 family have been shown to bind monosaccharides and disaccharides found at the termini of mammalian *N*-linked glycans (reviewed by [283]). The putative CBM from EndoS showed homology to the galactose binding domain of a sialidase from *Micromonospora viridifaciens* [284] and the galactose binding domain of galactose oxidase from *Dactylium dendroides* [251]. Both of these CBM domains bind galactose, as their name suggests, with mM affinity. This low affinity binding of CBMs to carbohydrate moieties is typical for this family of proteins [260]. The function of this domain in these homologues is likely to aid avidity and selectivity of the enzyme. In the case of EndoS, a CBM could indeed aid the function of the enzyme and may be responsible for conferring specificity to IgG glycans. In addition to these two galactose-binding domains, another homologue identified was the CBM32 domain of the endo-β-1,4-*N*-acetylglucosaminidase, EndoD. EndoD is produced by the related bacteria *Streptococcus pneumonia* which can cleave some oligomannose-type glycans, such as $Man_{5-4}GlcNAc_2$, and paucimannose-type structures such as $Man_3GlcNAc_2$ [285]. A ligand for this CBM could not be identified, but structural features of the putative binding site suggested its function could involve both substrate capture or adhesion of *S. pneumonia* to host tissues via carbohydrates. Considering the role of these homologous proteins, it is likely that this CBM is able to aid in substrate recognition and enzyme avidity thus contributing to EndoS function.

Considering the potential role of a CBM in aiding the specificity and a avidity of EndoS, further structural and functional investigation of the CBM from EndoS was carried out and is described in Chapter 4. The contribution of this domain to the catalytic activity of EndoS is explored further in Chapter 5.

Analysis of EndoS using SAXS indicated that the enzyme had an overall L-shape in solution. Further SAXS analyses of subdomains of EndoS were carried out with an aim

to understand how the predicted domains of EndoS were organised within the structure. Although these subdomains could not be unambiguously assigned to regions of the full-length EndoS envelope, these studies of EndoS fragments still provided useful information about these individual domains. Specifically, SAXS data for the predicted CBM of EndoS was consistent with that of a small globular domain, further compounding the evidence suggesting this region encoded a CBM. In addition, SAXS analysis of the CBMCT protein suggested that the CT domain was appended to one end of the putative CBM region as the calculated envelope for the CBMCT construct formed an elongated shape as opposed to a globular unit.

As mentioned in the results section of this chapter, an X-ray crystal structure of EndoS was published during the course of this project. Encouragingly, this structure provided support for all of the bioinformatic and biophysical data described in this chapter. The structure indeed contained a CBM in its C-terminal region where the bioinformatic prediction that this region spanned residues 760–925 closely matched that of the actual crystal structure (residues 764–924). The crystal structure also confirmed that the last 70–100 residues of EndoS formed a largely α-helical component, consistent with the CD analysis of the CBMCT construct. Overall, it appears that the programme FUGUE was successful in identifying both homologous regions and domain boundaries for the multiple subdomains of EndoS. As the amount of structural information within databases like the PDB increases, in combination with increases in computing power and efficiency, bioinformatic methods become increasingly powerful and useful as an initial step towards understanding protein structure. This work provides an example of how a combination of bioinformatic and biophysical analysis can significantly contribute towards the understanding of protein structure in the absence of atomic resolution information.

Once this structure of EndoS was made available, the SAXS data for EndoS was reanalysed in light of this new structural information. The L-shape of EndoS predicted using SAXS was consistent with the conformation seen in the crystal structure. Further analysis using EOM suggested that EndoS largely formed this L-shape conformation in solution but that it may have some flexibility which allows formation of more extended conformations. The conformations of the domains within the crystal structure of EndoS are likely to be stabilised by the presence of various salt-bridges. However, considering the large proportion of

lysines (10%) and acidic residues (15%) within the EndoS sequence it is perhaps feasible that salt bridges could break and reform to allow these domains to move with respect to one another. Overall, it appears that EndoS can form a stable conformation in solution, but perhaps explores other conformations also.

Initial attempts to purify EndoS in complex with Fc were hampered by the fact that the published inactivating mutation did not produce an entirely inactive form of EndoS and due to EndoS showing activity against what was thought to be a resistant Fc glycoform. Nevertheless, several mutants were made in order to produce a completely catalytically inactive protein. The catalytically inactive EndoS protein with the mutation E235L was used for further attempts at forming and EndoS-Fc complex. EndoS^{E235L} was combined with purified Fc before being analysed by SEC to look for the formation of a complex. SEC analysis revealed that these two proteins coeluted in a single peak that was consistent with 1:1 complex of EndoS to Fc. Attempts to crystallise this complex were unfortunately unsuccessful as were trials using a putative complex of EndoS with the full-length IgG antibody 2G12. A structure of this complex would be of great value for understanding which aspects of EndoS are crucial for its specific activity against IgG. In particular, information concerning the interaction interface of EndoS and IgG would be of use for receptor refocusing strategies which require the therapeutic antibody to be resistant to EndoS deglycosylation. By understanding the structural basis of this specificity, rational mutations or truncations of IgG could be designed to confer EndoS resistance. Despite the potential worth of a structure of the complex of these two proteins, further work towards this goal was abandoned as we became aware of a competing group working on the structure of this complex (Prof. Eric Sundberg, personal communication).

Alongside crystallisation trials, SAXS analysis of the putative EndoS-Fc complex was carried out. Despite the apparent coelution of EndoS^{E235L} and Fc during SEC analysis, it is possible that complexes of EndoS-Fc are not stable in solution. Indeed, the similarity of the $I(0)$ of EndoS and the $I(0)$ of the putative EndoS^{E235L}Fc complex may be caused by heterogeneity in the sample. This heterogeneity is likely a result of dissociation of EndoS-Fc complexes such that a mixture of free EndoS^{E235L} and free Fc exists with the population of EndoS^{E235L}Fc. In future, cross-linking experiments could be pursued to stabilise the complex formed by EndoS and Fc so that a more homogeneous sample could be obtained

for use in SAXS experiments. Further analysis of the SAXS data collected for the EndoS-Fc complex also suggested that a rigid complex with a single conformation was not formed by these two proteins. This was exemplified by the variation observed in the envelopes generated using the programme DAMMIF (Fig. 3.32). Considering how different these envelopes were from one another, the averaged model produced by DAMMIN is unlikely to be an accurate representation of the complex of EndoS and Fc as it does not take into account this flexibility nor potential heterogeneity in the sample. Given the apparent flexibility and potential sample heterogeneity of EndoS^{E235L}Fc as observed using SAXS, this may partially explain the inability to crystallise this complex. If these studies were to be further pursued, a critical first step would be to perform SEC-MALLS (multi-angle laser light scattering) to verify the size and homogeneity of the complex in the eluted peak. Secondly, quasi-electron light scattering (QELS) could be used to infer whether the complex generated here forms a single conformation or is a series of conformations.

Considering that atomic resolution structures for both components of this complex are now available, molecular dynamics simulations could be pursued. Indeed, an initial docking study was presented here using the programme ClusPro. This programme was able to filter protein-protein docking results using the SAXS in an attempt to define a model of EndoS in complex with Fc which was consistent with experimental data. Although a model was produced, the active site of EndoS was positioned too far away from the Asn297 residue Fc (where the glycan is positioned) for this to be a plausible model. Crucially, ClusPro is unable to model glycans into the supplied structures nor is it able to account for flexibility or conformational changes of the components upon complex formation. These two factors are both likely to influence the formation of the EndoS^{E235L}Fc complex and would need to be accounted for in further modelling experiments. In addition to computational methods, it may be of use to pursue hydrogen-deuterium exchange (HDX) MS methods. HDX-MS can be used to investigate protein–protein interactions, specifically, it can be used to identify crucial interfacing residues (reviewed in [286]). Any residues identified by HDX-MS could be further probed using mutagenesis in concert with binding or activity experiments in an effort to confirm their involvement in the EndoS-Fc interaction.

Overall, EndoS is a large multi-domain protein. Newly identified domains in the C-terminus of EndoS were described here and have been confirmed by the recently published

crystal structure. A catalytically inactive form of EndoS was used to form a complex of EndoS with Fc however, the studies presented here suggest that EndoS and Fc do not appear to form an ordered or rigid complex. Overall, the studies carried out in this chapter were able to provide further insight into the structure of EndoS providing clues as to which regions of this enzyme may confer IgG specificity.

4

Biophysical and structural investigation of the EndoS carbohydrate-binding module

As reviewed by Boraston *et al.* [260] CBMs have been implicated in aiding both the specificity and avidity of proteins which manipulate carbohydrates. Considering the potential role of a CBM in conferring IgG specificity to EndoS, the putative CBM of EndoS was further investigated using X-ray crystallography and other biophysical techniques. The bioinformatic prediction of a CBM in the C-terminal of EndoS was confirmed by the X-ray crystal structure of EndoS residues 98–986 [149]. This structure was subsequently used to determine the structure of EndoS residues 761–924 (CBM) to 3.00 Å resolution by molecular replacement. In addition, the monosaccharide binding functionality of CBM was probed using NMR, ITC and X-ray crystallography ligand-soaking experiments.

4.1 X-ray crystallography

Bioinformatic analysis of the sequence of EndoS revealed the presence of a potential CBM in the C-terminal of EndoS (Chapter 3). In support of this bioinformatic prediction, the results of CD and SAXS experiments suggested that this putative CBM was a globular domain (R_g=18.8 Å) with predominantly β-sheet secondary structure. Given the potential role of a CBM in dictating EndoS specificity for IgG, further verification of the structure of this region was sought. To this end, structural studies using X-ray crystallography were pursued for the CBM construct from EndoS (residues 761–924). X-ray crystallography was chosen as a method due to its potential to resolve atomic level detail of protein structures.

4.1.1 Crystallisation

CBM protein was purified according to the protocol described in Chapter 8, Section 2.4.4. Briefly, CBM was purified from *E. coli* cell lysate using IMAC followed by SEC. An example SEC chromatogram and corresponding SDS-PAGE gel of purified CBM are shown in Figure 4.1.

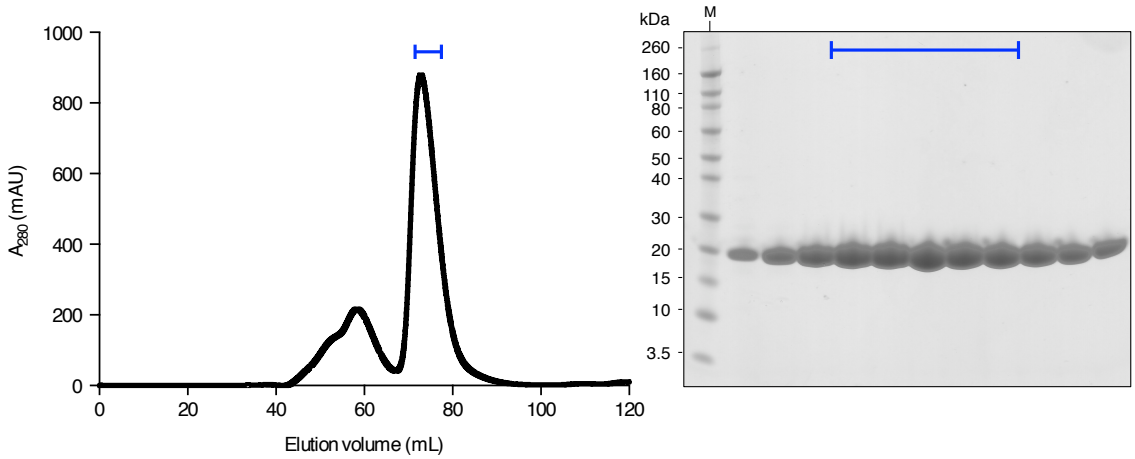


Figure 4.1: Purification of CBM. (A) A chromatogram from the final SEC step of purification using a HiPrep Superdex 75 16/60 column shows a peak at approximately 75 mL which corresponds to CBM. The horizontal blue bar above the peak indicates the fractions pooled for crystallography. (B) SDS-PAGE analysis shows the purity of the peak fractions where the horizontal blue bar above the wells indicates the fractions which were pooled for crystallisation trials.

To determine the concentration required for crystallisation, a pre-crystallisation test (PCT) was carried out using the Hampton Research PCT Kit. Given the precipitation results

of the PCT, purified CBM was concentrated to 13.8 mg mL^{-1} (A_{280} = 20.5) for use in crystallisation trials. Initial crystallisation trials were set up in 96-well plates as sitting-drop vapour diffusion experiments at both 4 °C and 18 °C. The crystallisation drops were formed of 100 nL protein solution and 100 nL of precipitant equilibrated against 50 μL of mother liquor. CBM crystallised readily across multiple crystallisation conditions (>100). Crystals grew in several needle-like morphologies ranging from small sea anemone-like clusters to long needles.

Although CBM crystallised readily, optimisation experiments were carried out with an aim to grow larger, single crystals for X-ray diffraction studies. The conditions containing long crystalline needles were used as a starting point for these optimisation experiments. Notably, several conditions from the Morpheus® HT-96 screen [224], which grew the largest needles, contained a cocktail of monosaccharides. These were chosen for optimisation with the rationale that if EndoS CBM does have carbohydrate binding functionality, these conditions may permit or encourage the growth of CBM crystals with bound substrate. Optimisation screens were made to survey both a narrow pH range and precipitant concentration around the original hit conditions described in Chapter 8, Section 2.7.1.

The crystals from optimisation trials consistently grew in the needle morphology first observed in the initial hit conditions. Needles that had larger cross-sectional dimensions were grown, but single crystals or different morphologies were not observed. In addition to native protein crystals of the CBM, crystals of SeMet CBM were grown from the optimised crystallisation conditions.

The SeMet CBM crystals used for structure determination were grown from a condition containing 0.1 M HEPES:MOPS pH 6.5, 120 mM Monosaccharides (20 mM D-galactose, 20 mM D-glucose, 20 mM D-mannose, 20 mM *N*-acetylglucosamine, 20 mM xylose and 20 mM L-fucose), 9.7% w/v PEG 8000 and 19.4% v/v ethylene glycol (Fig. 4.2A). Crystals were observed in as little as 16 hours and were harvested approximately 48 hours after initial set up of the crystallisation experiment.

4.1.2 Diffraction data collection and processing

Upon harvesting, the CBM crystals were cryoprotected by transferring them into mother liquor with an increased ethylene glycol concentration totalling 25% v/v. The crystals were then cryocooled by direct immersion into liquid nitrogen. The cryocooling process was carried out to reduce potential damage caused by synchrotron radiation during diffraction data collection (as reviewed in Garman *et al.* [270]).

Data collection from the SeMet CBM crystal used for structure determination was kindly carried out by Dr Edward Lowe (Department of Biochemistry, University of Oxford). However, prior to successful structure determination using this data, X-ray diffraction data were collected for native crystals of CBM. The native CBM crystals gave a dataset of primitive trigonal symmetry to a resolution of 3.3 Å with cell dimensions $a = b = 89.895$, $c = 59.823$ and $\alpha = \beta = 90°$, $\gamma = 120°$. Attempts at molecular replacement were unsuccessful so heavy atom derivitisation of CBM crystals was carried out in an effort to gain experimental phase information. Despite fluorescence signal being observed for Lu-soaked crystals, a lack of anomalous signal in the diffraction data and a subsequent lack of phasing power suggested there was no Lu bound to the protein structure in an ordered fashion. As a result, this data could not be used to gain experimental phase information. None of the remaining heavy atom derivatives gave sufficient diffraction for further analysis. SeMet derivatised crystals gave a higher resolution dataset (3.00 Å), however, these data were not isomorphous with data from the native crystals. At this point, the publication of the structure of EndoS [149] allowed for successful determination of the structure of the isolated CBM by molecular replacement. The highest resolution dataset from SeMet CBM crystals was used for structure determination so further data processing is discussed below with regards to this dataset.

An example X-ray diffraction image from the SeMet CBM dataset used for structure determination is shown in Figure 4.2. In general, crystals were irradiated using a helical linescan approach as depicted in Figure 4.3. Here, the X-ray beam was centred at each end of the needle-like crystal to indicate the 'start' and 'end' points of data collection. During data collection, the crystal was not only rotated but was also translated from the 'start' to the 'end' point with the result that the X-ray beam irradiated the crystal in a helical fashion. This data collection strategy was first implemented on beamline ID23-2

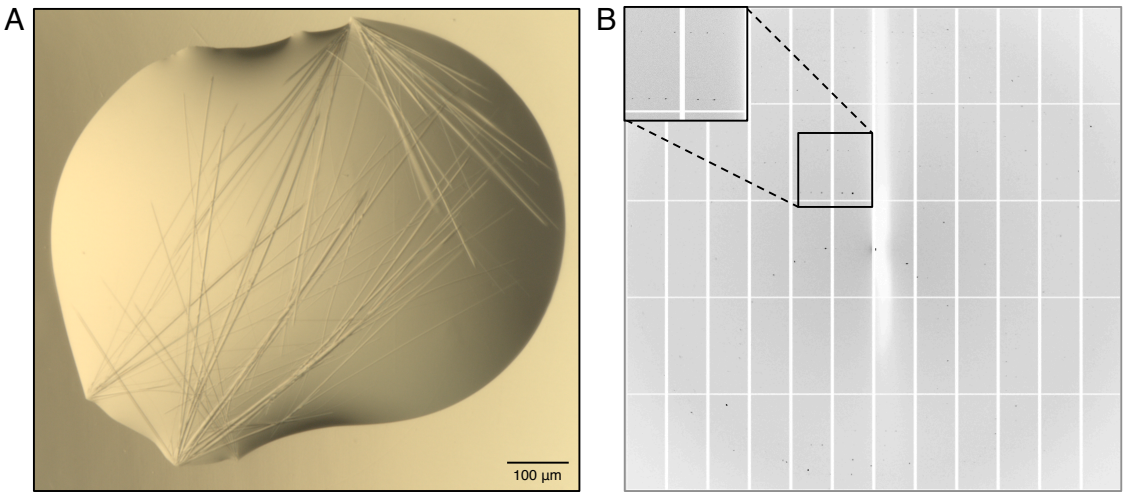


Figure 4.2: Crystallisation of CBM. (A) Needle-like crystals of CBM grown from a condition containing 0.1 M HEPES:MOPS pH 6.5, 120 mM Monosaccharides (20 mM D-galactose, 20 mM D-glucose, 20 mM D-mannose, 20 mM *N*-acetylglucosamine, 20 mM xylose and 20 mM L-fucose), 9.7% w/v PEG 8000 and 19.4% v/v ethylene glycol pictured using a light microscope. (B) An X-ray diffraction pattern from SeMet CBM crystals reveals reflections from a 0.2° oscillation. The inset panel shows a magnification of the reflections.

at the ESRF and is particularly suited for data collection from needle-like crystals. Data collection from the entire crystal can be maximised whilst the dose of radiation is spread over the length of the crystal, reducing potential radiation damage [287].

Data processing and structure determination were carried out in collaboration with Dr Pietro Roversi (Glycobiology Institute, University of Oxford). X-ray diffraction data were integrated and scaled using the programmes XDS [288] and CCP4-Aimless [289], run using the data processing suite autoPROC [225]. The presence of a 3-fold screw axis (either 3_1 or 3_2) was supported by the systematic extinctions in a plot of $I/\sigma(I)$ vs. l for Miller indices $0, 0, l$ (Fig. 4.4A). Inspection of the self-rotation function (SRF) [290] suggested that the data were consistent with the space group $P3_221$ (or its enantiomorph $P3_121$) (Fig. 4.4B).

Although crystallographic merging R factors of the diffraction data were not used to decide the resolution cut off, R_{meas} was used to aid in the assignment of a space group. R_{meas} is a statistical measure of the quality of data merging which is independent of multiplicity [291, 292] and is described in the footnotes of Table 4.1. The scaled intensity from the measurement of an individual reflection $I_{\mathbf{h}i}$ is compared to the average intensity, $\langle I_{\mathbf{h}} \rangle$, from multiple observations of that unique reflection $\mathbf{h}$ or its symmetry equivalents. A

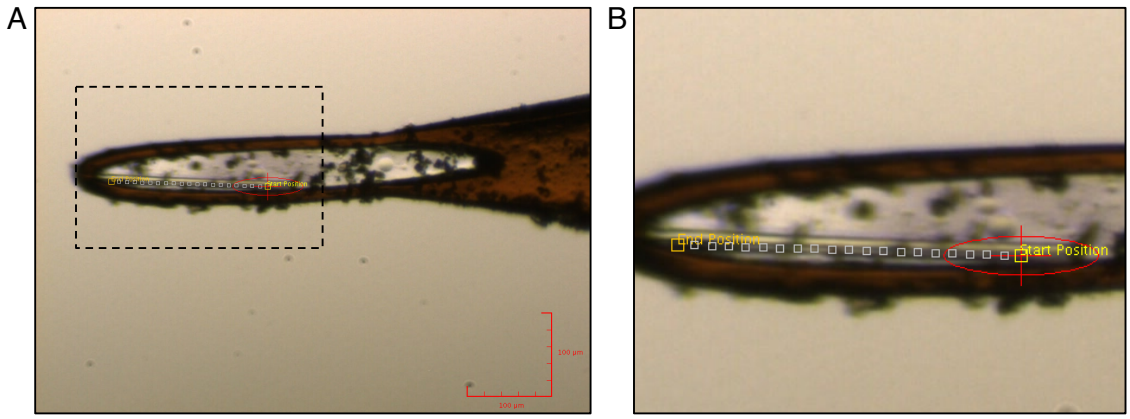


Figure 4.3: CBM linescan data collection. (A) An Elliptical LithoLoop (Molecular Dimensions) containing a SeMet CBM crystal with a dashed box indicating the position of the crystal in the loop. (B) The boxed area from panel A is magnified here to show the start and end positions of the linescan used for data collection (yellow text). White boxes indicate the trajectory of the scan with the location and approximate size of the beam indicated by the red oval and crosshairs.

term denoting the number of times that $\mathbf{h}$ and its symmetry equivalents are observed ($n_{\mathbf{h}}$) is used to account for the multiplicity of the data. If a space group is assigned to the data with higher symmetry than is actually present in the crystal, measurements of *different* reflections will be averaged to give $\langle I_{\mathbf{h}} \rangle$. Averaging reflections which are not multiple observations of the same unique reflection or its symmetry equivalents but are in fact different reflections, is liable to cause an increase in the R_{meas} value for a given dataset.

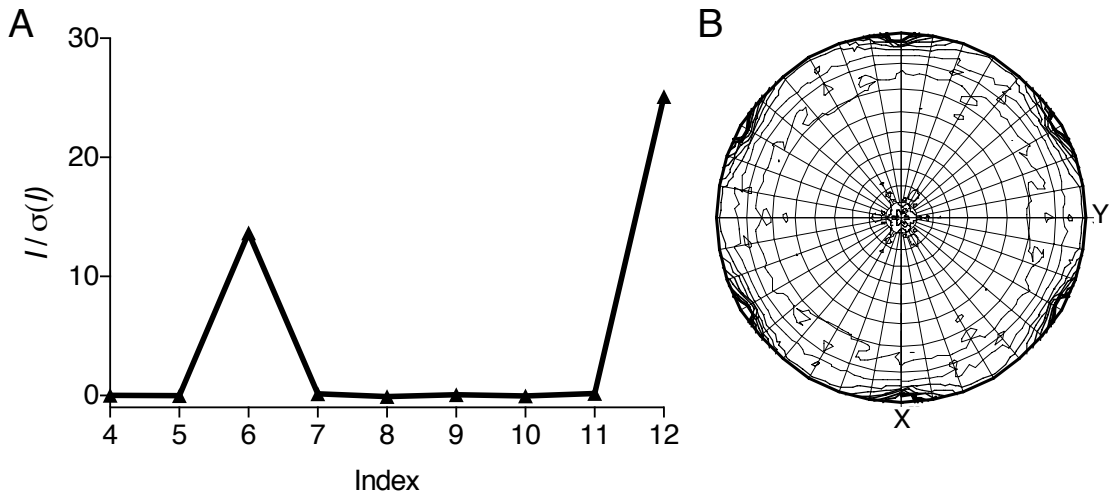


Figure 4.4: Selection of a space group for the SeMet CBM diffraction data using systematic absences and the SRF. (A) A plot of $I/\sigma(I)$ vs. l for Miller indices $0, 0, l$ reveals peaks at $l = 3n$. (B) The SRF for the data scaled in $P3_221$ is shown for $\kappa = 180°$.

A plot of R_{meas} vs. resolution showed no significant difference between the data when scaled in space group $P3_221$ (or $P3_121$) or when scaled in the lower symmetry space group $P3_2$ (or $P3_1$) (Fig. 4.5). Therefore, to make use of the increased symmetry and multiplicity, the data scaled in $P3_221$ were used for further experiments. Data quality was also assessed using the measures $CC_{1/2}$ and $\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$ as described by Karplus and Diederichs [293] and as implemented by CCP4-Aimless [289]. The resolution limit of 3.0 Å was chosen using values of $CC_{1/2} > 0.5$ and $\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle \geq 2.0$ as a guide. The resolution limit was also dictated by the crystal to detector distance in this instance as the outer edge of the detector was set at 3.00 Å. The crystallographic data collection and processing statistics are given in Table 4.1 for the data scaled in space groups $P3_2$ and $P3_221$.

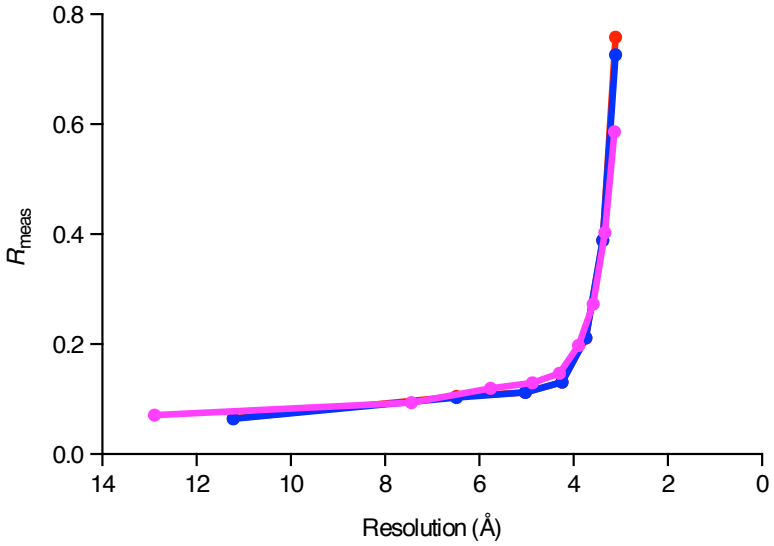


Figure 4.5: Selection of a space group for the SeMet CBM diffraction data using R_{meas}. R_{meas} was plotted against resolution for the CBM data scaled in different space groups. Data were scaled using CCP4-Aimless [289] with data shown for $P1$ (magenta), $P3_2$ (blue) and $P3_221$ (red).

4.1.3 Phasing

Initial attempts to determine the structure of CBM by molecular replacement were carried out using MrBUMP [294]. In addition, the FFAS server [295] was used to identify homologous structures which were then used as search models for molecular replacement with Phaser [227]. The top hit from FFAS was a CBM from the *Streptococcus pneumoniae* endoglycosidase, EndoD, which showed only 16% sequence similarity to the putative EndoS CBM. Despite extensive efforts using individual structures, ensembles and trunca-

Table 4.1: CBM X-ray diffraction data collection statistics.

	SeMet CBM	
Beamline	DLS-I02	
Wavelength (Å)	0.97856	
Resolution range (Å)	80.58–3.00 (3.24–3.00)	
Cell dimensions (Å)		
a, b, c (Å)	93.05, 93.05, 60.00	
α, β, γ (°)	90.0, 90.0, 120.0	
Spacegroup	$P3_2$	$P3_221$
Total reflections	92261 (16327)	96050 (20158)
Unique reflections	11650 (2400)	6242 (1258)
Multiplicity	7.9 (6.8)	15.4 (16.0)
Mosaicity (°)	0.12	0.12
Completeness (%)	99.9 (99.87)	99.9 (99.9)
$\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$	11.5 (3.1)	15.6 (4.6)
R_{merge}[†]	0.165 (0.671)	0.177 (0.734)
R_{meas}[‡]	0.176 (0.726)	0.183 (0.758)
$R_{\text{p.i.m.}}$[*]	0.062 (0.275)	0.046 (0.188)
$CC_{1/2}$	0.995 (0.932)	0.997 (0.971)

Values in parentheses refer to the outer resolution range.
[†] $R_{\text{merge}} = \sum_{\mathbf{h}} \sum_i \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}}\rangle \mid / \sum_{\mathbf{h}} \sum_i I_{\mathbf{h}}$, where $I_{\mathbf{h}i}$ is the scaled intensity of an individual measurement and $\langle I_{\mathbf{h}}\rangle$ is the average intensity from multiple observations of the unique reflection $\boldsymbol{h}$.
[‡] $R_{\text{meas}} = \sum_{\mathbf{h}} \sum_i \sqrt{(n_{\mathbf{h}}/n_{\mathbf{h}} - 1)} \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}}\rangle \mid / \sum_{\mathbf{h}} \sum_i I_{\mathbf{h}}$
[*] $R_{\text{p.i.m.}} = \sum_{\mathbf{h}} \sum_i \sqrt{(1/n - 1)} \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}}\rangle \mid / \sum_{\mathbf{h}} \sum_i I_{\mathbf{h}}$

tions of the most homologous structures, both MrBUMP and Phaser were unsuccessful in determining a molecular replacement solution.

During attempts to phase the diffraction data experimentally using SeMet derived crystals (Fig. 4.6) and heavy atom soaked crystals (Lu, Gd, Tb and Sm), the crystal structure of EndoS residues 98–995 was published (PDB ID: 4NUY, [149]). As predicted by bioinformatic analyses presented in Chapter 3, the structure contained a CBM assigned to residues 765–923. This domain of the structure, which has 100% sequence identity to CBM, was used as a search model for molecular replacement of the CBM using the programme Phaser [227].

A single monomer in the asymmetric unit was placed in $P3_221$ and gave a corresponding solvent content of 70.16% (V_m = 4.12 Å^3 Da^{-1} [296]). Although the protein monomers formed a continuous lattice, the solvent content was towards the upper limit of that found in protein crystals, and upon inspection of the initial electron density maps calculated with Phaser, a large cluster of positive density present in the electron density maps indicated

that there was likely to be a second copy of the CBM protein in the asymmetric unit. However, attempts to fit a second copy using Phaser failed due to steric clashes. At this stage the space group symmetry was lowered to $P3_2$, removing the symmetry axis upon which the cluster of positive density was positioned. Molecular replacement using Phaser [296] in $P3_2$ successfully placed three copies of CBM in the asymmetric unit giving a solvent content of 55.24% (V_m = 2.75 Å^3 Da^{-1} [296]). The symmetry relationships between these monomers are described in Section 4.1.5 and a summary of the molecular replacement statistics from Phaser is given in Table 4.2.

Table 4.2: Phaser molecular replacement statistics.

Space group	Monomer	Rotation function Z-score	Translation function Z-score	LLG*	LLG_{RB}†
$P3_221$	A	10.1	27.9	536	696
$P3_2$	A	3.9	2.9	6	268
	B	8.4	27.6	910	1271
	C	3.9	22.9	1280	1596

*Log-likelihood gain before model refinement.
†Log-likelihood gain of the model after rigid body refinement.

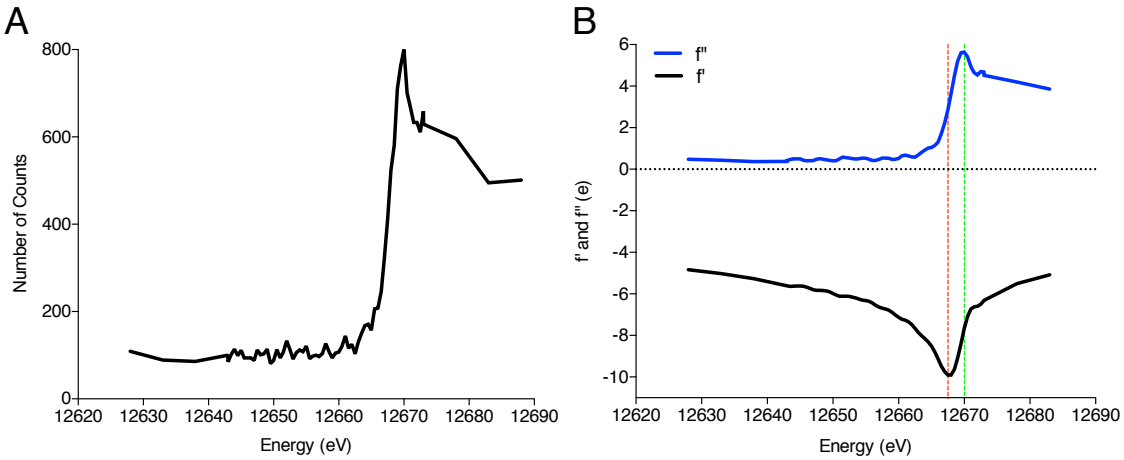


Figure 4.6: Fluorescence scan of SeMet CBM crystals. (A) The fluorescence counts from the SeMet CBM crystal were plotted against beamline energy (eV). (B) The dispersive ($f\prime$) and anomalous ($f\prime\prime$) scattering factor contributions were plotted against beamline energy (eV) with peak values at 12670.00 eV of -7.67 e and 5.64 e and inflection values at 12667.51 eV of -9.91 e and 2.86 e for $f\prime$ and $f\prime\prime$, respectively. The vertical dashed lines indicate the inflection point energy (red) and the peak energy (green). Plots of $f\prime$ and $f\prime\prime$ were calculated using CHOOCH [297] and plotted using GraphPad Prism (GraphPad Software Inc., La Jolla, U.S.A.).

The native and SeMet CBM crystals were not isomorphous, showing > 1.0% difference between their respective cell dimensions. Considering this, single isomorphous replace-

Table 4.3: SHARP refinement of SeMet residue positions

Residue		Initial coordinates			SHARP refined coordinates				
		x	y	z	x	y	z	B	Occ.
A	764	0.03797	-0.08304	-0.20969	0.0596	-0.0734	fixed*	50.2170	0.0484
	765	0.09423	0.04034	-0.21574	0.0951	0.0394	-0.2155	50.1648	0.2650
B	764	-0.38083	0.41351	0.25494	-0.3883	0.4082	0.2470	50.0614	0.0985
	765	-0.29321	0.42711	0.18164	-0.2931	0.4311	0.1791	50.0219	0.2342
C	764	0.34549	0.74379	-0.54995	0.3455†	0.7438†	-0.5460†	ND	0.0182
	765	0.32762	0.7511	-0.45088	0.3269	0.7521	-0.4493	50.1326	0.2050

*Indicates the fixed coordinate used to reference the refinement positions given the monoaxial space group of the crystal.
†Coordinates and B factors could not shift due to the occupancy falling to near zero.
ND, Not determined.

ment with anomalous scattering (SIRAS) was not able to be used to determine the phases. However, single-wavelength anomalous diffraction (SAD) was used to successfully obtain experimental phase information. Initial positions of the six SeMet residues in the asymmetric unit were assigned using the coordinates of residues SeMet 764 and SeMet 765 from the Phaser molecular replacement solution [298]. The programme SHARP [299] was then used to refine the positions, B factors and occupancies of these SeMet residues against the SAD anomalous differences. The results of this SHARP refinement are given in Table 4.3. The occupancy of residue SeMet 764 was less than 0.1 for all three molecules of the asymmetric unit. For this reason, further refinement of the heavy atom positions and calculation of the phases was carried out using only SeMet 765 from each of the three molecules in the asymmetric unit.

The quality of the phase information was assessed using both the phasing power (PP) and the figure of merit (FOM) in resolution shells (Fig. 4.7 and Table 4.4). These parameters are discussed in *'Outline of Crystallography for Biologists'* (Blow, 2002 [300]). PP relates the magnitude of the heavy atom structure factors to the error of the phases. Phasing power is considered low when PP is less than 1.0 for a given resolution shell because in this instance, the error of the phases exceeds the magnitude of the structure factor contributions of the heavy atom. The FOM is a measure of precision of the phase distribution. It can be used as a measure of confidence in a centroid phase value. For the SeMet SAD derived phases, both the PP and the FOM drop below 1.0 for the resolution bin 7.7-6.7 Å. Both statistics decrease from this resolution bin all the way to the highest resolution bin (3.1-3.0 Å) indicating that the phasing power of these data is low.

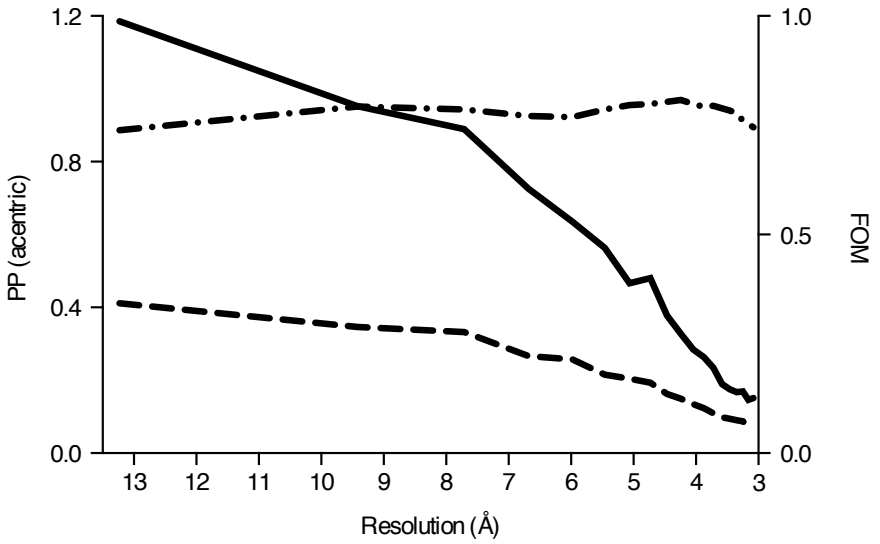


Figure 4.7: PP and FOM vs. resolution for SAD phases. The trend of PP vs. resolution is depicted as a solid black line. The FOM vs. resolution before and after density modification with SOLOMON is shown as a dashed line and a dashed and dotted line, respectively. This graph was produced with help from Dr Pietro Roversi (Glycobiology Institute, Department of Biochemsitry).

To further improve the electron density maps, density modification was carried out using the programme SOLOMON [301]. A solvent content of 51%, corresponding to 3 molecules per asymmetric unit in $P3_2$, was used for solvent flattening of the SHARP generated electron density maps (Table 4.4). Density modification using SOLOMON was monitored by assessing I1 and I4. I1 is a measure of 'contrast' between protein and solvent electron density local RMSD and should decrease during refinement. I4 is indicative of map quality and should increase during refinement. Both I1 and I4 followed the expected trend during solvent flattening. In contrast to the SHARP phases, the FOM across all resolution bins was maintained at a value close to 1.0 for the SOLOMON phases. However, upon inspection of the solvent flattened electron density maps, it was apparent that a significant proportion of the electron density for monomer C had been truncated as a result of solvent flattening. Despite the improvement in the FOM that was observed across all resolution bins (Fig. 4.7) and the correct trends of I1 and I4 during refinement, the solvent flattened phases were not used in refinement due to this loss of electron density for monomer C. Instead, the unmodified SHARP phases were used for model refinement as discussed below.

Using the SeMet SAD derived phases from SHARP, the anomalous electron density map

Table 4.4: SAD phasing statistics from SHARP and after density modification with SOLOMON.

Phasing statistic	Value
Resolution range	80.59–3.00
SHARP	
⟨PP⟩	0.32
⟨FOM⟩	0.15474
SOLOMON	
⟨FOM⟩	0.933
I1 (initial/final)†	0.7659/0.1522
I4 (initial/final)‡	0.3271/0.7147

†I1 = Local RMSD of solvent electron density (before flattening) / Local RMSD of protein electron density.
‡I4 = Overall correlation on E^2.

was calculated. Looking at this map in conjunction with the unrefined molecular replacement model, the most significant peaks of anomalous signal localised to the coordinates of the side chain of residue SeMet 765 for all three molecules in the asymmetric unit (Fig. 4.8).

4.1.4 Structure building and refinement

Iterative rounds of model refinement were carried out using the computer programme autoBUSTER [229] whilst the programme Coot [228] was used for manual model building steps. After initial rigid-body refinement of the Phaser solution, three Ca^{2+} ions were modelled into the structure (one per monomer). After initial cycles using individual atomic B factors, subsequent rounds of refinement used translation/libration/screw (TLS) restraints determined using the TLSMDweb server [302]. A total of 12 TLS bodies were defined, with 4 per monomer in the asymmetric unit. Automatically defined non-crystallographic symmetry (NCS) restraints were also applied to account for NCS present in the asymmetric unit [303]. The NCS in the asymmetric unit is discussed in the context of the model in Section 4.1.5. The Hendrickson-Lattman coefficients [304] calculated using SHARP (Section 4.1.3) were used throughout refinement to improve the model generated phases. For the last rounds of refinement, hydrogen atoms were added to the model using the programme REDUCE via the MolProbity server [305]. Refinement with riding hydrogens was carried out with an aim to improve the R factors of the model [306]. After refinement,

the geometry of the model was assessed using MolProbity [307]. The Ramachandran plot for the model of the CBM is shown in Figure 4.9. There were no Ramachandran outliers and 96.3% of residues were in Ramachandran favoured regions.

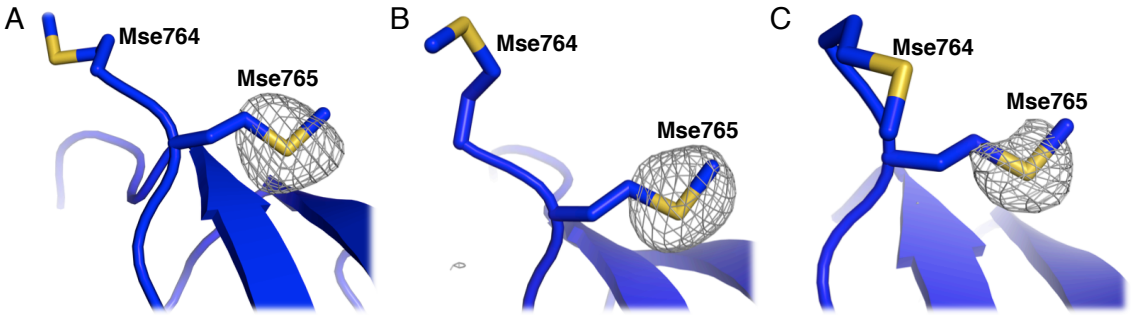


Figure 4.8: Anomalous difference map for SeMet CBM. The map was calculated using SAD phases from SeMet CBM diffraction data. Anomalous difference peaks are shown to localise to the side chain of residue 765 of (A) monomer A, (B) monomer B and (C) monomer C. The anomalous difference map is contoured at 3.0 σ.

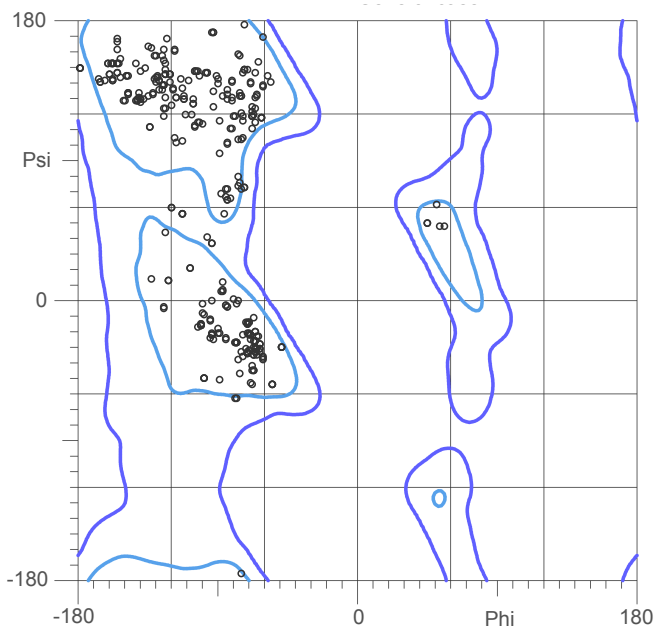


Figure 4.9: Ramachandran analysis of the refined CBM model. The plot was calculated using MolProbity [307] where black circles indicate each residue with respect to its psi angle (°) (y-axis) and phi angle (°) (x-axis). Favoured regions are outlined in light blue and allowed regions are outlined in dark blue.

4.1.5 Model of CBM

The model and refinement statistics for CBM are given in Table 4.5. The model of CBM contains 3 monomers in the asymmetric unit as shown in Figure 4.10. Each pair of monomers (AB, BC and AC) is related by a unique two-fold rotational NCS axis. The SRF

section for κ=180° was calculated from the data scaled in $P3_2$ and the peaks corresponding to each monomer pair were identified (Fig 4.11).

Table 4.5: CBM refinement statistics.

	SeMet CBM
Resolution range (Å)	80.59-3.00 (3.29-3.00)
Spacegroup	$P3_2$
Observations	11610 (2756)
Observations in free set	551 (119)
R	0.2260 (0.2209)
R_{work}†	0.2242 (0.2184)
R_{free}‡	0.2624 (0.2753)
RMSD bond lengths (Å)	0.008
RMSD bond angles (°)	1.00
Ramachandran outliers	0
Ramachandran favoured	96.3%
Residues modelled (chain)	764-927* (A)
	764-927* (B)
	764-876, 883-904, 912-926* (C)
Waters modelled	46
Wilson B factor (Å^2)	77.56
Average B, protein (Å^2)	64.83
Average B, water (Å^2)	31.46
Non-protein molecules	7 ethylene glycol, 3 Ca^{2+}

Values in parentheses refer to the outer resolution range.
†$R_{work} = \sum_{hkl} \mid \mid F_o \mid - \mid F_c \mid \mid / \sum_{hkl} \mid F_o \mid$.
‡R_{free} was calculated in an identical manner to R_{work} using 4.75% of the data which were excluded prior to refinement.
*Residues > 924 form the sequence which precedes the C-terminal hexahistidine tag and arise due to expression with the pET-303 vector.

Inspection of monomer A from the asymmetric unit shows a β-sandwich which forms a jelly-roll motif made of 4-stranded and 5-stranded sheets (Fig 4.12A, B). The strands are largely connected by unstructured loops with a small α-helical component. The secondary structure and topology of the monomer is shown schematically in Figure 4.12C. As was observed in the full-length EndoS structure, a single Ca^{2+} ion is coordinated by the backbone carbonyl groups of Lys786, Gln791 and Pro915 and the side chains of Asp789 and Glu916. Further coordination groups for the Ca^{2+} were not observed in the electron density maps, however a hydrophilic pocket adjacent to the Ca^{2+} ion is large enough to hold a water molecule.

In Figure 4.13 each monomer is coloured according to B factor showing that, in general,

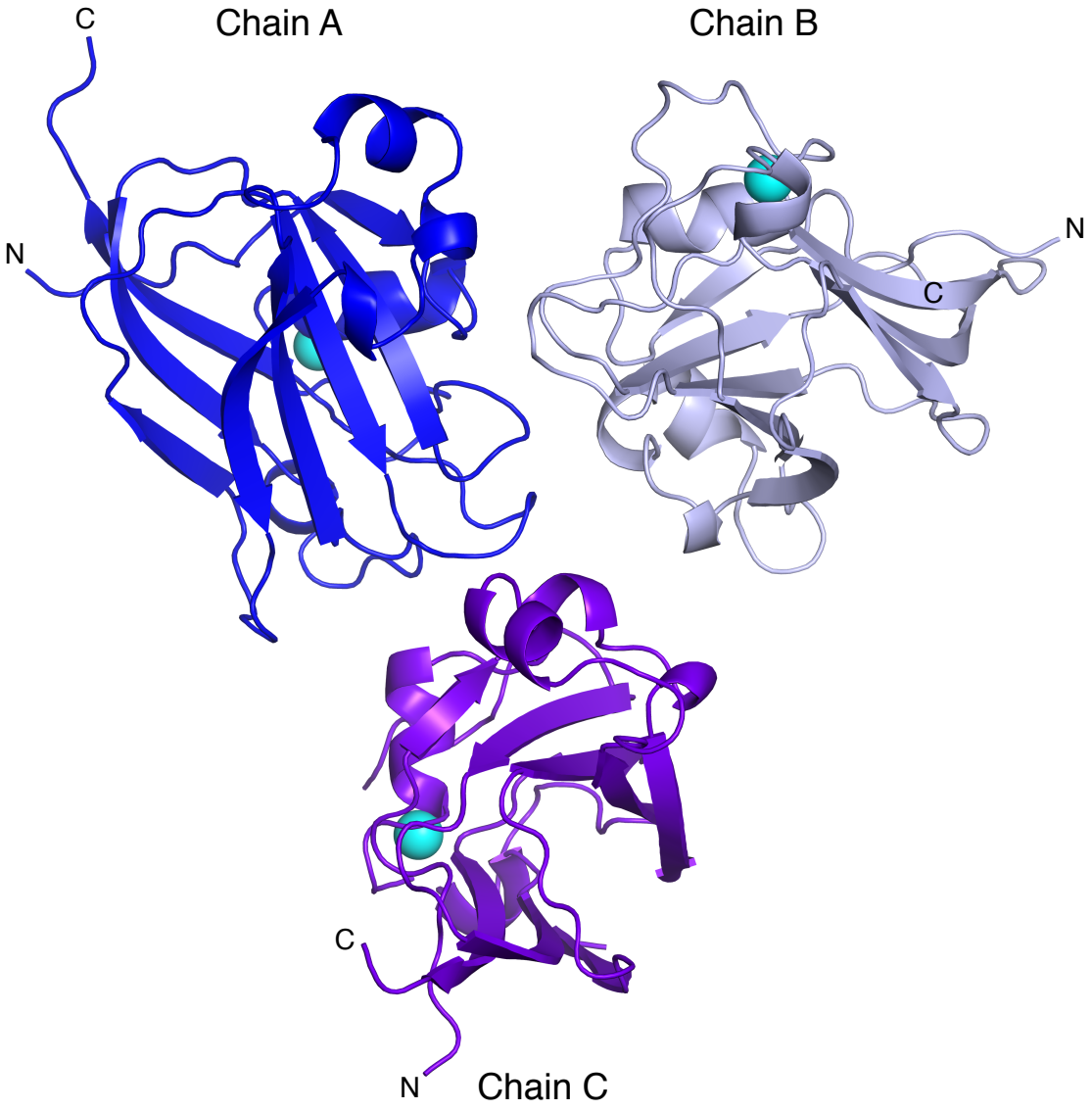


Figure 4.10: The asymmetric unit of the CBM crystal structure in $P3_2$. The protein chains are shown in cartoon representation: chain A, blue; chain B, light blue; chain C, purple. The bound Ca^{2+} ions of each monomer are shown as cyan spheres.

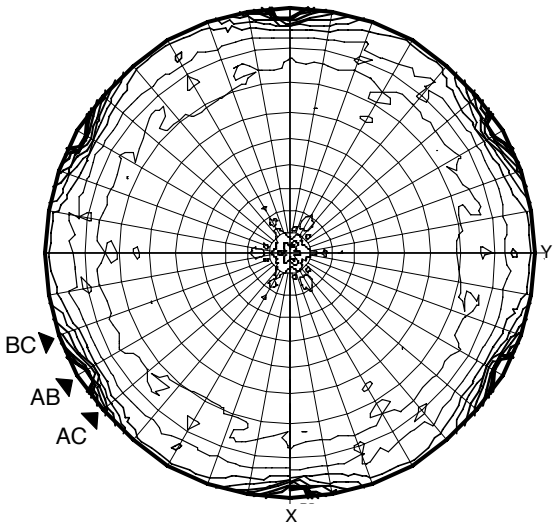


Figure 4.11: Analysis of the SRF of CBM calculated using F_o for data scaled in $P3_2$. The SRF is shown for κ=180°. Black triangles indicate which peaks correspond to each monomer pair. The directions of the NCS two-folds (to the nearest 10°) are as follows: A to B, $\theta = 90°$, $\phi = -60°$; B to C, $\theta = 90°$, $\phi = -70°$; A to C, $\theta = 90°$, $\phi = -50°$.

the loop regions of CBM have higher B factors than the regions of secondary structure. Compared to monomers A and B, the electron density for monomer C was not as well defined. This is reflected by the high B factors associated with this monomer and the weak or non-existent electron density for a number of the loops in this monomer.

The monomers of the asymmetric unit were aligned using the maximum likelihood alignment programme THESEUS [308] to give an overall RMSD of 0.01 Å (Fig. 4.13D). The loop conformations were similar for all copies in the asymmetric unit with the exception of monomer C, where several loop residues were not modelled due to poor electron density for these regions. When aligned to the CBM domain from the recently published EndoS structure (PDB ID: 4NUY, [149]) the structures have an overall RMSD of 0.07 Å. The secondary structure elements of the CBM in 4NUY aligned well with those of the CBM structures described here. The differences between the structures are all localised to the unstructured loop regions, in particular, to the loop from residues 877–885. This structural alignment shows that residues 764–924 of EndoS maintain the structure of a CBM when expressed as an isolated domain.

A structural similarity search using the DALI server was performed using the refined coordinates for monomer A [309]. Behind the structures for EndoS (both native EndoS (PDB ID: 4NUY) and SeMet EndoS (PDB ID: 4NUZ)) the best match to the CBM was

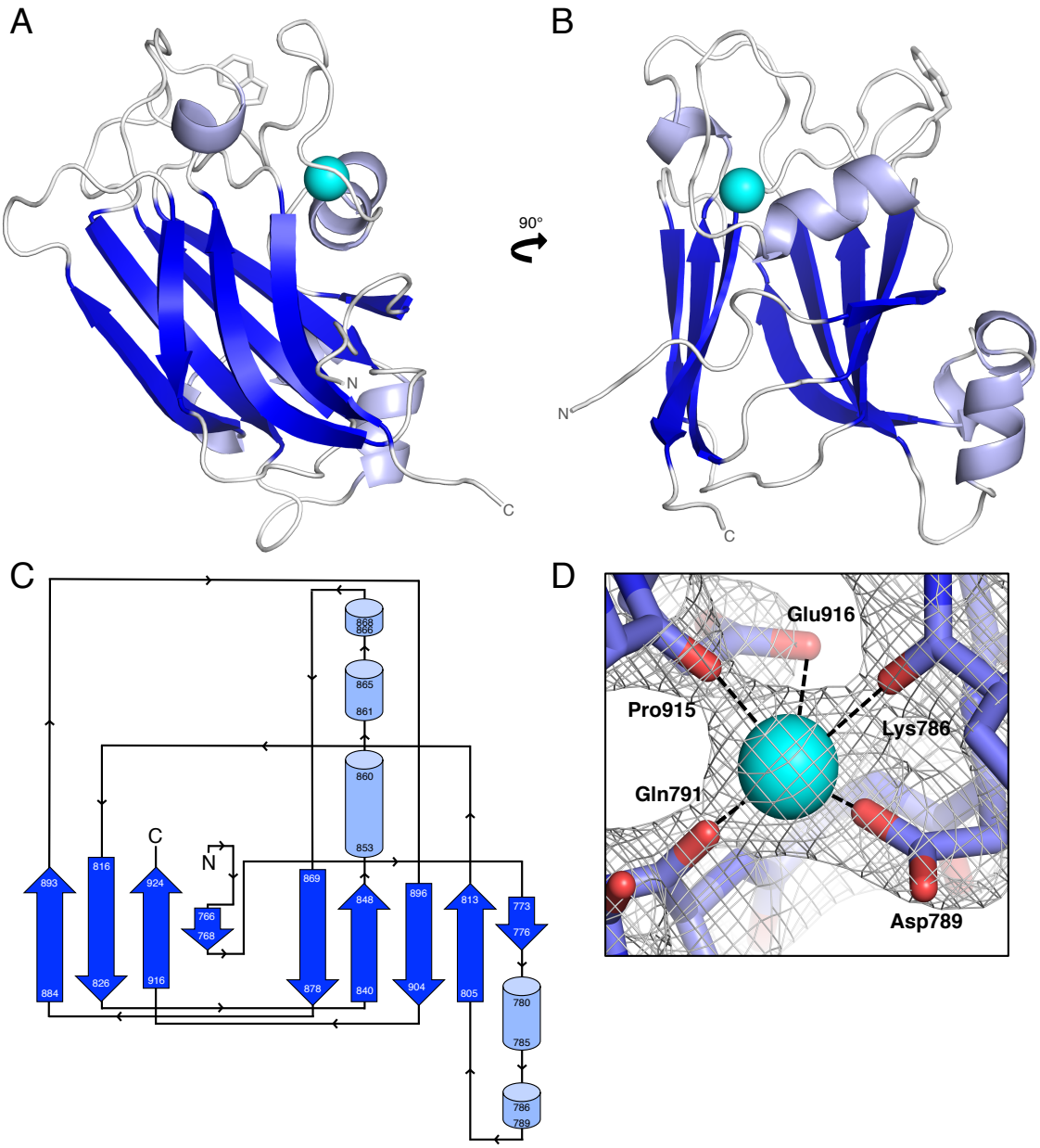


Figure 4.12: The crystal structure and topology of CBM. (A) The crystal structure of CBM monomer A is shown in cartoon representation coloured according to secondary structure (β-sheets, blue; α-helices, light blue) with the bound Ca^{2+} ion shown as a cyan sphere and Trp803 of the putative carbohydrate binding site shown in stick representation with (B) a second view showing the protein rotated through 90°. (C) A schematic representation of the topology of CBM was produced based on an output from PDBsum [48]. (D) A view of the bound Ca^{2+} (cyan sphere) of monomer A is shown with labels indicating the coordinating residues and the $2F_o - F_c$ electron density contoured at 2.0 σ.

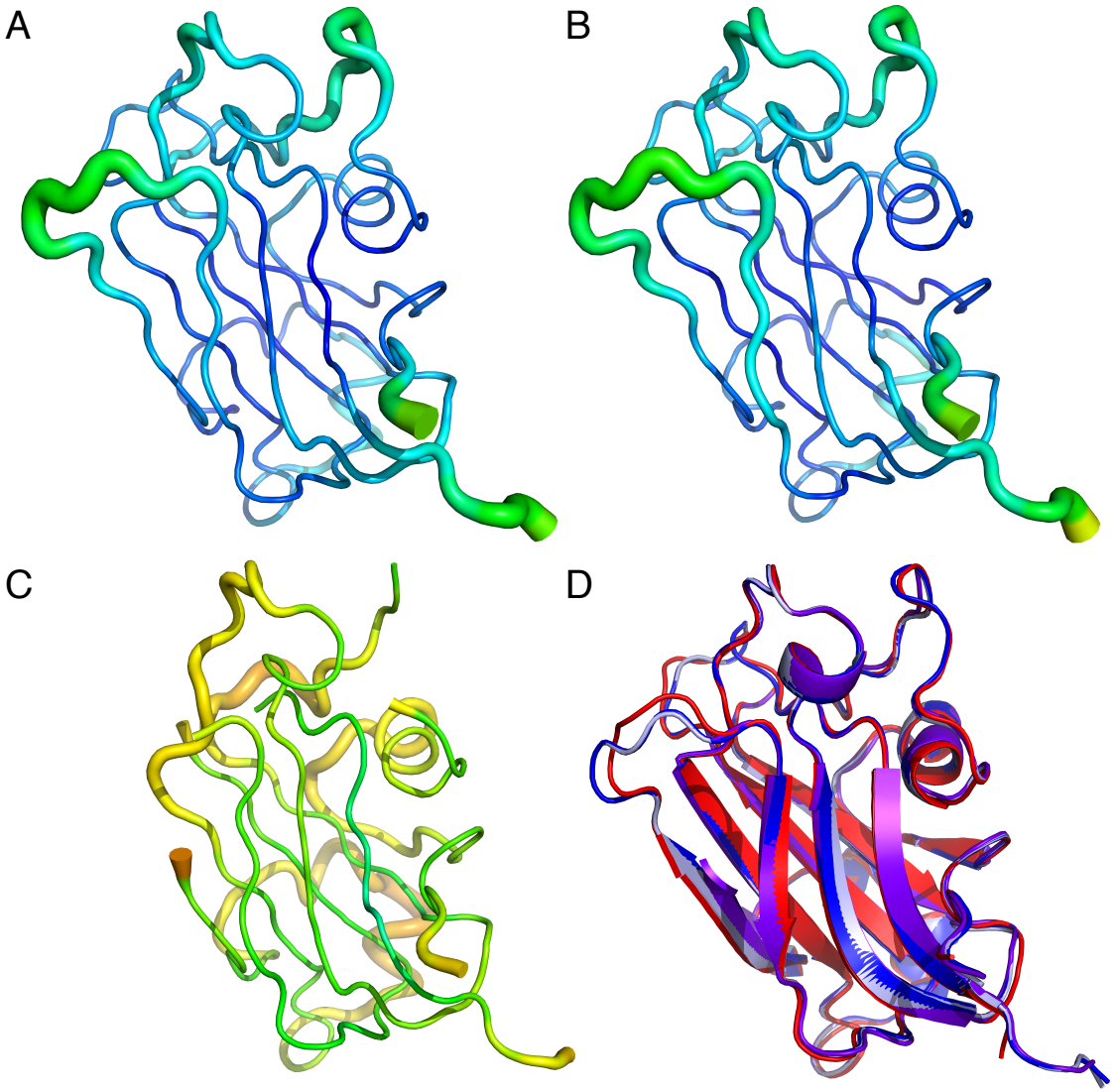


Figure 4.13: A representation of the *B* factors of the three CBM monomers (A) monomer A, (B) monomer B and (C) monomer C. All monomers are represented as tubes by drawing only the C^{α} of the polypeptide chain. Both the colour and thickness of the tubes is indicative of the *B* factor associated with the structure. Dark blue, ($\sim$ 20 Å^2); cyan, ($\sim$ 50 Å^2); green, ($\sim$ 80 Å^2); yellow, ($\sim$ 100 Å^2); red, ($\sim$ 130 Å^2). A thicker tube indicates a higher *B* factor. (D) Monomer A (blue), B (light blue), C (purple) and the CBM domain from the EndoS crystal structure (PDB ID: 4NUY [149], red) were aligned using THESEUS [308].

the structure of the CBM from EndoD, *Sp*CBM32 (PDB ID: 2XQX, [252]) with a Z-score of 14.9%, a sequence identity of 16% and an RMSD of 2.3 Å over 147 residues. EndoD, like EndoS, is an endo-β-1,4-*N*-acetylglucosamidase, but is instead produced by the close relative of *S. pyogenes*, *Streptococcus pneumoniae*. Although *Sp*CBM32 shows a number of the conserved residues required for carbohydrate-binding, no functional evidence has been documented to suggest that *Sp*CBM32 does in fact have carbohydrate-binding activity. Other top hits from the DALI search are shown in Table 4.6.

Table 4.6: DALI results.

PDB ID	Z-score	RMSD	Id.%	Protein
2XQX	14.9	2.3	16	*Sp*CBM32 from EndoD [252]
2YC2	13.4	2.6	13	Intraflagellar transport protein 25 [310]
4FI9	13.1	2.7	10	Sad1-UNC-84 homology domain [311]
1O59	12.8	3.0	10	AR2 from YIR029W allantoincase [312]
4A42	12.5	2.6	21	CBM32-4 of exo-α-D-*N*-acetylglucosaminidase [253]

4.2 Investigating the monosaccharide binding function of CBM

The potential carbohydrate-binding capacity of CBM was probed using NMR, ITC and X-ray crystallography ligand-soaking experiments.

4.2.1 Nuclear magnetic resonance

NMR spectroscopy was used to assess chemical shift perturbations of the protein spectra upon titration of monosaccharides. Isotopically ^{15}N labelled protein was expressed in minimal media using an *E. coli* expression system. The labelled CBM protein could be purified in an identical manner to the unlabelled sample for subsequent use in NMR spectroscopic measurements. The CBM gave a dispersed ^{15}N-1H HSQC spectrum indicative of a folded domain, in support of both the results of the CD experiments and the bioinformatic analysis (Fig. 4.14).

A series of HSQC spectra were collected for the CBM in the presence of increasing concentrations of monosaccharides (Fig. 4.15). When aligned against an untreated sample, chemical shift perturbations were apparent upon addition of D-galactose and D-glucose.

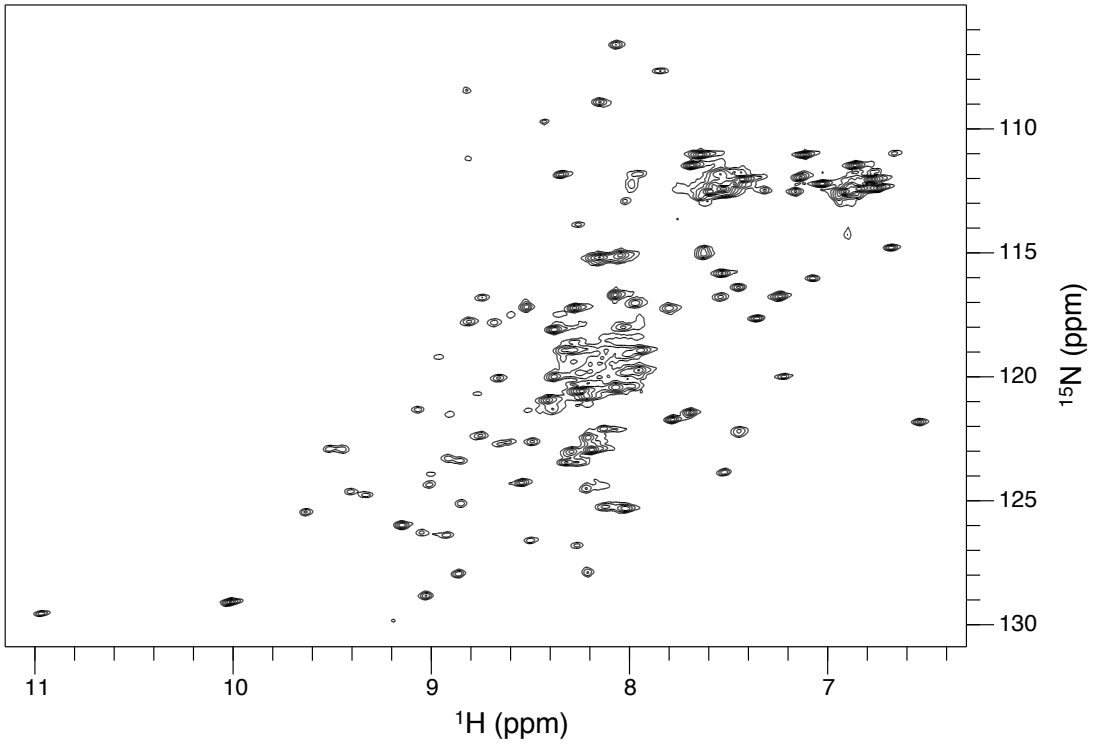


Figure 4.14: A ^{15}N HSQC spectrum of CBM. The spectrum was collected at 37 °C using a 0.15 mM solution of CBM in 10 mM BisTris pH 6.5, 15 mM NaCl.

Considering the large molar excess of monosaccharide used in these experiments, L-fructose was used as a control to ascertain if these changes were specific for these monosaccharides. It was apparent that some of these perturbations to the spectra also occurred with titration of L-fructose. These non-specific perturbations are illustrated in Figure 4.16. However, 3 distinct peaks appeared upon D-galactose and D-glucose titration, which were visible from the lowest titration point. These peaks did not appear in the L-fructose treated sample and were not present in the untreated sample.

Titration with *N*-acetylglucosamine did not result in any specific chemical shift perturbations. The two peaks that appeared upon titration with *N*-acetylglucosamine could be attributed to the chemical shifts visible for *N*-acetylglucosamine alone. The smeared appearance of these peaks was due to shimming difficulties during data acquisition.

4.2.2 Isothermal titration calorimetry

ITC was used to further investigate the potential D-galactose binding ability of the CBM (Fig. 4.17). Although an endothermic event was evident upon titration of D-galactose to

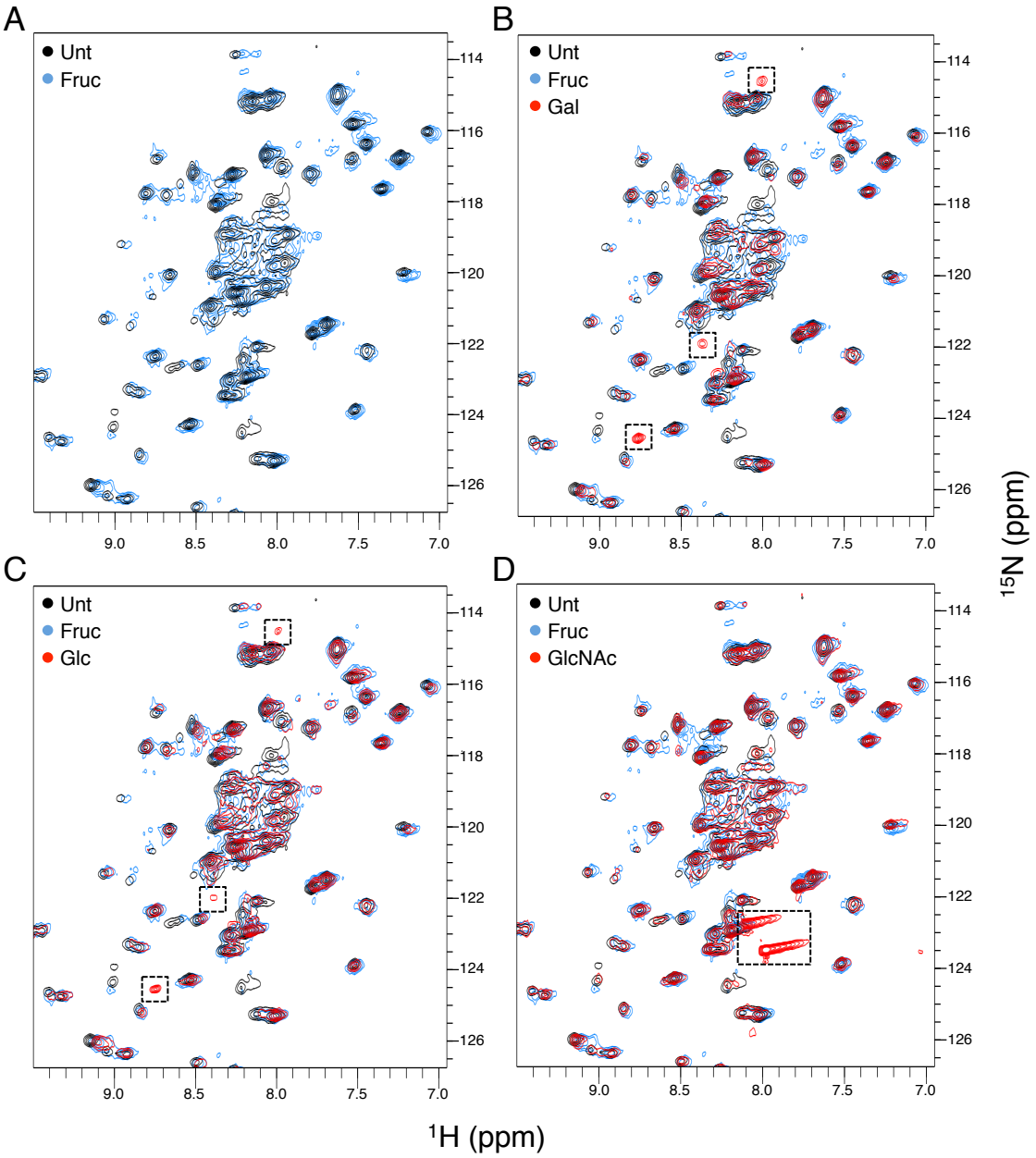


Figure 4.15: NMR analysis of CBM in the presence of monosaccharides. (A) A magnified region of the ^{15}N-HSQC spectra is shown for the untreated CBM (Unt, black) and the CBM titrated with 1 M L-fructose (blue). The same spectra are shown in the remaining panels with addition spectra in red overlaid for the CBM titrated with (B) 1 M D-galactose, (C) 1 M D-glucose and (D) 250 mM *N*-acetylglucosamine. The protein chemical shifts which are unique to the D-galactose and D-glucose treated samples are indicated by dashed black boxes. The monosaccharide chemical shifts due to the N atom of *N*-acetylglucosamine are indicated by a dashed black box in panel D.

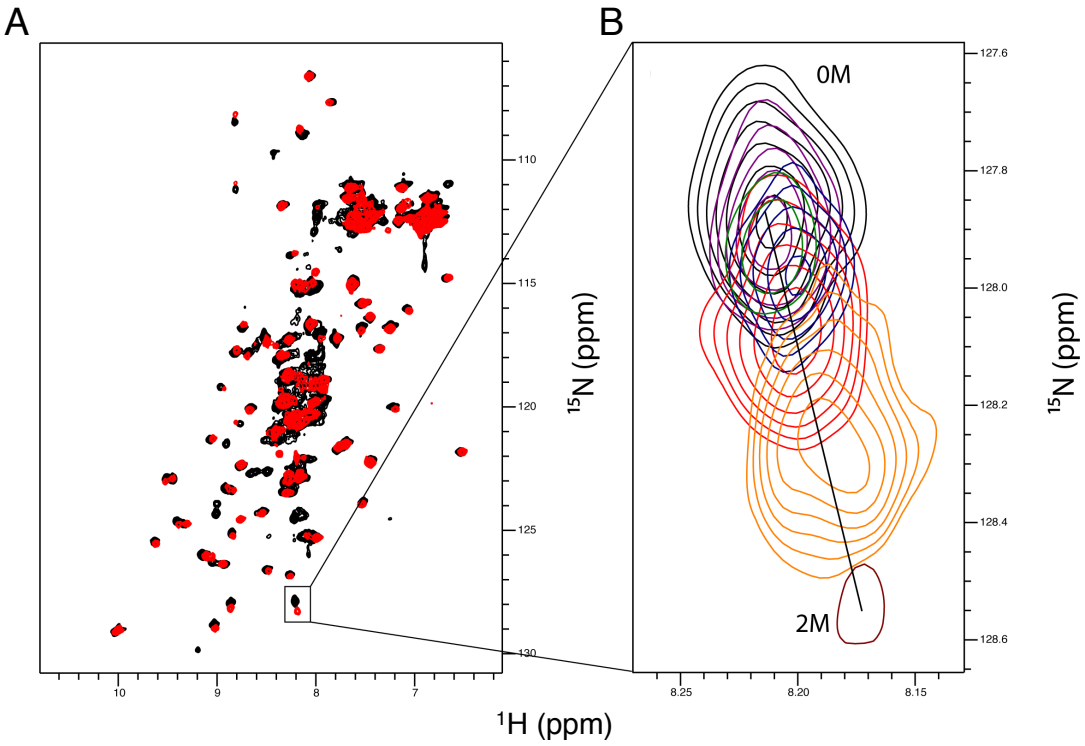


Figure 4.16: Non-specific perturbation of CBM chemical shifts induced by monosaccharide titrations. (A) The ^{15}N-HSQC spectra for untreated CBM (black) and CBM with 1 M D-galactose (red) showing non-specific chemical shift perturbations upon addition of D-galactose with (B) a magnification of one such example over the titration points of 0 mM, 62.5 mM, 125 mM, 250 mM, 500 mM, 1 M and 2 M D-galactose.

the CBM, this change could be attributed to the heat of dilution of the CBM sample in the cell, which displayed a similar endothermic profile when PBS alone was added.

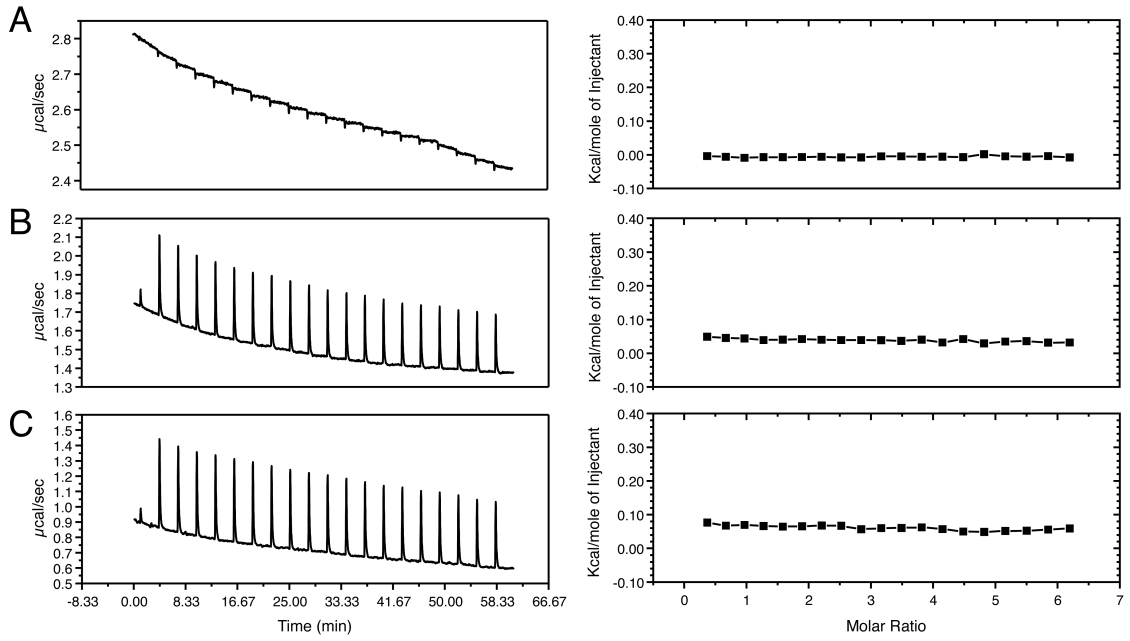


Figure 4.17: Isothermal calorimetry titrations of the CBM with D-galactose. The left hand panels indicate the raw data collected during titration of CBM (1.4 mM) with repeated injections of 42 mM D-galactose. The right hand panels show the corresponding enthalpy changes of the integrated peaks plotted as the change in kcal mole^{-1} of injectant (PBS or D-galactose) vs. the molar ratio of CBM to D-galactose. (A) The titration of D-galactose into PBS shows a negligible exothermic change in enthalpy attributable to the heat of dilution of the D-galactose solution into PBS. (B) The titration of D-galactose into a solution of CBM gives endothermic changes in enthalpy at each injection point. (C) The titration of PBS into a solution of CBM also shows endothermic changes in enthalpy at each injection point, comparable to those, which occur upon addition of D-galactose. None of the titrations give an isotherm curve indicative of a binding event (right hand panels).

4.2.3 X-ray crystallography

Having successfully crystallised and determined the structure for CBM using X-ray crystallography, experiments were carried out in an attempt to crystallise CBM with potential substrates bound. This method was pursued considering that homologous CBM domains have previously been crystallised with either di- or monosaccharides bound in their active sites [284, 313].

For these experiments, crystallisation trials were set up using the original crystallisation condition supplemented with 120 mM solutions of either D-galactose, D-glucose or D-mannose in the place of the cocktail of monosaccharides. Crystals were grown for all three

new conditions and showed the same needle-like morphology as observed for the original crystallisation condition (Fig. 4.18).

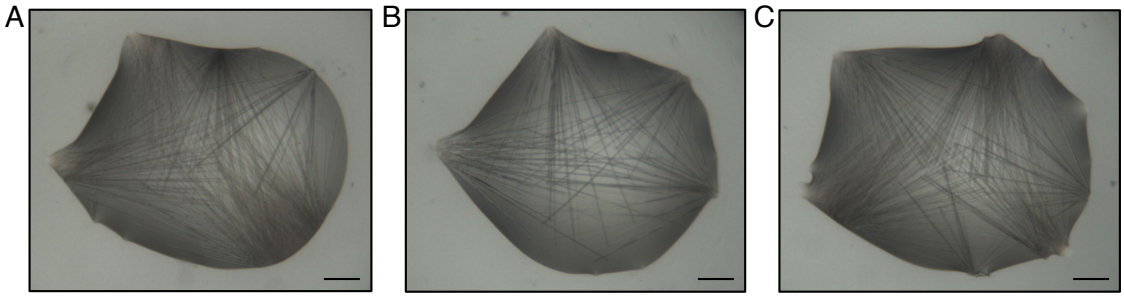


Figure 4.18: Crystallisation of CBM using individual monosaccharides. Needle-like crystals of CBM are shown for conditions containing 0.1 M HEPES:MOPS pH 6.5, 9.7% w/v PEG 8000 and 19.4% v/v ethylene glycol supplemented with (A) 120 mM D-galactose, (B) 120 mM D-glucose or (C) 120 mM D-mannose. The pictures were taken using a Minstrel^{TM} HT (Rigaku) and the scale bar on each image represents 100 μm.

The crystals were harvested and cryoprotected in a similar manner to the original crystals described earlier in this chapter. Data collection was carried out at DLS beamline I03 using the helical linescan method described in Section 4.1.2. For the crystal grown from the condition containing 120 mM mannose, diffraction data was collected using a standard approach as the crystal was too small to carry out linescan data collection. The data collection statistics are shown in Table 4.7 for crystals grown from conditions containing D-galactose (CBMGal), D-glucose (CBMGlc) and D-mannose (CBMMan). Data processing was carried out following the same protocol as for the original diffraction data (Section 4.1.2) where data were processed to 2.99 Å (CBMGal), 2.90 Å (CBMGlc) and 2.91 Å (CBMMan).

The structures were determined by molecular replacement using the programme Phaser [227]. In all cases, the entire asymmetric unit of the refined CBM structure described in Section 4.1.5 was used as the search model. Refinement of the Phaser derived models was carried out using BUSTER [229]. Here, an initial rigid body refinement was performed followed by a round of refinement which introduced TLS and NCS restraints. The TLS bodies and NCS restraints were defined in an identical manner to those used in refinement of the CBM structure described earlier in this chapter (Section 4.1.4). The refinement and model statistics for these structures are listed in Table 4.8 and the Ramachandran plots for each structure (calculated using the MolProbity server [307]) are shown in Figure 4.19. Inspection of the electron density maps for these structures showed no positive density in

the putative binding site of the CBM, for this reason, further refinement of the structures was not pursued.

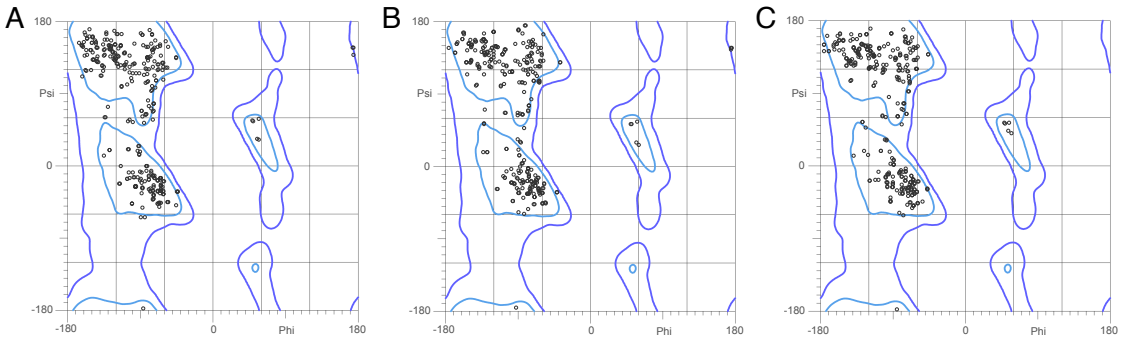


Figure 4.19: Ramachandran analysis of CBM monosaccharide soaks. The plots for (A) CBMGal, (B) CBMGlc and (C) CBMMan were calculated using MolProbity [307] where black circles indicate each residue with respect to its psi angle (°) (y-axis) and phi angle (°) (x-axis). Favoured regions are outlined in light blue and allowed regions are outlined in dark blue.

Table 4.7: Diffraction data collection statistics for CBM monosaccharide soaks.

	CBMGal	CBMGlc	CBMMan
Beamline	I03	I03	I03
Wavelength (Å)	0.9763	0.9763	0.9763
Resolution range (Å)	59.88-2.99 (3.66-2.99)	46.68-2.90 (3.13-2.90)	48.10-2.91 (3.56-2.91)
Cell dimensions (Å)			
a, b, c (Å)	93.33, 93.33, 59.88	93.36, 93.36, 59.87	93.15, 93.15, 59.91
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Spacegroup	$P3_2$	$P3_2$	$P3_2$
Total reflections	34319 (15627)	58328 (10629)	41920 (19122)
Unique reflections	11683 (5315)	12918 (2675)	12697 (5794)
Multiplicity	2.9 (2.0)	4.5 (4.0)	3.3 (3.3)
Mosaicity (°)	0.13	0.21	0.12
Completeness (%)	99.2 (99.4)	99.8 (99.6)	99.4 (99.3)
$\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$	7.4 (4.0)	11.5 (2.8)	8.1 (3.8)
R_{merge} [†]	0.117 (0.256)	0.130 (0.616)	0.122 (0.320)
R_{meas} [‡]	0.164 (0.361)	0.166 (0.788)	0.172 (0.452)
$R_{\mathrm{p.i.m.}}$ [*]	0.086 (0.189)	0.077 (0.374)	0.090 (0.236)
$CC_{1/2}$	0.987 (0.948)	0.985 (0.759)	0.988 (0.922)

Values in parentheses refer to the outer resolution range.
[†] $R_{\mathrm{merge}} = \sum_{\mathbf{h}} \sum_i \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}} \rangle \mid / \sum_{\mathbf{h}} \sum_i I_{\mathbf{h}}$, where $I_{\mathbf{h}i}$ is the scaled intensity of an individual measurement and $\langle I_{\mathbf{h}} \rangle$ is the average intensity from multiple observations of the unique reflection $\boldsymbol{h}$.
[‡] $R_{\mathrm{meas}} = \sum_{\mathbf{h}} \sum_i \sqrt{(n_{\mathbf{h}}/n_{\mathbf{h}} - 1)} \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}} \rangle \mid / \sum_{\mathbf{h}} \sum_i I_{\mathbf{h}}$
[*] $R_{\mathrm{p.i.m.}} = \sum_{\mathbf{h}} \sum_i \sqrt{(1/n - 1)} \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}} \rangle \mid / \sum_{\mathbf{h}} \sum_i I_{\mathbf{h}}$

Table 4.8: Refinement statistics for CBM monosaccharide soaks.

	CBMGal	CBMGlc	CBMMan
Resolution range (Å)	48.12-2.99 (3.27-2.99)	46.68-2.90 (3.13-2.90)	48.10-2.91 (3.19-2.91)
Spacegroup	$P3_2$	$P3_2$	$P3_2$
Observations	11670 (2754)	12900 (2634)	12668 (3006)
Observations in free set	559 (123)	706 (149)	613 (136)
R	0.2269 (0.2367)	0.2226 (0.2345)	0.2240 (0.2463)
R_{work}[†]	0.2244 (0.2337)	0.2197 (0.2298)	0.2213 (0.2440)
R_{free}[‡]	0.2761 (0.3022)	0.2728 (0.3123)	0.2776 (0.2960)
RMSD bond lengths (Å)	0.010	0.010	0.010
RMSD bond angles (°)	1.10	1.11	1.11
Ramachandran outliers	0	0	0
Ramachandran favoured	94.6%	94.4%	96.3%
Residues modelled (chain)	764-927* (A)	764-927* (A)	764-927* (A)
	764-927* (B)	764-927* (B)	764-927* (B)
	764-876, 883-904, 912-926* (C)	764-876, 883-904, 912-926* (C)	764-876, 883-904, 912-926* (C)
Wilson B factor (Å^2)	74.33	81.61	72.44
Average B, protein (Å^2)	55.88	61.11	53.63

Values in parentheses refer to the outer resolution range.
[†]$R_{work} = \sum_{hkl} \mid \mid F_o \mid - \mid F_c \mid \mid / \sum_{hkl} \mid F_o \mid$.
[‡]R_{free} was calculated in an identical manner to R_{work} using approximately 5% of the data which were excluded prior to refinement.
*Residues > 924 form the sequence which precedes the C-terminal hexahistidine tag and arise due to expression with the pET-303 vector.

4.3 Discussion

Bioinformatic analysis in Chapter 3 identified a putative CBM in the C-terminal region of EndoS. This region was shown to have homology to CBM family 32, in particular to CBMs which bind galactose with mM affinity [284, 251]. In this chapter, further investigation of the putative CBM from EndoS using X-ray crystallography revealed that residues 764–924 of EndoS do form a CBM domain. The carbohydrate-binding functionality of the CBM of EndoS was investigated using NMR, ITC and X-ray crystallography and the data presented in this chapter suggest that if EndoS CBM does bind carbohydrates it is likely to do so with low affinity.

The purified CBM protein was readily crystallisable but only ever formed needles or needle-like crystal morphologies. Despite the small size and narrow dimensions of the crystals, X-ray diffraction data were collected using a linescan approach. Diffraction data were originally scaled in space group $P3_221$, however, when phasing attempts using molecular replacement were carried out, it became apparent that the crystals were of lower symmetry and were rescaled to $P3_2$. The NCS of the asymmetric unit contributed to the ambiguity of the SRF. The asymmetric unit of the CBM crystals contained three monomers where a different two-fold rotational symmetry operator relates each monomer pair but all the NCS two-folds are very close to the crystallographic two-fold of $P3_221$. Nevertheless, the correct space group of $P3_2$ was eventually assigned to the data for further processing.

Initial attempts to phase the CBM diffraction data by molecular replacement were unsuccessful. This was likely due to both the poor sequence identity of the homologous structures produced using FFAS where the best hit shared only 16% sequence identity. The incorrect assignment of the space group is likely to have also impeded molecular replacement efforts. Although anomalous signal was observed for SeMet CBM crystals, the data were not sufficient to phase the entire CBM structure using SAD. This was due to the small number of SeMet residues (2 per monomer) and the partial occupancy of the SeMet residues (20.5–26.5%). The low anomalous signal of the SeMet residues is potentially due to the N-terminal location of these residues and may also be due to partial incorporation of SeMet during expression of CBM.

During attempts to determine the phases of the CBM diffraction data, the structure of

EndoS was published. This structure confirmed the presence of a CBM in the C-terminal of EndoS and also provided a molecular replacement model for the diffraction data collected here. Structure determination of the isolated CBM domain was pursued in the light of the full-length EndoS structure to ascertain if CBM maintains the fold of a CBM when expressed as an isolated domain and with a hope that the structure may provide further insight towards carbohydrate-binding functionality.

The structure of CBM was determined using molecular replacement to 3.00 Å resolution with $R_{\text{work}} = 0.22$ and $R_{\text{free}} = 0.26$ and 96.3% of residues in Ramachandran favoured regions. Although the SeMet SAD phases were not sufficient to determine the phases, the Hendrickson-Lattman coefficients [304] calculated using SHARP were used throughout refinement of the CBM model. The structure was highly similar to the structure of EndoS CBM in the context of the full-length enzyme (PDB ID: 4NUY) with an RMSD of 0.07 Å. This is supportive of the 'beads on a string' model of EndoS domain organisation as described by Trastoy *et al.* [149]. The stability of the structure of the isolated CBM domain indicates how the individual domains of EndoS may have existed as individual proteins before evolution to give the multi-domain conformation of EndoS.

DALI analysis of monomer A from the CBM structure suggests that the CBM from EndoS has homology to a number of non-carbohydrate-binding CBMs from proteins with various biological roles. However, amongst the homologous proteins are CBMs from *Streptococcus pneumoniae* EndoD [252] and from a *Micromonospora viridifaciens* sialidase [313]. The later displays galactose binding affinity and is important for substrate avidity and enzyme activity. Given the potential role of CBM in determining EndoS specificity for IgG glycans, the ability of CBM to bind carbohydrates was further investigated using NMR, ITC and further crystallography experiments.

Analysis of CBM by NMR gave a dispersed ^{15}N-^{1}H HSQC spectrum indicative of a folded domain. NMR titrations with monosaccharides indicated possible monosaccharide binding capabilities, manifested as the appearance of new peaks; however, ITC experiments could not detect measurable binding of D-galactose to the CBM assuming a Kd of 1 mM. Furthermore, none of the CBM crystals grown in the presence of monosaccharides showed any indication of bound substrate in their putative binding sites. Although the D-galactose and D-glucose induced peak appearances in the HSQC spectra were indicative of a specific

interaction, the inability to detect an interaction via ITC suggests that any interaction was low affinity, as is typical of CBMs [260]. Alternatively, these peaks may arise as an artifact of the large molar excess of D-galactose and D-glucose used in NMR titrations. However, considering these peaks were not evident in either the D-galactose, D-glucose only (data not shown) or the L-fructose CBM control experiments, it is conceivable the CBM can specifically bind monosaccharides, but with low affinity.

The inability to detect a reliable interaction between CBM and monosaccharides may be due to the fact that CBM interacts with larger oligosaccharides. Considering the *N*-linked glycan substrate of EndoS, it seems reasonable to suggest that multiple interactions with individual monosaccharide subunits of the glycan may contribute to potential CBM carbohydrate-binding affinity. These potential interactions could be explored and tested using glycan arrays with the caveat that the typically low affinity interactions of CBMs with their carbohydrate substrates may complicate the use of this type of technology. Alternatively, a larger repertoire of monosaccharides and disaccharides could be explored using UV based assays which have previously been used to investigate CBM carbohydrate binding (reviewed in [260]).

Overall, the structure of CBM was able to corroborate the bioinformatic prediction that EndoS contained a C-terminal CBM. Efforts were made to try and identify potential ligands of CBM, however, further investigation is required to consolidate the observations seen in these experiments and to determine the ligands, if any, of the EndoS CBM.

5

Substrate specificity and functional characterisation of EndoS

Considering the potential scope for EndoS use in both therapeutic and transglycosidase applications, experiments in this chapter were carried out with an aim to understand how its apparent specificity is conferred and what factors might be important for EndoS glycoside hydrolase activity towards IgG. In this chapter, EndoS was shown to be catalytically active against a number of smaller IgG fragments including the isolated CH2 domain of Fc. Functional analysis of truncations of EndoS revealed that, whilst the C-terminal of EndoS is dispensable for activity, its deletion impedes the hydrolysis of IgG glycans. The glycan specificity of EndoS was explored in the context of a number of engineered Fc glycoforms. Further to this, the specificity of EndoS for Fc glycans over Fab IgG glycans was investigated. Finally, it was shown that EndoS has glycoside hydrolase activity against glycoproteins other than IgG.

The HPLC and MS methods used in this chapter represent the structures of glycans following Oxford glycan nomenclature [83] combined with the colour scheme of the Consortium for Functional Glycomics as previously implemented [91, 87]. Figure 5.1 summarises this notation.

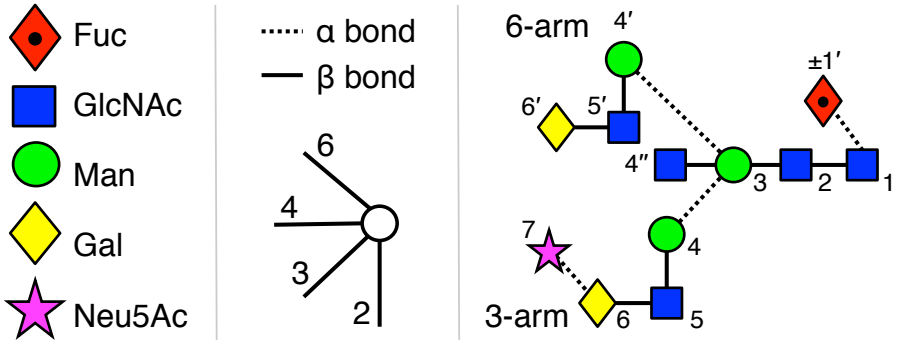


Figure 5.1: Schematic representation of *N*-linked glycans. The left panel indicates the symbols with their respective colours for the monosaccharide components of *N*-linked glycans discussed in this thesis. The central panel depicts the notation used for bonds where dashed lines indicate α linkages and solid lines indicate β linkages. The angle of the line connecting two residues indicates the type of linkage as either 1–6, 1–4, 1–3 or 1–2 (from the reducing to non-reducing end of the glycan). The right hand panel describes the numbering system used to identify glycan positions within the 3-arm and 6-arm on a typical *N*-glycan.

5.1 Investigation of enzyme and substrate components required for activity

5.1.1 EndoS activity against IgG fragments

As an initial step towards understanding the specificity of EndoS for IgG, the minimal protein unit required for deglycosylation of IgG by EndoS was considered. A number of constructs of the IgG1 Fc domain were designed to elucidate protein-protein interactions that might be important for EndoS activity against IgG. In addition to an already described Fc domain construct [87], further constructs were made to encompass the CH2 domain inclusive of the hinge region, capable of disulphide bonding (CH2-H) and the CH2 domain excluding the hinge region, predicted to exist as isolated monomeric domain (CH2). A schematic of these constructs is shown in Figure 5.2A. All of these constructs could be transiently expressed in HEK 293T cells. In addition, all constructs were soluble and could be purified to $> 95\%$ purity as judged by SDS-PAGE (Fig. 5.2B).

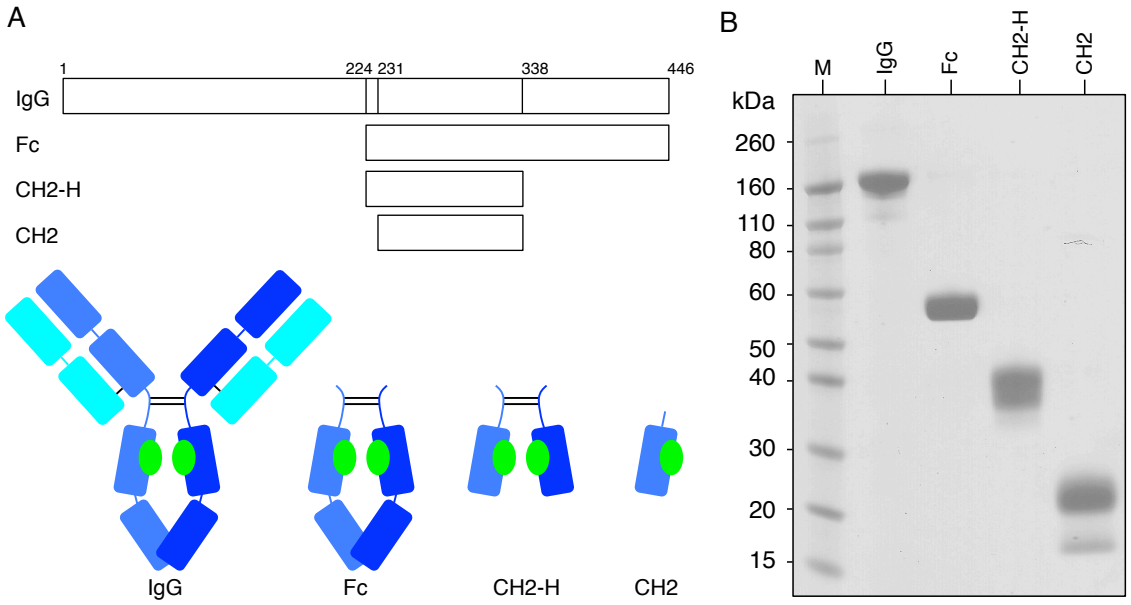


Figure 5.2: Purification of IgG fragments. (A) The upper diagram shows the residue boundaries for the IgG fragment constructs which numbered in reference to the full-length sequence of the heavy chain of human IgG1. Cartoon illustrations of the protein fragments indicate the predicted dimerisation state of each construct. (B) SDS-PAGE analysis of purified IgG fragments indicates the purity of the samples after the final SEC purification step.

Glycoprofiling of IgG fragments

To assess the relevance of these constructs as targets for EndoS deglycosylation, we determined their overall pattern of glycosylation (Fig. 5.3). The total glycan pool was released from each of the IgG fragments using PNGaseF. The resultant glycans were then assessed using HPLC. More detailed analyses of the glycans present on each construct were carried out via MALDI-TOF MS and negative ion nano-ESI MS (discussed below). The glycan repertoire of the IgG fragments included the majority of the complex biantennary glycans routinely identified on human serum IgG. Compared to serum IgG, the IgG fragments had varying degrees of both galactosylated and bisected glycans. The CH2-H and CH2 fragments both contained hybrid-type glycans and a peak corresponding to either a bisected triantennary or tetraanntenary species. The expressed Fc fragment shows reduced glycan processing when compared to human serum IgG.

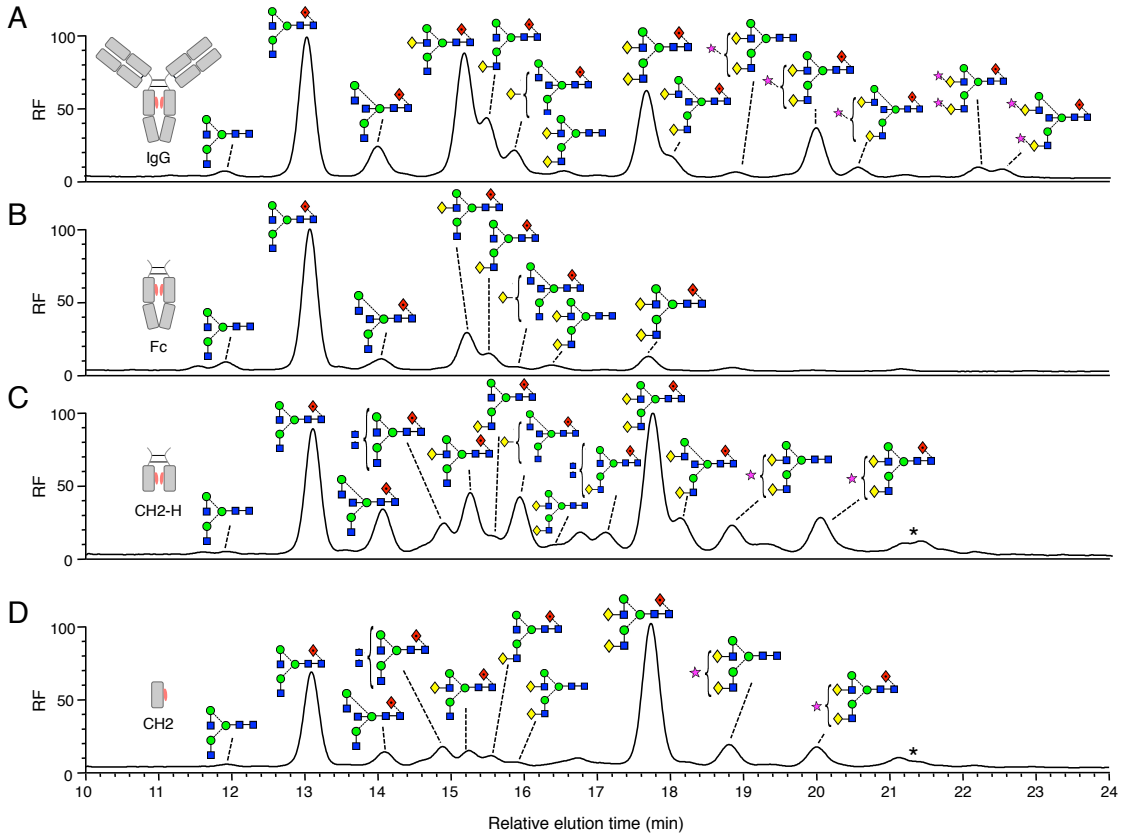


Figure 5.3: Glycan profiling of serum IgG and recombinantly expressed IgG fragments. HPLC analysis of fluorescently labelled glycans released from (A) serum IgG and recombinantly expressed (B) Fc, (C) CH2-H and (D) CH2 using PNGaseF. Peaks were assigned using MALDI-TOF MS and negative ion ESI-MS/MS. The '$*$' denotes a glycan with the composition $Neu5Ac_2Hex_5GlcNAc_4$ (probably $Neu5Ac_2Gal_2GlcNAc_2Man_3GlcNAc_2$) which was present in too low a quantity for fragmentation analysis. Glycans were separated using the $LudgerSepN_2$ A method (Chapter 8, Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

Fragmentation MS of glycans from IgG fragments

Compositions of the glycans in terms of hexose (Hex) and *N*-acetylhexose (HexNAc) residues were deduced from the masses measured by MALDI-TOF ($[M+Na]^+$ ions) and negative ion nano-ESI MS ($[M+H_2PO_4]^-$ and $[M-H]^-$ ions). These studies were done in collaboration with Prof. David Harvey (Glycobiology Institute, University of Oxford). After mobility separation, detailed structures were determined by negative ion collision-induced decomposition (CID) in the transfer cell of a Waters Synapt G2 ion mobility Q-TOF instrument [232]. These structures were confirmed, where possible, by comparison with *N*-glycans released from well-characterized glycoproteins such as ovalbumin and porcine thyroglobulin.

Representative spectra are shown in Figure 5.4 with glycans containing one core GlcNAc residue (released with EndoS) in the left-hand column and those with two GlcNAc residues (released with PNGaseF) in the right-hand column.

The di-*N*-acetylchitobiose core region of the PNGaseF-released glycans was characterised by the production of abundant $^{2,4}A_R$, B_R and $^{2,4}A_{R\text{-}1}$ ions (the subscript R in this discussion refers to the reducing terminal residue and R-1 refers to the penultimate residue). In Figure 5.4, R and R-1 correspond to 4 and 5 or to 5 and 6 depending on the length of the antennae). Where a core fucose residue was present at the 6-position (Fig. 5.4F, G, H, J), the $^{2,4}A_R$ ion appeared at 405 rather than 259 mass units below that of the molecular ion. For the EndoS-released glycans, only a $^{2,4}A_R$ ion was present.

The composition of the 6-antenna and, by difference, the 3-antenna, was defined by the mass of the D and D-18 ions that are formed by formal loss of the di-*N*-acetylchitobiose core and the 3-antenna, together, in most cases, with the $^{0,3}A_{R\text{-}2}$ and $^{0,4}A_{R\text{-}2}$ ions from the core branching mannose residue. The D and D-18 ions appear at *m/z* values of 526 and 508, respectively, in the spectra of the biantennary glycans without galactose residues (Fig 5.4A, F). These ions appear at *m/z* values of 688 and 670 in glycans with galactose on the 6-antenna (Fig 5.4C, H) and where one galactose residue was present, (Fig 5.4B, G) the presence of both sets of ions indicated the existence of isomers with galactose on either antenna. The $^{0,3}A$ cross-ring cleavage ion defined the compositions of the antennae with the C1 ions defining the terminal residues. The hybrid glycans (Fig 5.4D, I) gave D and

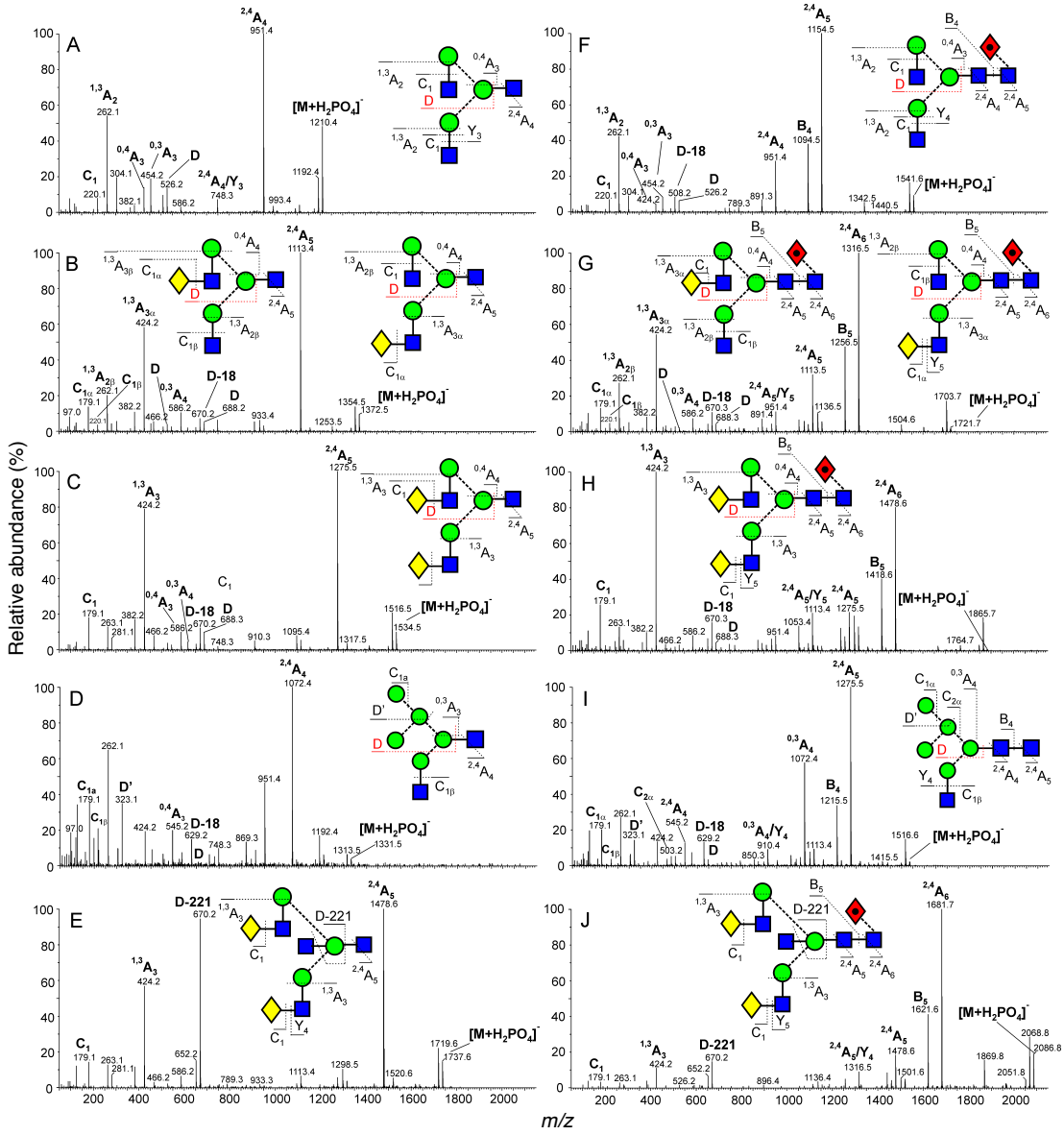


Figure 5.4: Fragmentation analysis of released glycans from IgG. Example fragmentation spectra are shown for EndoS released glycans (A, B, C, D, E) and PNGaseF released glycans (F, G, H, I, J) from serum IgG. Fragments are labelled following the Domon and Costello nomenclature [234]. The spectrum shown in D, contains a contribution from ions (particularly the ions at m/z 262, 424 and 951) derived from another adduct of the glycan of composition $Man_3GlcNAc_3$ appearing at the same mass. Only partial separation of the ions from these compounds was achieved by ion mobility. This figure was made by Prof. David Harvey (Glycobiology Institute, Department of Biochemistry).

D-18 ions at m/z values of 647 and 629 respectively with additional $^{0,3}A_3$ and $^{0,4}AA_3$ ions characterising the presence of the three mannose residues in the 6-antenna. The bisected glycans (Fig 5.4E, F) displayed prominent ions at an m/z of 670 which corresponds to D-221 (*i.e.* the loss of the bisecting GlcNAc residue from the D ion). For further information on the interpretation of these negative ion CID spectra see [232].

Susceptibility of IgG truncations to EndoS activity

Having confirmed that isolated domains of IgG1 could be expressed as soluble glycoproteins, it was of interest to then test their susceptibility to EndoS deglycosylation. These experiments were conducted with an aim of determining if any protein-protein interactions involving missing domains were crucial to EndoS glycoside hydrolase activity against IgG1. Figure 5.5 shows that glycans could be hydrolysed from the expressed IgG fragments in all cases.

In terms of human serum IgG, the higher molecular weight sialylated glycans appear less susceptible than the more abundant neutral glycans. Subsequent analysis detailed later in this chapter reveals that these glycans derive from the Fab domains of IgG (Section 5.3.1). The glycans of the Fc domain were all able to be released by EndoS, however, small populations of undigested bisected biantennary glycans could be seen by MALDI-TOF MS analysis upon treatment of the EndoS digested product with PNGaseF. The same observation was made for the CH2-H and CH2 glycoproteins where there was a population of uncleaved bisected biantennary glycans. The peaks representing either tetraanntenary or bisected-trianntenary glycans (in both galactosylated and agalactosylated forms) from CH2-H and CH2 also appear more resistant to EndoS cleavage than other glycan species present on these isolated domains (Fig. 5.5C, D).

5.1.2 The role of the EndoS C-terminus in glycoside-hydrolase activity

Considering the evidence for the presence of a CBM within EndoS, we sought to assess whether deletion of this region had an effect on EndoS activity. In addition, the effect of deleting the CT domain was investigated to obtain an overall understanding of the role of the C-terminal domains of EndoS in its activity against IgG.

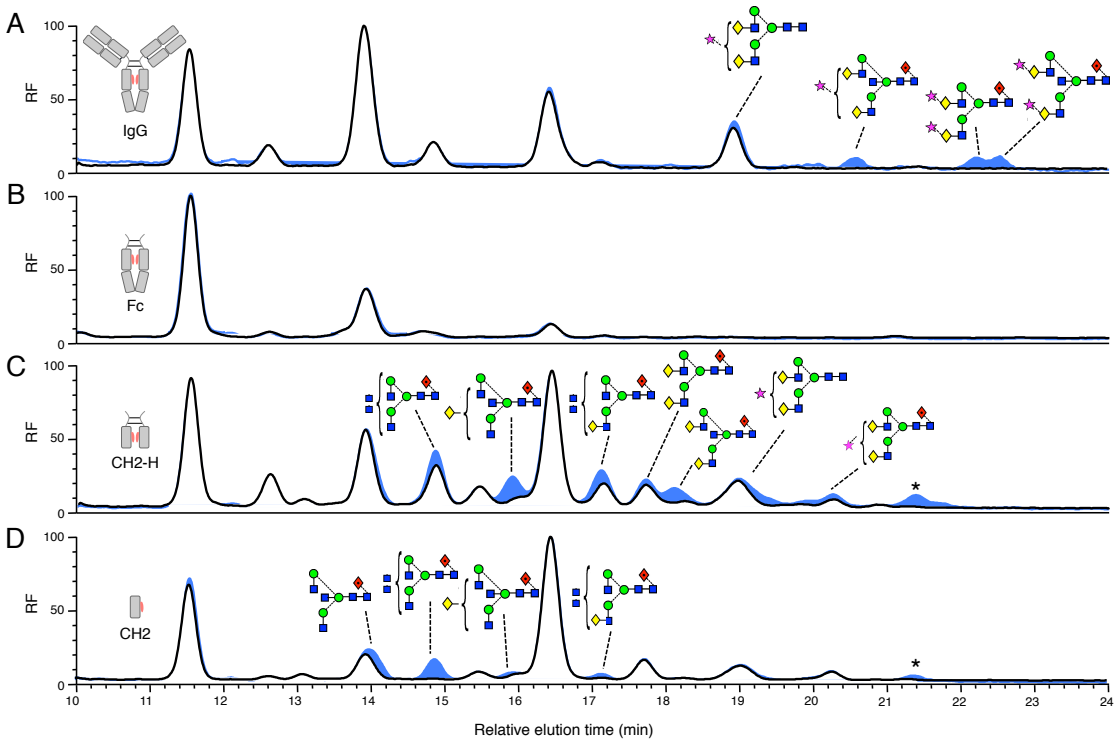


Figure 5.5: Analysis of EndoS-released glycans from serum IgG and IgG fragments. HPLC analysis of fluorescently labelled glycans released from (A) serum IgG and recombinantly expressed (B) Fc, (C) CH2-H and (D) CH2. Glycans were released using EndoS (black) or a sequential digest consisting of EndoS treatment followed by digestion with PNGaseF (blue). The peaks were assigned using negative ion ESI-MS/MS. Blue glycan peaks corresponding to glycans which remained after treatment with EndoS, which were only released as a result of sequential digest with PNGaseF are labelled using Oxford glycan nomenclature as described in Fig. 5.1. The '$*$' denotes a glycan with the composition $Neu5Ac_2Hex_5GlcNAc_4$ (probably $Neu5Ac_2Gal_2GlcNAc_2Man_3GlcNAc_2$) which was present in too low a quantity for fragmentation analysis. Glycans were separated using the $LudgerSepN_2$ A method (Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

Activity of EndoS fragments against IgG

The EndoS truncations, CBM-KO (a deletion of residues 761–924) and ChitLRR (residues 1–760), were incubated with human serum IgG to assess their ability to cleave the glycans from the IgG heavy chain. The CBM and CBMCT constructs were also tested to confirm that these domains had no isolated glycoside hydrolase activity. After incubation of the reaction at 37 °C, the components were analysed by reducing SDS-PAGE. A molecular weight shift of the band representative of the IgG heavy chain was used to judge the glycosylation state of the IgG.

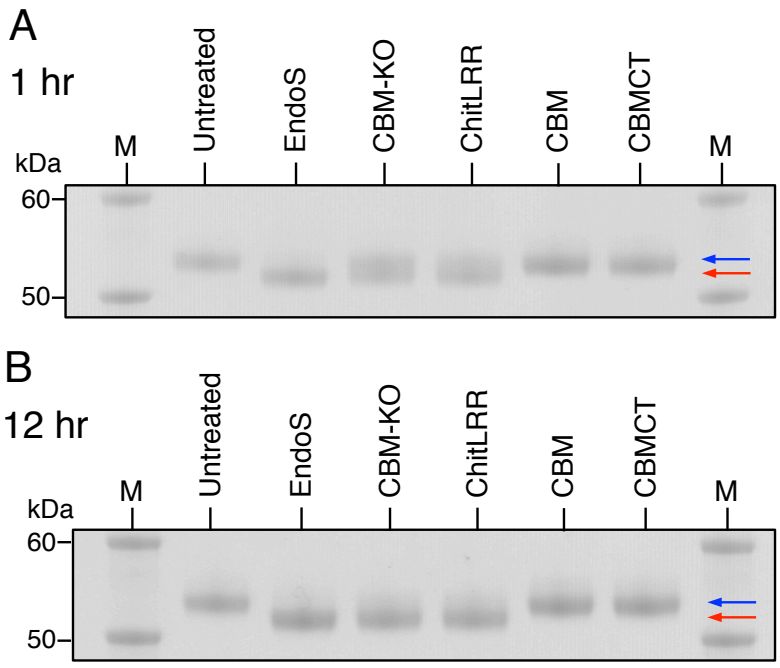


Figure 5.6: Deletion of the CBM or C-terminal region of EndoS affects glycoside hydrolase activity against serum IgG. Both panels show SDS-PAGE analyses of the reaction components from the enzymatic digest of serum IgG with either EndoS or the EndoS truncations. The constructs used for each reaction are denoted above each lane of the gel where M indicates molecular weight marker. SDS-PAGE analyses after (A) 1 hour or (B) 12 hours of incubation with EndoS at 37 °C are shown. The coloured arrows indicate the position of the upper (glycosylated, blue) and lower (deglycosylated, red) bands representative of the heavy chain of serum IgG.

After a 1 hour incubation, a reduction in molecular weight (approximately 3 kDa) of the IgG heavy chain band could be seen for the EndoS treated sample when compared with the untreated IgG sample. At this 1 hour time point, both the CBM-KO and ChitLRR-treated samples revealed the presence of two bands for the IgG heavy chain (Fig. 5.6A). Upon longer incubation, it was apparent that these truncated forms of EndoS were able to

produce an SDS-PAGE profile similar to full-length EndoS. This could be seen 12 hours after the digestion was initiated. Here, a single band for the IgG heavy chain was present for both CBM-KO and ChitLRR treated samples, which ran at a lower apparent molecular weight compared to untreated IgG sample (Fig. 5.6B).

Figure 5.7 shows glycan analysis of the upper band from the 1 hour reactions of IgG when treated with CBM-KO and ChitLRR. Both CBM-KO and ChitLRR left a residual population of bisected glycans at this time-point in addition to the di-sialylated Fab glycans, which remain uncleaved even by full-length EndoS after 24 hours (refer to Fig. 5.5A). When compared to CBM-KO, ChitLRR left an additional population of complex glycans uncleaved after 1 hour.

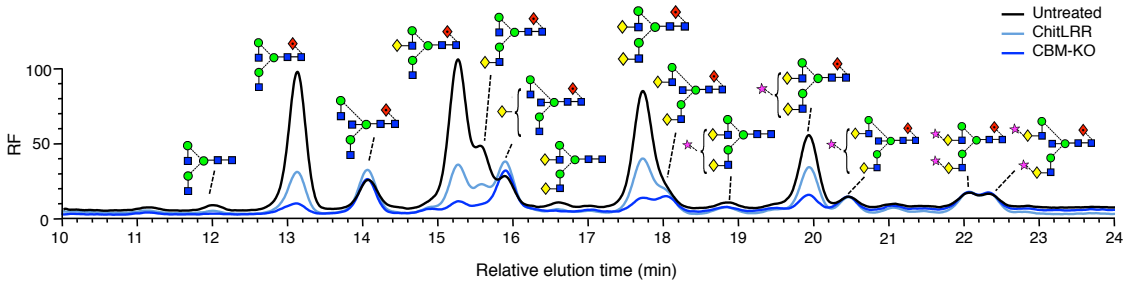


Figure 5.7: HPLC analysis of residual serum IgG glycan species present after a 1 hour incubation with either CBM-KO or ChitLRR. The HPLC chromatograms show fluorescently labelled glycans released from the upper gel bands corresponding to the glycosylated IgG heavy chain from Figure 5.6A. The gel bands were excised and PNGaseF digested for IgG from the lanes corresponding to the reactions with CBM-KO and ChitLRR. The chromatograms for glycans that remain after treatment with ChitLRR (light blue) and CBM-KO (dark blue) are shown in comparison to PNGaseF released glycans from untreated IgG (black). The chromatograms were normalised to the EndoS resistant Fab glycan peak at 22.4 minutes. Glycans were separated using the LudgerSepN_2 A method (Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

IgG binding capability of C-terminal fragments of EndoS

From the experiments described above, it was apparent that deletions to the C-terminal domain of EndoS were detrimental to its glycoside hydrolase activity. Considering this, an ELISA based assay was used to assess whether the C-terminal domains of EndoS were directly involved in IgG binding by assessing their ability to bind IgG1 Fc. Attempts were made to express and purify the CT region alone (residues 924–995) for use in these experiments. Despite successful expression of a soluble construct corresponding to this fragment of EndoS, the CT domain could not be purified to sufficient homogeneity for

further analyses and was not used in subsequent experiments.

An IgG1 Fc construct was expressed in HEK293 T cells with no C-terminal hexahistidine tag (FcnoH) for use in binding assays. The purification of this construct from the cell culture supernatant was achieved using a Protein A affinity column. A SEC step was then used to further purify the protein and to remove aggregates. As a negative control to be used in appreciating background binding levels to FcnoH, the globular bacterial protein MBP was cloned with a C-terminal hexahistidine tag. MBP was then expressed using *E. coli* cells and was purified using IMAC and SEC. The proteins CBM and CBMCT as well as the control MBP were then tested for their ability to bind FcnoH using an ELISA based approach.

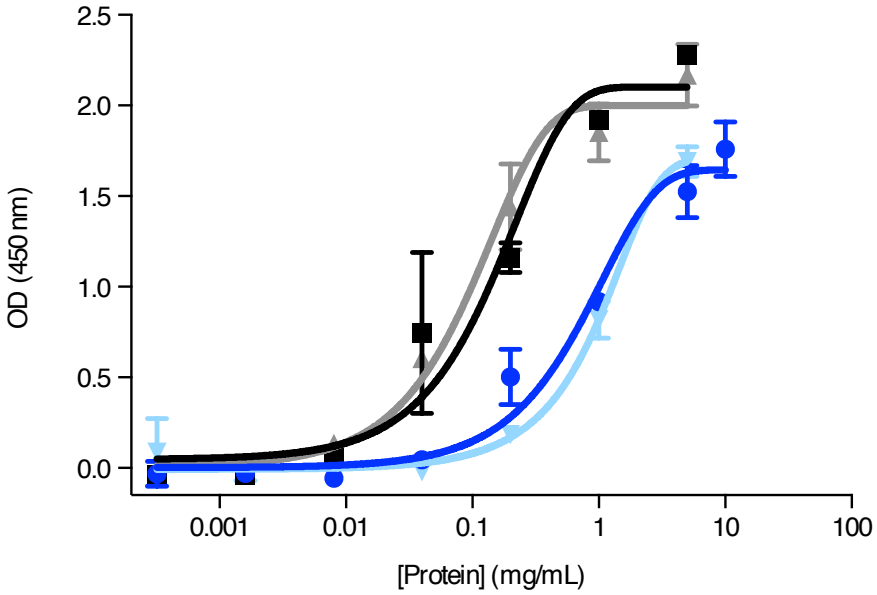


Figure 5.8: Assessing CBM and CBMCT binding to IgG1 Fc. Curves generated for CBM (blue), CBMCT (black), CBMCT C986S (grey) and MBP (light blue) when bound to recombinantly expressed IgG1 Fc (FcnoH). Bound protein is measured by the optical density (OD) at 450 nm observed by colorimetric assay of the detection agent HisProbe-HRP. Data are represented as the mean with the standard deviation and are fitted with a Boltzmann sigmoidal curve.

FcnoH was immobilised onto a 96-well plate which was then incubated with solutions of CBM, CBMCT and MBP. Unbound CBM, CBMCT and MBP were washed from the plate before bound protein was detected using HisProbe-HRP. HisProbe-HRP is a Ni-conjugated HRP which binds to histidine tags. Once bound, HisProbe-HRP can be detected using a standard colorimetric assay for HRP. The binding curves for CBM, CBMCT and MBP to FcnoH are shown in Figure 5.8. The CBM and MBP curves are not significantly different.

The CBMCT gives a curve indicative of higher affinity binding to Fc than is seen for either CBM or MBP, the same is true for CBMCT C986S. This latter construct was made to control for the effect of non-specific interactions between CBMCT monomers through potential dilsulphide bonds of the unpaired cysteine at position 986.

5.2 Investigation of IgG glycan specificity

In the recent paper by Baruah *et al.* [142], it was shown that the oligomannose glycoform $Man_5GlcNAc_2$ was more resistant to EndoS-mediated deglycosylation than the usual repertoire of complex biantennary glycans present on serum IgG. Although more resistant, experiments carried out as part of this thesis revealed that this glycoform was still susceptible to EndoS hydrolysis (see Chapter 3, Section 3.4). In an effort to find alternative glycoforms which were more resistant to EndoS for use in receptor-refocusing experiments, the glycan specificity of EndoS was explored.

To investigate the repertoire of glycans that EndoS was able to hydrolyse from IgG Fc, a number of Fc glycoforms were engineered using *in vivo* and *in vitro* methods. The oligomannose glycoform $Man_9GlcNAc_2$ was produced by treatment of Fc expressing HEK 293T cells with the α-mannosidase I inhibitor kifunensine [204]. Fc expressed in the presence of the α-mannosidase II inhibitor, swainsonine, resulted in a series of hybrid-type glycans [91, 204]. The GnTI deficient cell line HEK 293S was used to produce the $Man_5GlcNAc_2$ glycoform [82]. Finally, a highly sialylated glycoform of Fc was produced by successive incubation of neuraminidase-treated Fc with the transglycosidases B4GalT1 and ST6GalT1 as previously described [87]. HPLC analysis of the glycans present on these engineered glyco-variants is shown in Figure 5.9.

These glycoforms were treated with EndoS for 24 hours at 37 °C. The glycans released during these digests were analysed by HPLC. To assess whether any EndoS resistant glycans were left on any of the Fc glycovariants, the EndoS digested Fcs were subsequently treated with PNGaseF to remove residual glycans. The results of these digests are shown in Figure 5.10. In the case of Fc expressed in the presence of kifunensine, a proportion of the $Man_9GlcNAc_2$ glycoform remained attached to Fc after digestion with EndoS for 24 hours. The same was true for several of the hybrid-type glycans present on the Fc expressed in the

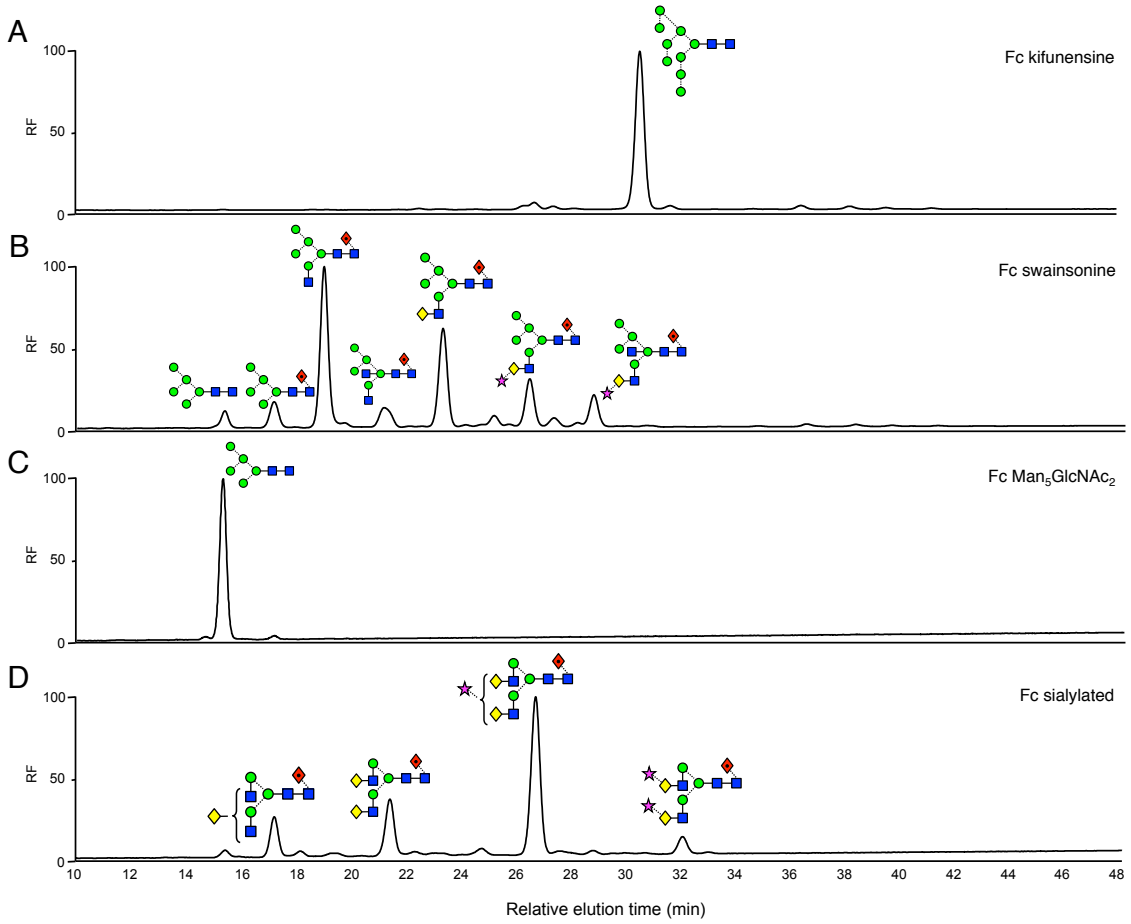


Figure 5.9: Glycan profiling of engineered Fc glycoforms. HPLC chromatograms are shown for fluorescently labelled glycans released with PNGaseF from the Fc glycoforms. The $Man_9GlcNAc_2$ and hybrid-type glycans are shown for Fc grown in the presence of (A) kifunensine and (B) swainsonine. (C) The predominantly $Man_5GlcNAc_2$ glycoform was produced by expression of Fc in GnT1 deficient HEK 293S cells. (D) A highly sialylated glycoform was produced by successive incubation of Fc with B4GalT1 and ST6GalT1. Glycans were separated using the XBridge method (Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

presence of swainsonine. These glycans were still hydrolysed by EndoS, but not as readily as the complex biantennary glycans usually found on IgG Fc. The $Man_5GlcNAc_2$ glycan present on Fc produced in HEK 293S cells was completely hydrolysed by EndoS in 24 hours. This was also the case for the highly sialylated glycoforms produced by treatment of Fc with transglycosidases. An EndoS resistant glycoform was not observed, but glycoforms that were more resistant to EndoS cleavage were identified, namely $Man_9GlcNAc_2$ and some hybrid-type glycans.

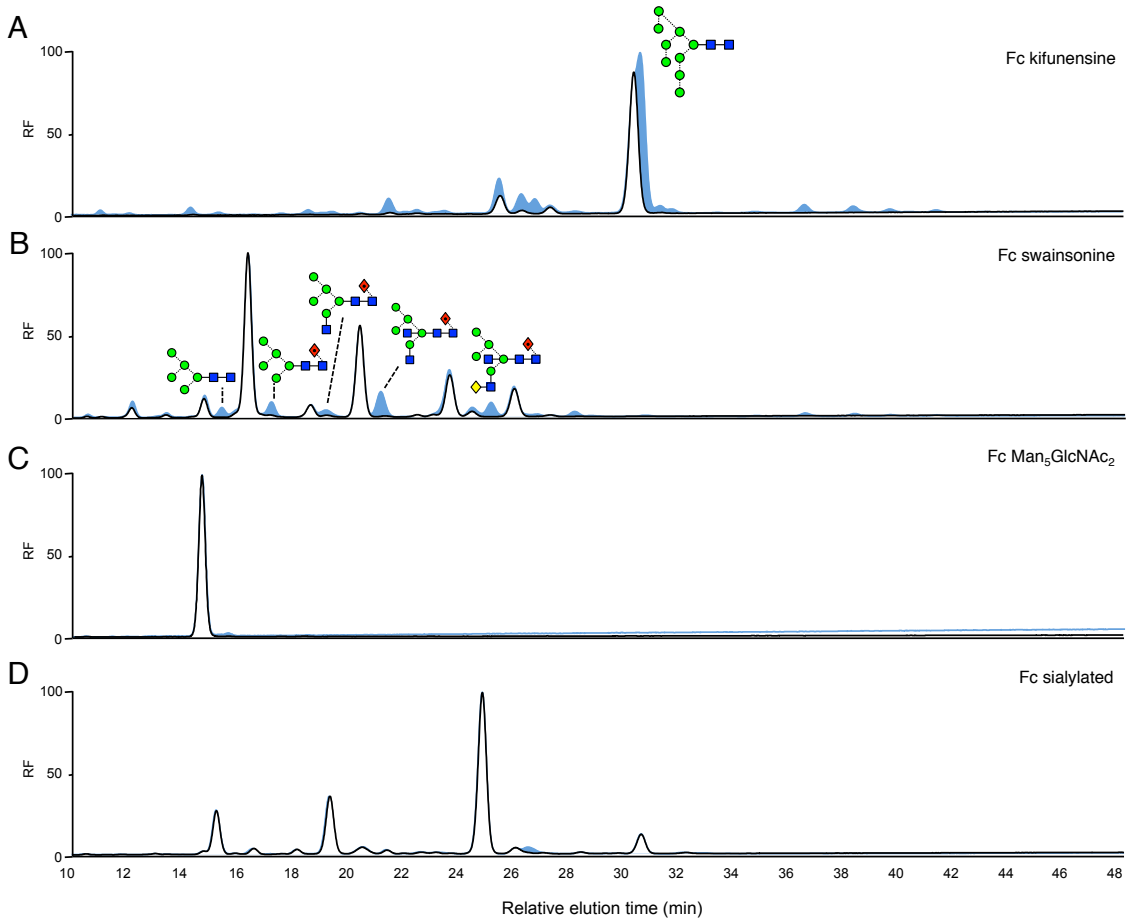


Figure 5.10: EndoS digestion of engineered Fc glycoforms. HPLC chromatograms are shown for fluorescently labelled glycans released with PNGaseF from the Fc glycoforms. The digested glycans are shown for Fc grown in the presence of (A) kifunensine and (B) swainsonine. (C) Digested glycans released from Fc expressed in GnT1 deficient HEK 293S cells and (D) a highly sialylated glycoform are also shown. Glycans were released using EndoS alone (black) or a sequential digest consisting of EndoS treatment followed by PNGaseF treatment (blue). Blue glycan peaks corresponding to glycans which remained after treatment with EndoS (which were only released as a result of the sequential digest with PNGaseF) are labelled using Oxford glycan nomenclature [83]. Glycans were separated using the XBridge method (Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

5.3 Investigation of alternative glycoprotein substrates

EndoS is able to cleave specifically the *N*-linked glycan from all four human IgG subclasses [145]. EndoS has also been shown to remove only the *N*-linked glycan from native but not denatured IgG [139]. The specificity of EndoS for IgG is further supported by a reported lack of EndoS activity against both IgA and IgM (unpublished data from [141]). Despite the widely reported specificity of EndoS for IgG, there is a lack of specific experimental evidence showing that EndoS does not cleave other glycoprotein substrates. Considering the potential therapeutic use of EndoS, characterising its specificity is crucial to ensure no off-target effects caused by deglycosylation of glycoproteins other than IgG. To this end, a series of experiments were carried out to investigate categorically the specificity of EndoS for IgG Fc glycans with reference to both IgG Fab domains and alternative glycoprotein substrates.

5.3.1 EndoS activity against IgG Fab glycans

As mentioned throughout this chapter, human serum IgG has a conserved *N*-linked glycan present at Asn297. In addition to Fc glycosylation, *N*-linked glycans can be found on approximately 15–20% of normal human IgG Fab domains [76, 77, 78]. These *N*-linked glycosylation sites arise due to the natural variation in sequences of the variable domains of IgG.

The specificity of EndoS for *N*-linked glycans of the Fc domain over the Fab domains of IgG has not been directly assessed. To ascertain whether EndoS showed activity for glycans present on the Fab domains of IgG, human serum IgG was digested into Fab and Fc fragments using immobilised papain. The digested fragments were subjected to purification via a Protein A column in an effort to separate both undigested IgG and Fc from the Fab fragments. To isolate the Fc from undigested IgG, the mixture was separated using SEC (Fig. 5.11).

The glycans of the purified Fab and Fc fragments were released using PNGaseF and were then analysed by HPLC (Fig. 5.12). The glycosylation profiles of the Fc and Fab fragments were comparable to those observed previously [77, 314]. The Fab glycans show a higher proportion of sialylated glycans and a higher proportion of glycans with a bisecting

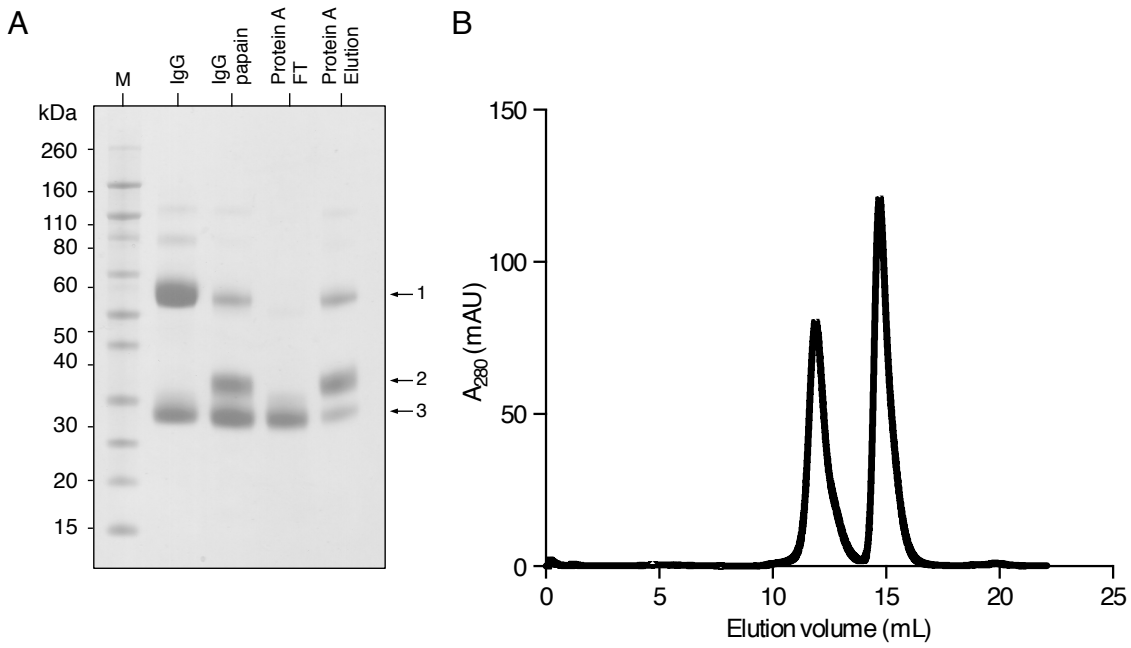


Figure 5.11: Preparation of Fab and Fc fragments from human serum IgG. (A) Reducing SDS-PAGE analysis of IgG and IgG fragments was carried out, the lanes show, 'IgG', intact serum IgG; 'IgG papain', serum IgG treated with immobilised papain for 18 hours at 37 °C; 'Protein A FT', the flow through of the Protein A column; 'Protein A Elution', the protein eluted from the Protein A column using 0.1 M citrate, pH 3.5. The right hand side of the gel is annotated to indicate (1) the IgG heavy chain, (2) the Fc monomer and (3) the band which represents both the IgG light chain and Fab fragments. (B) A SEC chromatogram is shown for the purification of Fc fragments from residual intact IgG. The Fc dimer eluted from the Supedex 200 10/300 GL column at approximately 15 mL. Only the leading edge of the peak was collected to avoid potential contamination with residual IgG light chain or Fab fragments.

GlcNAc than Fc. The level of neutral biantennary glycans was larger than expected for the Fab fragments, perhaps indicative of contamination of the Fab sample with residual IgG or Fc.

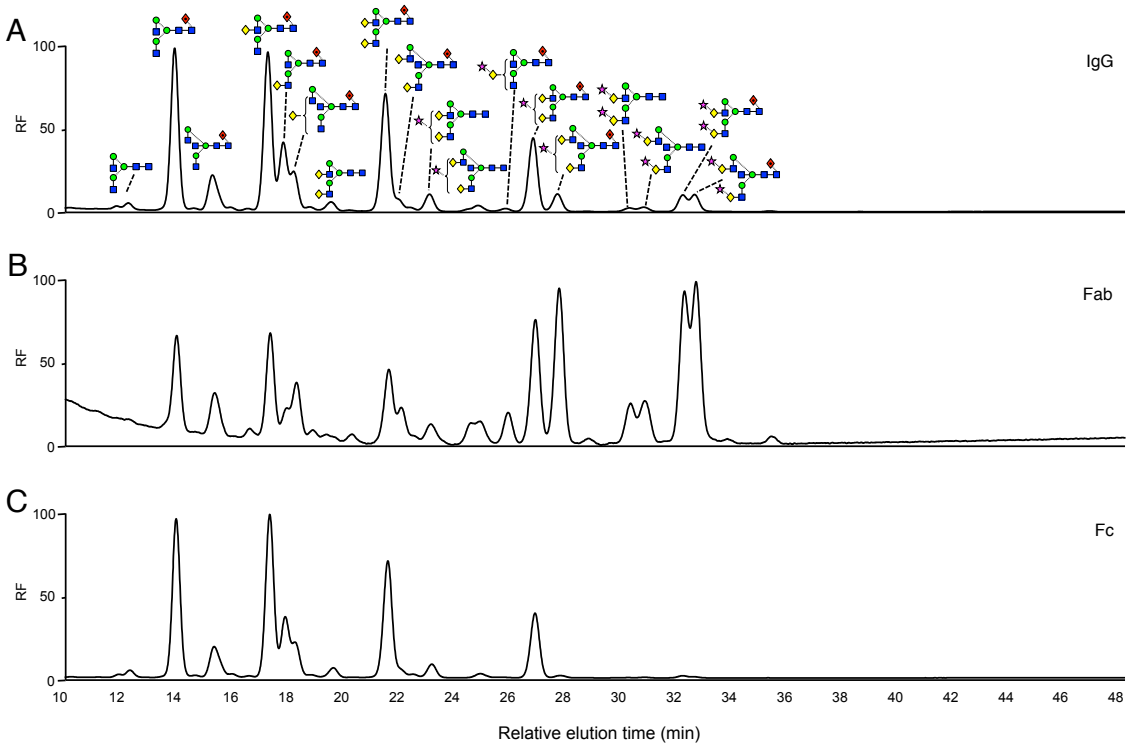


Figure 5.12: Glycan profiling of IgG Fab and Fc fragments from human serum IgG. HPLC chromatograms are shown for fluorescently labelled glycans released with PNGaseF from human serum IgG (A), IgG Fab fragments (B) and IgG Fc fragments (C). The glycans are labelled for the IgG chromatogram using Oxford glycan nomenclature [83] as described in Figure 5.1. Glycans were separated using the XBridge method (Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

To test the ability of EndoS to cleave glycans from the Fab fragment of human serum IgG, both the Fab and Fc fragments as well as intact serum IgG were incubated with EndoS for 24 hours at 37 °C. The released glycans were analysed by HPLC (Fig. 5.13, black chromatograms). Fab, Fc and IgG were then treated with PNGaseF to remove any residual glycans from the proteins that had not been hydrolysed by EndoS (Fig. 5.13, blue chromatograms).

The HPLC chromatograms in Figure 5.13 revealed that EndoS was able to release glycans from both Fab and Fc (Fig. 5.13, black chromatograms). In the context of IgG, a population of mono- and di-sialylated glycans with and without bisecting GlcNAc were resistant to EndoS hydrolysis in this assay (Fig. 5.13A). In the case of Fab, it was apparent that although some glycans were hydrolysed from this protein sample, the majority of the sia-

lylated glycans remained attached to the protein. The same was true for a population of the neutral glycans with bisecting GlcNAc (Fig. 5.13B). These uncleaved Fab glycans correspond well to those that remained uncleaved from full length serum IgG. No further glycans were released by PNGaseF treatment of EndoS-hydrolysed Fc, indicating that all glycans present on Fc were hydrolysed by EndoS (Fig. 5.13C).

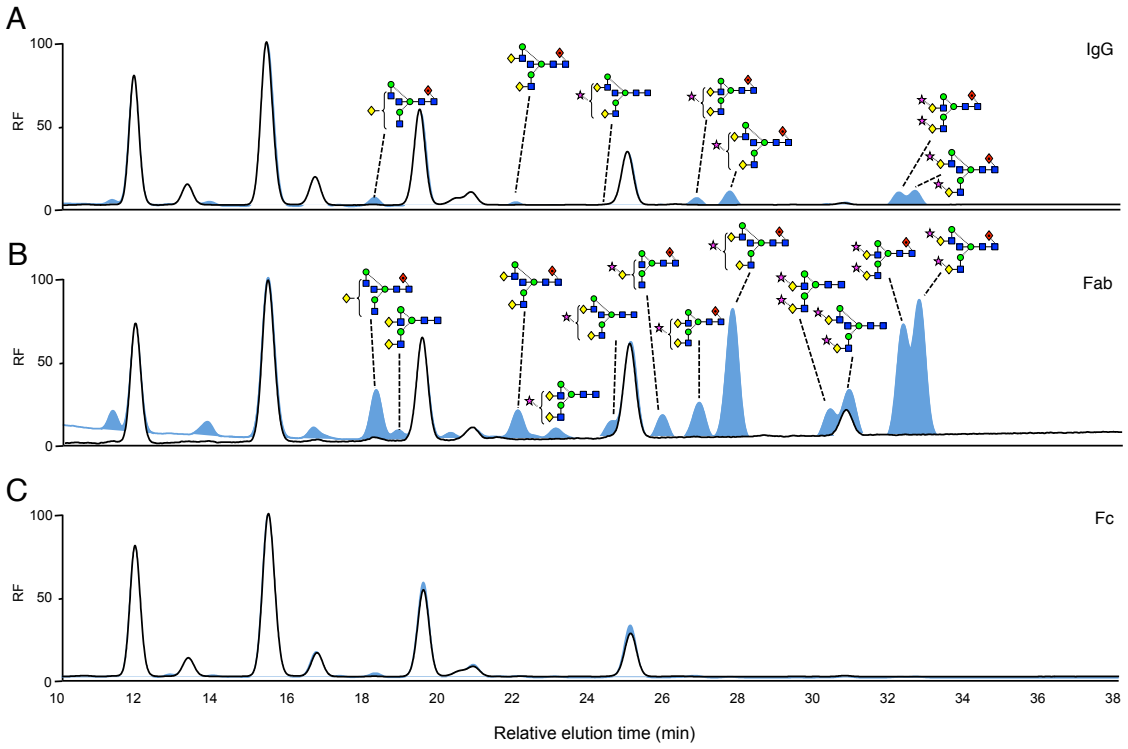


Figure 5.13: EndoS activity against Fab and Fc fragments from human serum IgG. HPLC chromatograms are shown for fluorescently labelled glycans released with EndoS (black chromatograms). Glycans which remained after EndoS treatment were released with a subsequent PNGaseF digest (blue chromatograms) and are shown on the same axes for each protein. Chromatograms are shown for (A) human serum IgG, (B) Fab fragments and (C) Fc fragments. The glycans which remained undigested by EndoS, which were released with subsequent PNGaseF treatment (blue peaks) are labelled following Oxford glycan nomenclature [83] as described in Figure 5.1. Glycans were separated using the XBridge method (Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

5.3.2 Human immunodeficiency virus gp120

To investigate EndoS activity against alternative glycoproteins, the large, heavily glycosylated HIV glycoprotein 120 (gp120) was used as a substrate. There are over 20 glycosylation sites on gp120 alone where the *N*-linked glycosylation sites are largely occupied by oligomannose type glycans [315]. Although EndoS is unlikely to encounter HIV gp120

as a natural substrate, the large and extensive glycan repertoire made it a useful tool for first assessing EndoS specificity.

The total glycan profile of HIV gp120 glycans released with PNGaseF is shown in Figure 5.14A. The typical series of oligomannose glycans dominate the profile with $Man_8GlcNAc_2$ and $Man_9GlcNAc_2$ glycoforms being the major species. Incubation of gp120 with EndoS reveals that EndoS is able to hydrolyse a subset of the glycans from HIV gp120 (Fig. 5.14B). The EndoS released glycans are dominated by glycoforms found on human serum IgG. The oligomannose glycans of HIV gp120 are untouched by EndoS as are the majority of other glycans found on this substrate. This is demonstrated by PNGaseF release of the population of HIV gp120 glycans which remain linked to the protein following the treatment with EndoS (Fig. 5.14C).

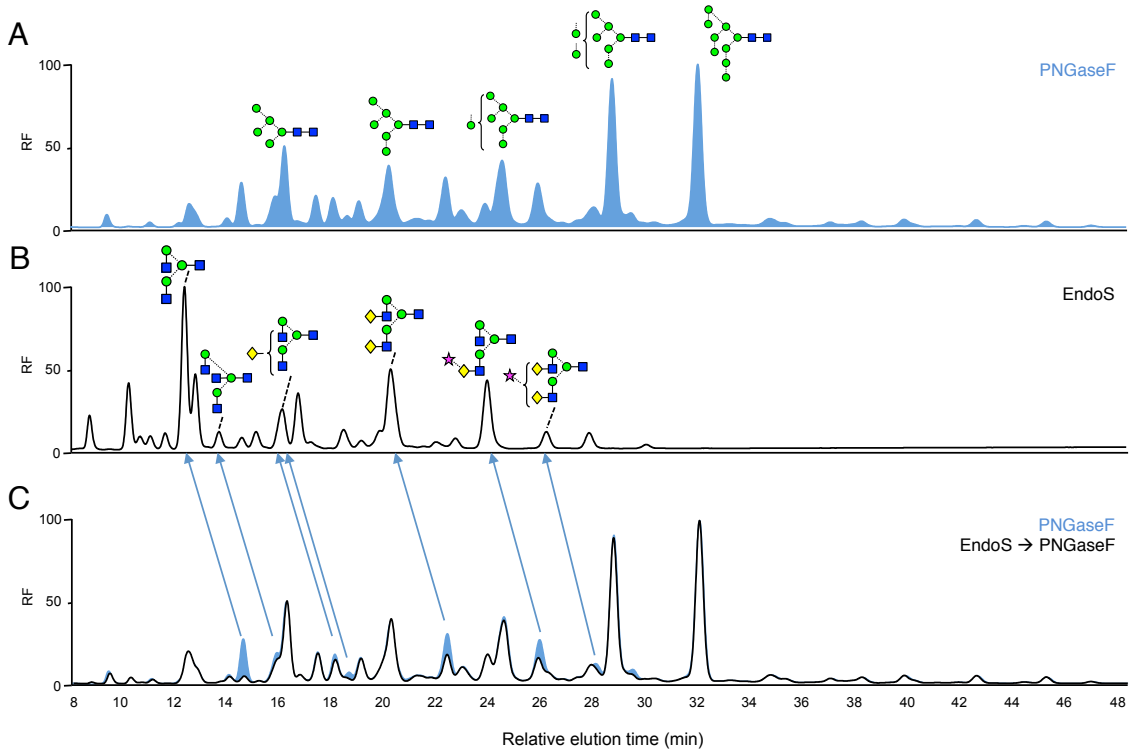


Figure 5.14: EndoS activity against HIV gp120 glycans. (A) An HPLC chromatogram is shown for the glycans released from HIV gp120 after PNGaseF digestion. Predominant glycans of the oligomannose series are labelled. (B) EndoS-released glycans from HIV gp120 were analysed by HPLC where the major glycan peaks are labelled. (C) To show the impact of EndoS deglycosylation on the total glycan pool of HIV gp120, the EndoS treated sample was subsequently treated with PNGaseF (black chromatogram). This chromatogram was overlaid on the chromatogram of total glycans released with PNGaseF (blue chromatogram). Blue arrows connect the visible blue peaks from panel C, with their digestion products in panel B. Glycans were separated using the XBridge method (Section 2.9.6) and are labelled according to Oxford glycan nomenclature [83] as described in Figure 5.1. Fluorescence is reported as relative fluorescence units (RF).

5.3.3 Serum glycoproteins

If EndoS is to be used intravenously as a treatment for autoimmune diseases, it would encounter not only IgG, but other serum glycoproteins as well. Having identified EndoS activity against HIV gp120, EndoS activity against five predominant serum glycoproteins was tested. This was done with an aim to explicitly explore EndoS specificity for IgG in the context of human serum glycoproteins and to look for potential off-target effects that could result from EndoS treatment. The abundant human serum glycoproteins transferrin, IgA, α_2-macroglobulin and α_1-antitrypsin were purchased as purified proteins to be tested for their susceptibility to EndoS. Figure 5.15 shows the structures of each of these proteins in an effort to describe both the size and shape of the proteins as well as the location of their N-linked glycosylation sites.

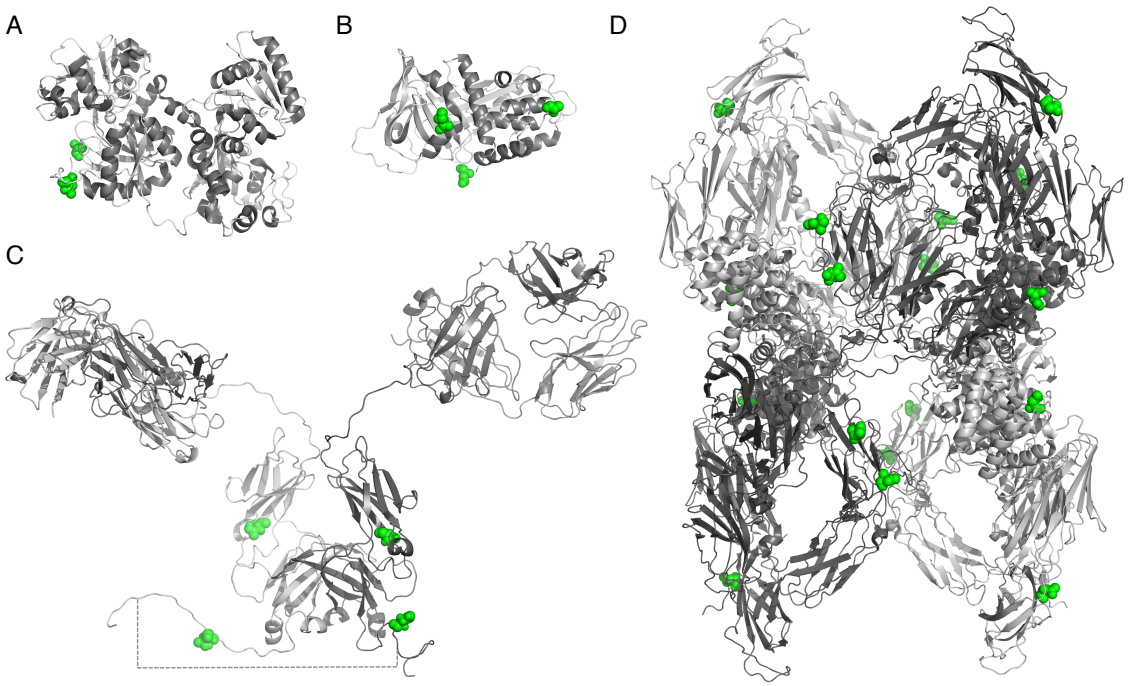


Figure 5.15: Protein structures of abundant serum glycoproteins. All structures show the protein component in cartoon representation with the asparagine residues of N-linked glycosylation sites shown as green spheres. The crystal structures are shown for (A) human transferrin (PDB ID: 2HAV) and (B) human α_1-antitrypsin (PDB ID: 3NE4). (C) The structure of IgA shown here is adapted from the model structure of dimeric secreted IgA (SIgA) from [316]. In human serum IgA, the position of the N-glycans is identical to SIgA, but an additional disulphide bond is formed that brings the C-terminal tails of each heavy chain closer together. The position of this bond is indicated by a dotted line. However, the C-terminal tails are shown in the conformation predicted for dimeric SIgA. (D) The homotetrameric crystal structure of human α_2-macroglobulin is shown with monomers A and C in dark grey and monomers B and D in light grey (PDB ID: 4ACQ).

Each of the proteins was incubated with EndoS to ascertain whether EndoS showed any glycoside-hydrolase activity against these alternative substrates. A preliminary analysis was carried out using an SDS-PAGE gel shift assay to look for changes in molecular weight induced by deglycosylation (Fig. 5.16). As expected, a detectable shift in the molecular weight of the IgG heavy chain can be seen upon treatment with EndoS. In addition, a visible shift in the molecular weight of the heavy chain of EndoS treated IgA was seen when compared to untreated IgA. For the EndoS treated IgA sample, there is also still a band at the molecular weight equivalent to that seen for untreated IgA protein. The remaining proteins, transferrin, α_2-macroglobulin and α_1-antitrypsin did not show any visible differences between the untreated and EndoS treated samples by SDS-PAGE analysis.

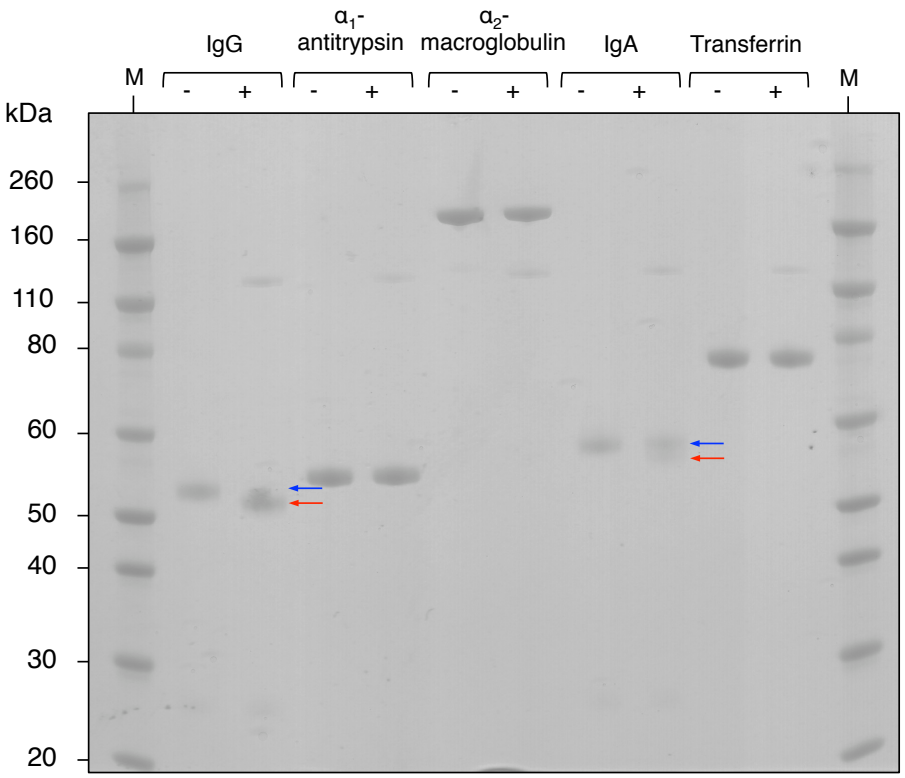


Figure 5.16: SDS-PAGE analysis of EndoS treated serum glycoproteins. The reducing SDS-PAGE gel shows (–) untreated samples and (+) EndoS treated samples in adjacent lanes for each serum glycoprotein. The coloured arrows indicate the position of the upper (glycosylated, blue) and lower (deglycosylated, red) bands representative of the heavy chain for both human serum IgG and IgA. 'M'; molecular weight marker.

Further analysis of glycans released from serum glycoproteins incubated with EndoS was carried out using HPLC (Fig. 5.17). Glycans were assigned based on relative elution time and comparison with both IgG glycan profiles and published glycosylation profiles

of the relevant serum glycoproteins [317, 318, 319, 320]. Further identification of these glycans was attempted using a Waters Synapt G1 ion mobility Q-TOF but glycans were present in too low a quantity to give sufficient signal to noise for formal identification and fragmentation analysis. Nevertheless, the chromatograms show that EndoS was able to hydrolyse glycans from all of the glycoproteins tested. However, less than 1% of the total glycan pools of transferrin and α_1-antitrypsin were hydrolysed by EndoS (as determined by integration of HPLC peaks). This is apparent from the similarity of the glycan profile of PNGaseF released glycans and the glycan profile of PNGaseF released glycans from the glycoproteins that had been pre-treated with EndoS (compare top and bottom panels of Fig. 5.17A and B). Notably, a significant pool of the IgA glycans were released by EndoS. Although the majority of glycans remain intact on IgA, a proportion of the di-sialylated complex biantennary glycans were hydrolysed. A smaller population of the mono-sialylated form was also digested by EndoS. EndoS also had an effect on the total glycan pool of α_2-macroglobulin where hydrolysis of the mono-galactosylated, mono-sialylated biantennary glycan was visible from the HPLC chromatograms.

5.3.4 EndoS activity against whole serum

Having shown that EndoS has the capacity to deglycosylate glycoproteins other than IgG, the effect of this non-specific activity was investigated in terms of whole human serum. Both IgG and whole human serum were incubated with EndoS for 1 hour at 37 °C. The HPLC profile of glycans released from serum was then compared to that of glycans released from IgG alone. To visualise the effect of using higher substrate to enzyme ratios (*i.e.* smaller amounts of EndoS), equivalent reactions were carried out using serial 10 fold dilutions of EndoS. The results of these assays were again analysed by HPLC and are shown in Figure 5.18.

In human serum, the concentration of IgG ranges from 10–15 mg mL^{-1} [26]. For these experiments, IgG was used at a concentration of 0.5 mg mL^{-1} and serum was diluted 20 fold with PBS to give an approximately equivalent IgG concentration. Using 0.05 mg mL^{-1} of EndoS (a 1:10 enzyme to substrate ratio), there were differences between the IgG and serum HPLC profiles (Fig. 5.18 Ai and Bi). Peaks were present in the serum profile that did not appear in the IgG profile. These peaks correspond to disialylated complex

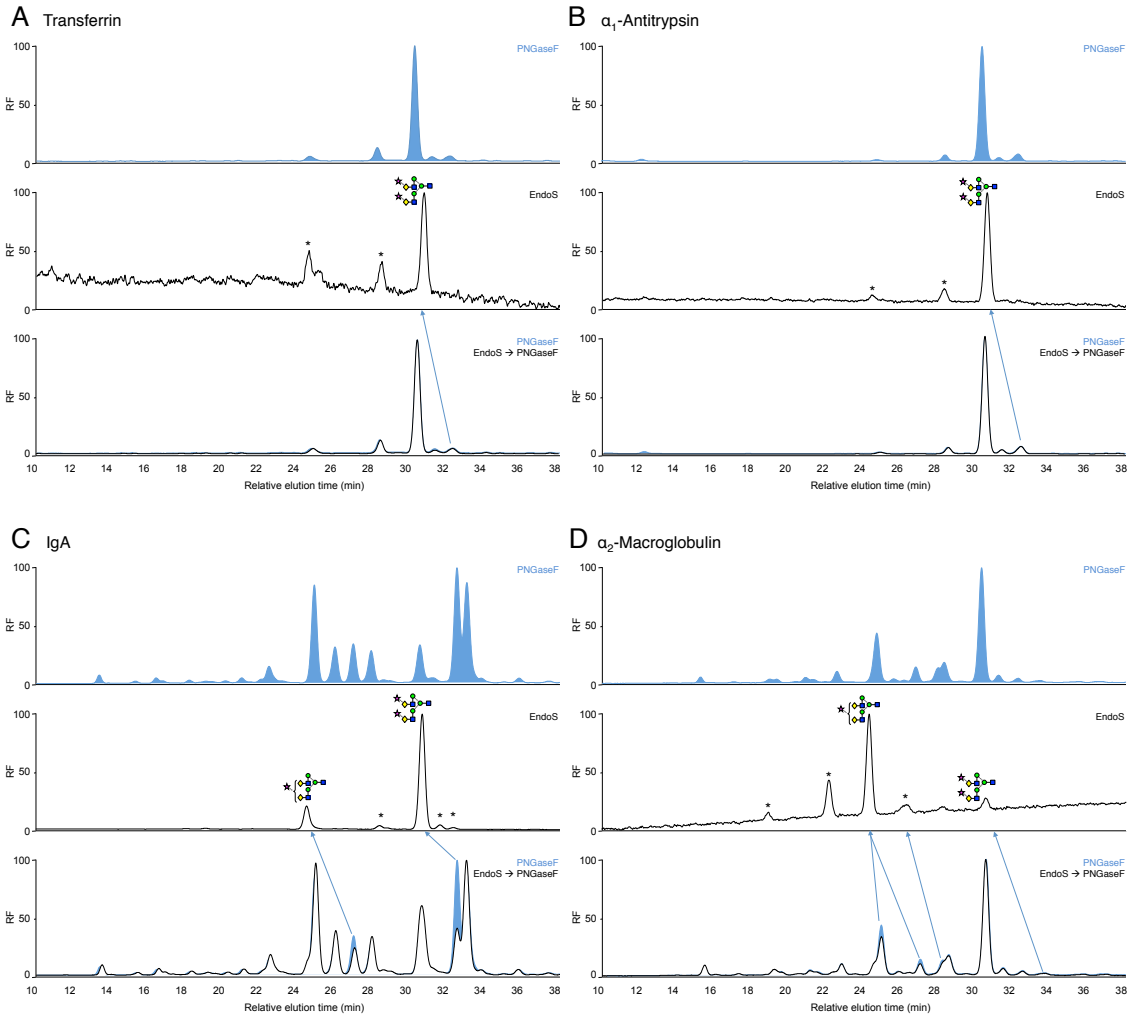


Figure 5.17: EndoS activity against serum glycoproteins. HPLC chromatograms used to assess EndoS activity against serum glycoproteins are shown for (A) transferrin, (B) α_1-antitrypsin, (C) IgA and (D) α_2-macroglobulin. For each protein, HPLC chromatograms are shown for the glycans released by (top) PNGaseF treatment and (middle) EndoS treatment. The final chromatogram (bottom) shows the impact of EndoS deglycosylation on the total glycan pool. Here, the EndoS treated sample was subsequently treated with PNGaseF (black chromatogram). This chromatogram was overlaid on the chromatogram of the total glycan pool released with PNGaseF (blue chromatogram). Blue arrows connect the peaks from panel iii, with their digestion products in panel ii. These 2AB labelled glycans were separated using the XBridge method (Section 2.9.6) and are labelled according to Oxford glycan nomenclature [83] as described in Figure 5.1. '*' indicates glycans that require further analysis in order to be assigned. Fluorescence is reported as relative fluorescence units (RF).

biantennary glycans. As the amount of EndoS used in the reaction was decreased, the appearance of these additional peaks in the serum profiles also decreased. In addition, peaks corresponding to released glycans with a bisecting GlcNAc reduced in size with reducing EndoS concentrations.

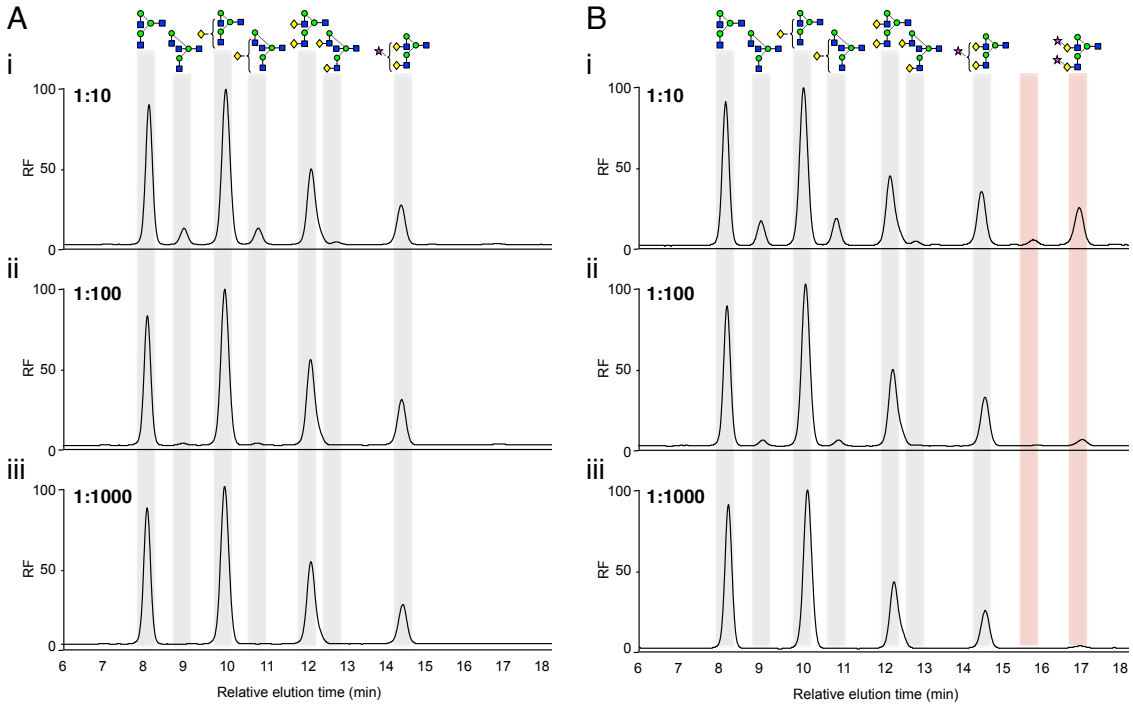


Figure 5.18: EndoS activity against whole human serum. HPLC chromatograms are shown for glycans released from (A) 0.5 mg mL^{-1} human serum IgG and (B) a 20 fold v/v dilution of whole human serum in PBS. For both IgG and serum, chromatograms are shown for reactions which used EndoS at a final concentration of (i) 0.05 mg mL^{-1}, (ii) 0.005 mg mL^{-1} and (iii) 0.0005 mg mL^{-1}. The calculated EndoS:IgG ratio is quoted in bold at the top left corner of each chromatogram. Glycan peaks that are unique to the serum chromatograms which are not seen in the IgG chromatograms are highlighted in red boxes. Glycans were separated using the LudgerSepN_2 B method (Section 2.9.6). Fluorescence is reported as relative fluorescence units (RF).

5.4 Discussion

In addition to its apparent contribution to immune evasion by *S. pyogenes*, EndoS is of interest as a tool for treatment of immunological diseases (reviewed in [180]) as a transglycosidase [143, 235] and as a tool for enhancing specific antibody-receptor interactions [142]. Thorough characterisation of therapeutic agents is a necessary and important step when preparing such agents for clinical use. This understanding is important in minimising unwanted off-target effects. Alternatively, discovering and understanding off-target

effects may offer new research avenues for further exploitation. Considering the potential therapeutic use of EndoS, experiments were carried out with an aim to obtain a more explicit understanding of EndoS specificity and its interaction with IgG.

In an effort to find the minimal IgG substrate unit from which EndoS was able to cleave glycans, a number of fragments of human IgG1 were expressed in mammalian cells. Fragments Fc, CH2-H and CH2 could all be expressed in the soluble fraction. All fragments were glycosylated although the levels of glycan processing were different between fragments. This suggested an effect of quaternary structure on glycan processing for these IgG fragments. In particular, the high proportion of digalactosylated glycans in the isolated CH2 fragment echoes that seen in previous analysis of truncated IgG domains of IgG expressed in Chinese hamster ovary cells [321]. The presence of hybrid glycan species in the CH2-H fragment and less so in the CH2 fragment was indicative that these protein structures were changing the accessibility of the *N*-linked glycans to Golgi α-mannosidase II.

Using these glycosylated IgG fragments, it was shown that EndoS could cleave the glycans from each of these fragments including the isolated CH2 domain from IgG. In all cases, EndoS could cleave the majority of the glycans from each of the constructs, which suggested that the Fab domains and the CH3 domains of IgG Fc were dispensable for EndoS activity against IgG. Looking at the CH2 domain in isolation, EndoS hydrolysis of glycans from this construct suggested that dimerisation of the IgG Fc region as well as the hinge of IgG were not required for EndoS activity. Despite this, it appeared that the CH3 region, the hinge and dimerisation may influence the activity of EndoS against IgG. Looking at CH2-H and CH2 deglycosylation, it was apparent that more residual glycans were left on the protein scaffold after a 24-hour incubation with EndoS when compared to Fc or serum IgG. However, this reduction in deglycosylation efficiency could be due to minor conformational changes or folding of the substrate due to expression of these domains in isolation. In addition, because the glycans were not of exactly the same composition on all expressed fragments, it was not possible to directly compare rates of deglycosylation. Furthermore, although glycans with bisecting GlcNAc could be cleaved by EndoS, it appeared that EndoS had reduced activity against these glycans compared to complex biantennary types. This was more apparent in the case of CH2-H and CH2 compared with Fc and

IgG and was supportive of the notion that the CH3, hinge and heavy chain dimerization may influence activity. EndoS was also less active against the bisected triantennary or tetratanntenary glycan types present on CH2-H and CH2.

In the context of *S. pyogenes* infection, IgG is likely to encounter a number of proteolytic and glycolytic enzymes including SpeB, IdeS and EndoS2 [322, 150]. The fact that EndoS appears to have activity on fragments of IgG, which do not contain either the hinge or Fab region, suggests that EndoS could continue to act on fragments produced by IdeS and SpeB cleavage. EndoS ability to act on these fragments may further compound the immune evasion strategy of *S. pyogenes*.

To assess the possible effects of the C-terminal domains on EndoS activity, truncations were designed where the CBM was removed (CBM-KO) and where both the CBM and CT regions were truncated (ChitLRR). As shown in Chapter 3, these constructs could both be expressed and purified as soluble proteins. Upon deletion of either the CBM or the CBM and CT regions, EndoS exhibited a difference in activity against serum IgG when compared with full-length EndoS. Using SDS-PAGE to assess serum IgG digested with these constructs, it was found that where EndoS could produce a single band representative of the deglycosylated heavy chain of IgG, CBM-KO and ChitLRR produced two bands, indicative of both deglycosylated and glycosylated forms of IgG heavy chain present in the reaction mixtures. After further incubation, it was apparent that both CBM-KO and ChitLRR could deglycosylate IgG to a comparable level to that achieved by EndoS. HPLC analysis of the glycans which remained on IgG after a 1 hour incubation with these constructs revealed residual populations comprising primarily bisected glycans. This observation perhaps indicates a role of the C-terminal region in IgG avidity, which allows these potentially more resilient glycoforms to be cleaved more effectively.

Further investigation into the role of the CBM and CT regions looked at whether these domains in isolation were able to bind to the Fc domain of IgG1. Sigmoidal binding curves were produced for both CBM and CBMCT when tested for binding against Fc. However, CBM bound with similar affinity to the negative control protein MBP. This suggests that if the CBM is able to bind specifically to Fc, the binding affinity of this region alone is likely to be poor. This is consistent with the potential low affinity interaction of CBM with monosaccharide subunits of the glycan at Asn297 as discussed in Chapter 4. In contrast to

the CBM construct, the CBMCT protein produced a sigmoidal binding curve indicative of a higher binding affinity to Fc than CBM and MBP. The same was true for the CBMCT C986S mutant, which was included to control for non-specific interactions via the unpaired cysteine residue in the CT domain. These data suggest that the CT region of EndoS is able to interact directly with the Fc domains of IgG. This is consistent with observations by Trastoy *et al.* [149] where no detectable binding of EndoS to IgG1 Fc was observed via surface plasmon resonance when a construct lacking the CBM and CT domains (residues 765–995) was tested.

The helical structure of the CT domain of EndoS has previously been shown to have structural homology to the substrate binding domain (SBD) of Hsp70 (PDB ID: 4JNE) [149]. The SBD of Hsp70 binds extended hydrophobic polypeptides which commonly occur in unfolded proteins [323]. The hinge region of IgG characteristically lacks secondary structure and contains hydrophobic stretches of proline residues and may be a potential binding site for CT. Interactions between CT and the hinge could help to localise EndoS to the CH2 domains of IgG Fc, facilitating cleavage of the *N*-linked glycan at position Asn297. An interaction of EndoS with the hinge of IgG may also account for its varying binding affinity against different IgG subclasses [149]. The sequence and length of the hinge region is highly variable between IgG subclasses and may partially dictate differing EndoS affinity.

Overall, the results of these experiments suggest that the C-terminal region of EndoS contributes to its activity. This contribution could conceivably occur via potential carbohydrate interactions with the CBM (as discussed in Chapter 4) and through additional interactions between the hinge region of IgG and the CT domain of EndoS. The differences seen between the specificity and activity of CBM-KO and ChitLRR against serum IgG glycans (which lack the CBM and CBMCT regions, respectively) are supportive of the additive contribution of both CT and CBM to the activity of EndoS. It is possible that deletion of the CBM domain in CBM-KO could disrupt its function due to a change in overall protein architecture rather than a loss of CBM function. However, the recently published crystal structure of EndoS shows that in the full-length structure, the N- and C- termini of the CBM domain are positioned close to each other [149]. Upon excision of the CBM domain, the protein backbone approaching from the upstream and downstream

N- and C-terminal fragments would likely still align to approximately the same location, causing minimal disruption to the overall architecture of the protein. In addition, the expressed CBM-KO and ChitLRR were both soluble and still exhibited glycoside hydrolase activity, suggesting that both constructs were correctly folded and that the effects seen are in fact due to the CBM and CT region of EndoS having a role in activity against IgG.

At least 30 different types of complex biantennary glycan are associated with human serum IgG [98]. In an effort to understand EndoS glycan specificity and in an attempt to find Fc glycoforms which were resistant to EndoS hydrolysis for use in receptor-refocusing experiments, EndoS was tested against a number of engineered Fc glycoforms which fall outside of this normal pool of IgG glycans. A number of Fc glycoforms were successfully engineered for this purpose. EndoS was able to hydrolyse the glycans of all of the engineered glycoforms tested, however, some glycans were more resistant to EndoS hydrolysis than others. In these cases, only a proportion of the total glycan population was removed by EndoS. The $Man_9GlcNAc_2$ glycoform, made by treating Fc-expressing HEK 293T cells with the α-mannosidase I inhibitor kifunensine, was the most resistant to EndoS. After a 24 hour incubation with EndoS at a 1:10 enzyme to substrate ratio, over half of the $Man_9GlcNAc_2$ Fc population remained glycosylated. Considering that the oligomannose glycoform $Man_5GlcNAc_2$ was essentially completely hydrolysed by EndoS under these conditions, it is likely to be the size and the extensive branching of the $Man_9GlcNAc_2$ that hinders EndoS hydrolysis as opposed to its oligomannose composition.

When serum IgG was treated with EndoS, a population of di-sialylated glycans remained intact on IgG. As is discussed later in this section, these glycans are associated with the Fab domains of IgG and cannot be hydrolysed as a result of their location rather than their structure. To specifically investigate whether EndoS could, in fact, hydrolyse these large, charged glycoforms from Fc, a highly sialylated population of Fcs was engineered using a combination of transglycosidases as described previously [88, 70]. The resultant glycan population predominantly consisted of a mono-sialylated glycoform but a small population of di-sialylated glycoforms were present. EndoS was able to successfully hydrolyse this di-sialylated glycan from the Fc scaffold suggesting that these glycoforms remain uncleaved in the context of serum IgG due to their localisation to the Fab domains rather than their composition.

It has previously been reported that glycans containing a bisecting GlcNAc residue are resistant to cleavage by EndoS [76]. It was shown here that bisected glycans are able to be digested by EndoS and are not resistant to hydrolysis. However, there are minor residual populations of uncleaved bisected glycans after digestion with EndoS as seen by both nano-ESI-MS and by HPLC analysis. Similarly, when C-terminal deletions were made to EndoS, these truncated forms of EndoS primarily left bisected glycan species intact. The presence of a bisecting GlcNAc residue is likely to be causing some steric hinderance to EndoS as it attempts to deglycosylate IgG Fc, but does not completely inhibit the catalytic activity of the enzyme.

Overall, it appears that the oligomannose glycoform $Man_9GlcNAc_2$ is the most resistant to EndoS hydrolysis out of those tested. Bisecting GlcNAc residues offer some resistance, but are still hydrolysed, contrary to previously published results. Essentially, as glycoforms deviate further from typical IgG Fc glycan structures, the more resistant they appear to be to EndoS hydrolysis. This correlation suggests that EndoS has evolved some steric specificity to IgG glycans and that avidity effects from protein-glycan interactions, through both the catalytic site and potentially the CBM, are contributing to the specificity of the deglycosylation reaction.

Experiments which looked explicitly at the ability of EndoS to hydrolyse glycans from the Fab domains of serum IgG showed that EndoS was able to cleave a population of glycans from Fab fragments. However, these may be due to some contamination of the Fab sample with Fc or IgG. Before conclusions are made about the ability of EndoS to cleave Fab glycans, additional purification steps of the Fab domain are necessary and a more stringent check for Fc or IgG contamination needs to be included to assess Fab purity before EndoS assays are carried out. Nevertheless, the majority of *N*-linked glycans on the Fab domains remained intact when treated with EndoS. In particular, populations of sialylated and bisected glycans were largely untouched. In contrast, the total glycan pool of IgG Fc was hydrolysed by EndoS. In context of full-length IgG, it is apparent that all Fc glycans can be successfully cleaved, whereas glycans associated with Fab are likely to be left intact.

Putting this information together, it was of interest how EndoS maintains an apparent specificity for IgG given its ability to cleave glycans from such a reduced protein scaffold

as the CH2 domain and its ability to hydrolyse varying glycoforms not commonly found on IgG Fc. It is possible that the multi-domain architecture of EndoS may sterically hinder it from cleaving other Ig substrates or it may be that the subtle differences in the CH2 region of IgG alone are responsible for distinguishing this substrate of EndoS from other Ig classes. However, to formally investigate this, a number of alternative glycoprotein substrates were tested for their susceptibility to EndoS.

HIV gp120 was used as an initial test protein for EndoS specificity. When incubated with EndoS, a population of glycans was cleaved from gp120 illustrating the first example of EndoS having the ability to deglycosylate an alternative glycoprotein substrate. The glycans which were hydrolysed from gp120 corresponded with glycan types commonly found, but not exclusive to those, on serum IgG. This was again suggestive that EndoS has some determinant of glycan specificity but importantly shows that EndoS has the capacity to deglycosylate other glycoproteins.

To further investigate the ability of EndoS to cleave alternative glycoprotein substrates, a panel of serum glycoproteins was tested. These were chosen with the rationale that serum glycoproteins will be encountered by EndoS if it is used intravenously as a therapeutic tool. The results showed that EndoS causes negligable deglycosylation of transferrin, α_1-antitrypsin and α_2-macroglobulin, although in each case, there was still some loss of glycan species from each protein. Importantly, a significant population of IgA glycans were hydrolysed by EndoS. This was visible as both a shift of the IgA heavy chain by SDS-PAGE gel analysis and through HPLC analysis of EndoS released glycans.

Much like IgG, monomeric IgA has two Fab domains and one Fc domain separated by an unstructured linker. Although no crystal structure exists for full-length monomeric serum IgA, a solution state model reveals its similar overall architecture to IgG [317]. As demonstrated by Mattu *et al.* [317], IgA is more heavily glycosylated than IgG with two conserved *N*-linked glycosylation sites and several potential *O*-linked glycosylation sites in the hinge region. One *N*-glycosylation site is present on the unstructured C-terminal tailpiece of the IgA heavy chain with the other on the IgA Cα2 (homologous to the IgG CH2 domains). The latter differs from the Asn297 glycan of IgG in that the glycan residue sits at the base of the Cα2 domain forming contacts with the Cα2 and Cα3 domains [324]. To further illustrate the location of the Fc glycans, the structure of the IgA Fc domain

in complex with its receptor, FcαRI, is shown in Figure 5.19. No direct interactions of the *N*-linked glycan from the Cα2 of IgA were revealed by the crystal structure of the Fc:FcαRI complex, suggesting that a loss of this glycan is unlikely to directly affect Fc:FcαRI interactions. EndoS is not as effective at deglycosylating IgA as it is IgG and is unlikely to disrupt FcαRI binding interactions, but importantly, EndoS still has the potential to interrupt IgA functions which rely on glycosylation.

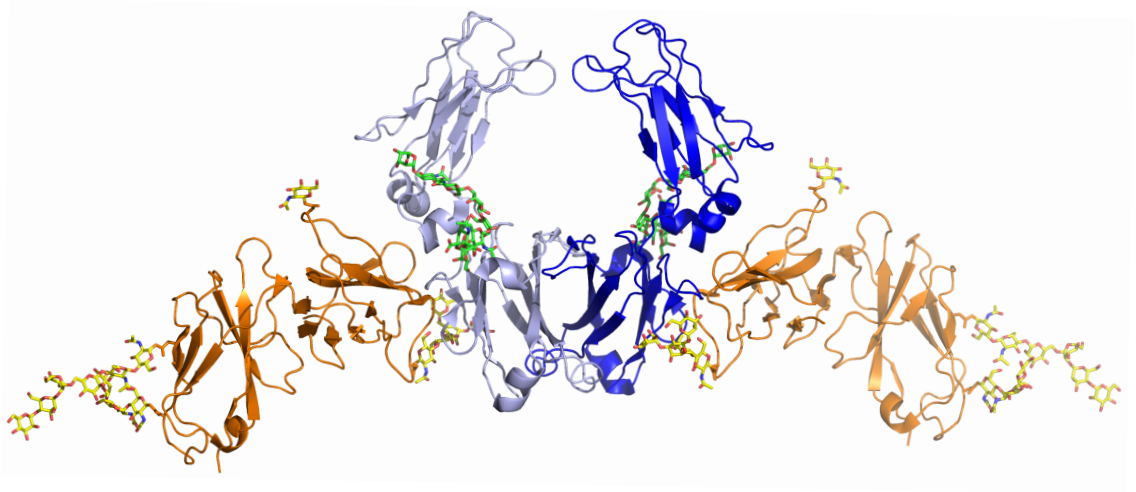


Figure 5.19: The crystal structure of IgA Fc in complex with FcαRI. The Fc is shown with the one monomer in blue and the other in light blue. The *N*-linked glycans at position Asn263 are shown in stick representation in green with oxygen atoms coloured red and nitrogen atoms coloured blue. The two FcαRI molecules are shown in orange with the four *N*-linked glycans on each receptor molecule shown in yellow in stick representation. The oxygen atoms of the glycans are coloured red and the nitrogens coloured blue (PDB ID: 1OW0, [324]).

Throughout these experiments, a 1:10 molar ratio of EndoS to substrate was used. This is a comparatively large amount of EndoS, compared to *in vivo* experiments which used as little as 500 μg to successfully deglycosylate the entire IgG pool of a rabbit within 4 hours [183]. This equates to an EndoS:IgG ratio of approximately 1:6000 assuming a serum IgG concentration of 10 mg mL^{-1} and a blood volume of between 160–480 mL [325]. A larger amount of EndoS was used in the experiments described here to ensure that all potential EndoS activity was observed, allowing for a more thorough characterisation of the enzyme. To look at the non-specific activity of EndoS in the context of whole serum, HPLC profiles of glycans released from IgG alone were compared to profiles of glycans released from whole serum. Additional peaks present for serum that are not present for IgG suggested that EndoS has non-specific effects in the context of serum. These additional peaks correlate to some of those observed when the individual serum glycoprotein substrates were

deglycosylated by EndoS. Despite this effect, as the amount of EndoS was decreased to give a calculated EndoS:IgG ratio of 1:1000, these additional glycan peaks were no longer detected for serum, and the profiles from IgG and from serum look almost identical. The same was true for the peaks corresponding to glycans with a bisecting GlcNAc residue. This result is supportive of the notion that EndoS is preferentially cleaving glycans from IgG, but has the capacity to cleave glycans from alternative glycoprotein substrates. The preference of EndoS for IgG is likely contributed to by both specific protein-protein and protein-glycan interactions as well as the high concentration of IgG in the serum. It is important to appreciate that EndoS has the potential to deglycosylate other proteins, but given the small amount needed to successfully deglycosylate the IgG pool, it may be that these potential off-target effects do not cause unwanted side-effects in a clinical setting. It must be noted however, that this method is only likely to detect the effect of EndoS on common serum glycoproteins, and that any glycoproteins with similar glycosylation profiles to IgG may not be detected. To circumvent these caveats, further investigation of EndoS specificity in the context of serum could be carried out using IgG depleted serum.

Overall, the experiments conducted here have given insight into the function and specificity of EndoS against IgG. Subunits of EndoS and IgG were identified that are important for EndoS specificity. In particular, the C-terminal domains of EndoS were implicated in conferring IgG-specificity through potential protein-carbohydrate and protein-protein interactions. Glycans more resistant to cleavage by EndoS have been identified, however no glycans were found which were completely EndoS resistant. Finally, EndoS was shown to cleave alternative glycoprotein substrates both in isolation and in the context of whole serum.

6

Structural effects of EndoS deglycosylation on IgG Fc

EndoS is a glycoside hydrolase produced by *S. pyogenes* which hydrolyses the *N*-linked glycan from the asparagine at position 297 of IgG Fc [141]. This *N*-linked glycan is crucial for the structural integrity and function of Fc (as reviewed in [79, 27]). In particular, deglycosylation of Fc by EndoS causes a significant reduction in Fc binding affinity to cellular FcγRs [145]. EndoS-mediated changes in FcγR binding are likely to be attributed to changes in the tertiary structure of the CH2 domain, with changes in quaternary structure of the Fc also playing a role. In this chapter, the impact of EndoS-mediated deglycosylation on Fc structure is discussed with reference to these two factors in an effort to determine which, if either, is contributing to the mechanism of Fc deactivation by EndoS. Particular attention is given to the murine IgG2a subclass, which, in contrast to murine IgG1 and IgG2b, maintains FcγR binding capability when deglycosylated with

EndoS [144, 182].

6.1 Assessing the effects of deglycosylation using SAXS

To understand the effects of EndoS-mediated deglycosylation on the structure of Fc, SAXS was used to compare the solution state behaviour of Fc and Fc which had been enzmatically deglycosylated with EndoS (Fc^{GlcNAc}). SAXS can be used to understand the overall shape and size of a protein in solution and can be further interpreted to give insight into to both the conformation and flexibility of the protein (for a recent review see [212]). SAXS has previously been used to explore the role of glycosylation on human IgG Fc [41]. Here, the authors noted an increase in the radius of gyration (R_g) for an aglycosylated Fc when compared to the glycosylated form. From this, the authors stipulated that aglycosylated Fcs are more open with respect to the CH2-CH2 domain distance than their glycosylated counterparts. This is contrary to the 'closed' CH2 conformations observed in the crystal structures of de- or aglycosylated Fcs [142, 41]. The results of the SAXS experiments generated here are discussed with respect to these previous studies.

SAXS requires protein in solution which is both homogeneous and monodisperse. To obtain high purity samples of Fc, recombinant human IgG1 Fc was produced in mammalian cell culture as previously described [142]. A construct encompassing the Fc region including the disulphide linked hinge (residues 224–446) from the heavy chain of human IgG1 was transiently expressed in HEK 293T cells using the mammalian expression vector pHLSec. Fc was expressed with a C-terminal hexahistidine tag and was then purified by IMAC directly from the cell culture supernatant. Finally, a SEC step was used to further purify the protein prior to SAXS experiments (Fig. 6.1A).

For the Fc^{GlcNAc} sample, purified Fc was treated with EndoS for 16 hours at 37 °C. Fc^{GlcNAc} was then purified from the reaction with a SEC step to remove residual EndoS (Fig. 6.1B). Both Fc and Fc^{GlcNAc} gave single symmetrical Gaussian peaks when eluted during the SEC step, indicative of pure monodisperse protein samples. Next, both Fc and Fc^{GlcNAc} were concentrated to 5 mg mL^{-1} for SAXS experiments. SAXS data were collected at the ESRF, BM29 where all measurements were recorded at 20 °C. Using the concentrated samples, a 2 fold dilution series of each protein was made. SAXS images

were collected for each dilution in the series. To collect these images, the samples were continuously flowed through a capillary where scattering images were collected for short (1.0 s) X-ray exposures. Before and after each protein sample, measurements of buffer alone were taken for subsequent buffer subtraction steps.

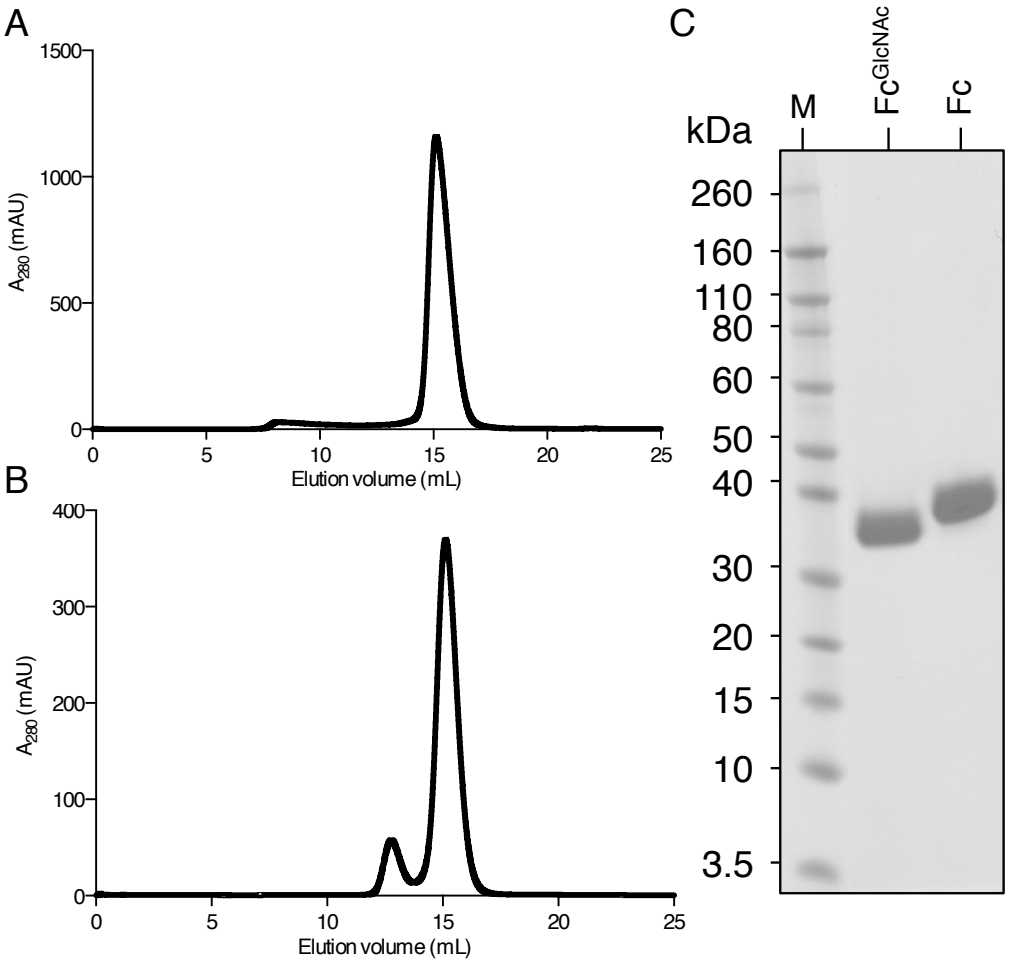


Figure 6.1: Purification of Fc and Fc^{GlcNAc} for SAXS. Chromatograms for the final SEC purification step of Fc (A) and Fc^{GlcNAc} (B) are shown. The A_{280} on the y-axis is given in arbitrary milli absorbance units (mAU). Both Fc and Fc^{GlcNAc} elute from the Superdex 200 10/300 GL column at approximately 15.5 mL. In the chromatogram for Fc^{GlcNAc}, the peak at 13 mL corresponds to the EndoS protein used in the deglycosylation reaction. Protein purity was observed by reducing SDS-PAGE analysis of pooled purified fractions (C).

Scattering images were processed at the beamline automatically using the computer programme PRIMUS [215]. For each protein concentration, curves of the log scattering intensity, $I(q)$ versus q from all exposures (frames) were plotted and averaged (see Section 2.6.4 for a more detailed description). The same was carried out for buffer measurements. Averaged buffer curves were then subtracted from the protein curves to account for the

scattering contribution of the solvent (Fig. 6.2).

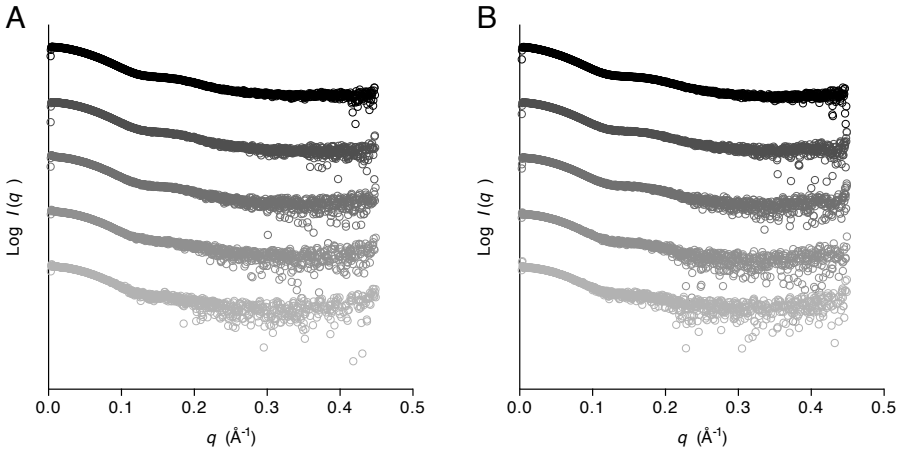


Figure 6.2: The solvent corrected SAXS data for Fc and Fc^{GlcNAc}. The solvent corrected scattered intensities are shown for the 2-fold dilution series of (A) Fc and (B) Fc^{GlcNAc}. The data collected from the highest concentration of both proteins (5 mg mL^{-1}) is shown in black with the lower concentrations shown in increasingly lighter shades of grey. Each dilution is offset along the y axis by a factor of 10^3 so that all curves can be visualised on the same axes.

Scattering data were then further assessed and processed using SCÅTTER. The scattering data from the serially diluted Fc and Fc^{GlcNAc} was used to extrapolate the curves to the intensity at zero concentration ($I(0)$). Further processing was carried out to produce the paired-distance distribution function, *P(r)* and to calculate the real space values for R_g and $I(0)$. The results of these analyses are given in Table 6.1.

Table 6.1: SAXS analysis of Fc and Fc^{GlcNAc}.

	Fc	Fc^{GlcNAc}
$I(0)^{\dagger}$	55.4 (53.0)	48.8 (46.7)
$R_g{}^{\dagger}$ (Å)	29.17 (28.74)	29.23 (28.79)
M_w (kDa)*	58.1	51.2
D_{max} (Å)	101	94
Porod volume (Å^3)‡ (Å^3)	86100	74900
V_c (Å^2)‡	425.0	409.3

†Values were calculated using the Guinier approximation [217].Values in parentheses are real space values calculated using SCÅTTER [277].
*Approximated using the $I(0)$ and the $I(0)_{BSA}$.
‡The Porod volume and the Volume-of-Correlation (V_c) were calculated using SCÅTTER as described in [214].

The values for R_g and $I(0)$ calculated using the Guinier approximation correlate well to the real space values which are calculated using data from the entire scattering curve.

The agreement between these values to ± 5% is an indication of good data quality in the Guinier region ($q \cdot R_g < 1.3$) which in turn is indicative of a non-aggregated sample with no significant radiation damage induced by SAXS.

The Guinier approximated R_g values for Fc and Fc^{GlcNAc} were 29.17 Å and 29.23 Å, respectively. The associated error when using the Guiner approximation can be up to ± 10% when data is selected for values of ($q \cdot R_g$) < 1.3 [213]. Given this precision of the approximated R_g, these values were not classed as being significantly different. For all other calculated parameters, Fc^{GlcNAc} gives values lower than Fc, where the Porod volume gives an approximation of the volume of the particle in solution and D_{max} is the maximum linear dimension in Å.

The $P(r)$ function for both Fc and Fc^{GlcNAc} is shown in Figure 6.3. The different curve shape indicates a different distribution of interatomic vectors. The curve for Fc^{GlcNAc} is skewed to right and shows a decrease in the frequency of interatomic vectors with distances ~20–40 Å.

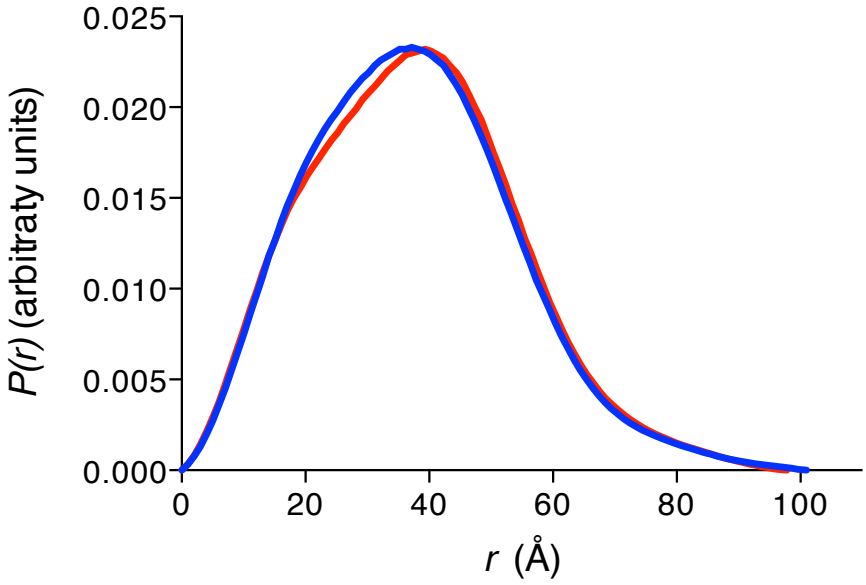


Figure 6.3: The *P(r)* of Fc and Fc^{GlcNAc}. Both curves were calculated using SCÅTTER [277]. Data for Fc is shown in blue (D_{max} = 101) and data for Fc^{GlcNAc} is shown in red (D_{max} = 94).

In a recent paper by Rambo *et al.* [214], further manipulations of SAXS data are described which result in a dimensionless Kratky plot. A traditional Kratky plot ($q^2 \cdot I/(q)$ versus q) describes the behaviour of a protein in solution in terms of foldedness and flexibility. Kratky plots for folded globular proteins follow a Gaussian peak at low q values and then

tend towards zero at higher q values. Unfolded proteins or random coil conformations give Kratky plots which do not tend to zero after the initial peak but instead increase or plateau at these higher q values. To be able to compare the Kratky plot of proteins of different size, measured at different concentrations, a dimensionless Kratky plot can be produced. These plots can be normalised using the Guinier approximated R_g, the V_c and the concentration [214]. This allows a semi-quantitative comparison of flexibility between different proteins. The dimensionless Kratky in Figure 6.4 shows curves which follow a similar shape. Both have a peak of the same amplitude at $q = \sqrt{3}$ Å^{-1} (indicated by the dotted line in Fig. 6.4), which indicates that the proteins are globular and that the Guinier approximation is satisfied. The second and smaller peak at $q = 5$ Å^{-1} is indicative of the multi-domain nature of the protein. The curves deviate from each other about the maximum point of this second peak.

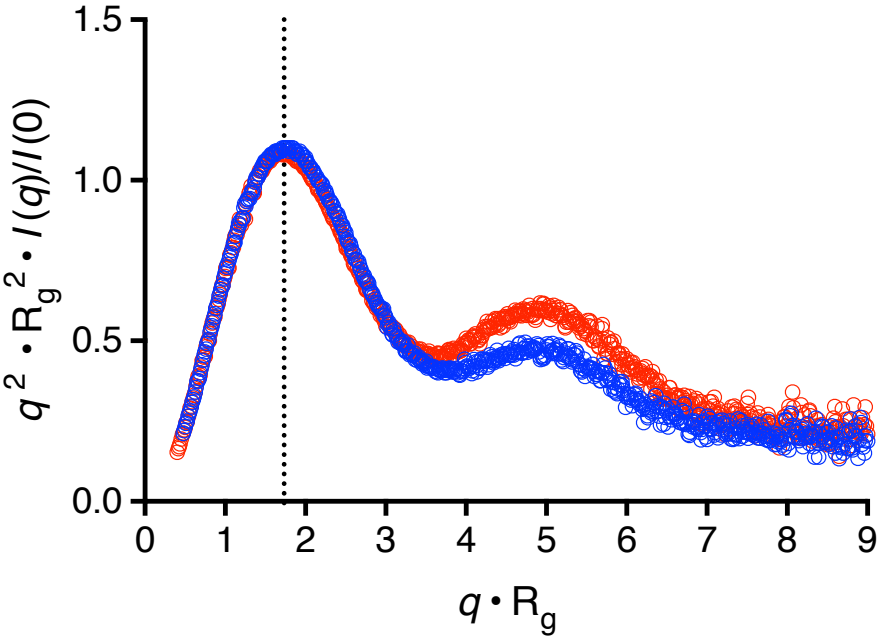


Figure 6.4: Dimensionless Kratky plot of Fc and Fc$^{\mathrm{GlcNAc}}$ normalised to the Guinier approximated R_g as described in [214]. The dashed vertical line indicates the $q \cdot {R_g}^2$ value $\sqrt{3}$ (1.732). Data for Fc is shown in blue and data for Fc$^{\mathrm{GlcNAc}}$ is shown in red.

6.2 Structural studies of deglycosylated murine IgG subclasses

This section describes the structures of the Fc region of deglycosylated murine 6A6-IgG1 and 6A6-IgG2a. Both structures were determined using X-ray crystallography with an aim to understand the ability of the IgG2a subclass to maintain FcγR binding upon EndoS deglycosylation [144].

In general, EndoS deglycosylation of IgG causes a significant reduction in FcγR binding affinity. However, this effect has been shown to be IgG subclass dependent [145]. In a previous study by Albert *et al.* [144] EndoS was used to deglycosylated IgG in mouse models of autoimmune disease. When autoantibodies were deglycosylated with EndoS, a reduction of autoantibody-mediated tissue inflammation was observed. This was observed when using IgG1 and IgG2b autoantibodies but was not observed for the murine IgG2a subclass of autoantibodies. The ability of murine IgG2a to induce an autoimmune response was maintained even upon EndoS deglycosylation. This was then attributed to the observed ability of murine IgG2a to maintain FcγR binding even when deglycosylated with EndoS [145].

In terms of FcγR binding affinity, the resistance of murine IgG2a to the effects of deglycosylation is of interest due to the potential applications for an EndoS resistant antibody (as discussed in Chapter 1, Section 1.4.3). Although murine IgG2a is not EndoS resistant *per se*, understanding how this antibody maintains receptor binding affinity may offer a route for developing antibodies resistant to the *effects* of deglycosylation as opposed to deglycosylation itself.

As its name suggests, the Fc (fragment crystallisable) is a good target for X-ray crystallographic studies. With this in mind, structural studies using X-ray crystallography were pursued for the deglycosylated Fc portions of *Mus musculus* 6A6-IgG1 and 6A6-IgG2a (*Mm*IgG1Fc and *Mm*IgG2aFc, respectively). These were carried out with an aim to identify structural features of *Mm*IgG2aFc which may confer resistance to the effects of deglycosylation.

X-ray crystallographic studies require the growth of protein crystals which, in general, require large amounts of pure protein. To this end, constructs were made to allow for re-

combinant expression of the Fc region (residues 224–446) of *Mm*IgG1Fc and *Mm*IgG2aFc. Both constructs encompassed at least one of the cysteine residues of the hinge region so that dimerisation of the Fc region via disulphide bonds could occur. Both constructs were transiently expressed in HEK 293T cells as C-terminal hexahistidine fusions. Similar to the human IgG1 Fc constructs, *Mm*IgG1Fc and *Mm*IgG2aFc were purified by IMAC and SEC.

The glycosylation state of the purified *Mm*IgG constructs was assessed by HPLC (Fig. 6.5). The glycans of the murine IgG Fc constructs showed lower levels of galactosylation when compared to recombinantly expressed human IgG1 Fc or human serum IgG. The constructs also showed no detectable sialylation through exoglycosidase assays with neuraminidase from *C. perfringens*. In addition to the complex biantennary glycans, the hybrid glycoform $Man_5GlcNAc_1GlcNAc_2Fuc$ was present on both *Mm*IgG1Fc and *Mm*IgG2aFc.

To prepare both isoforms of murine IgG Fc for crystallisation, both proteins were deglycosylated with EndoS for 16 hours at 37 °C. Comprehensive analysis of EndoS against hybrid glycoforms had not been carried out at this stage, therefore *Mm*IgG1Fc and *Mm*IgG2aFc were also treated with EndoH to ensure hydrolysis of the hybrid glycoform. EndoH, like EndoS, cleaves *N*-linked glycans between the two GlcNAc residues of the chitobiose core and can hydrolyse both high-mannose hybrid glycoforms [204]. Treatment of *Mm*IgG1Fc and *Mm*IgG2aFc with both EndoS and EndoH was therefore carried out to ensure homogeneously deglycosylated samples.

6.2.1 X-ray crystallography

For crystallisation trials, pure and monodisperse protein samples are desirable, therefore further purification of the deglycosylated samples was carried out. Both deglycosylated murine Fcs were purified by SEC using a HiPrep Superdex 200 16/60 column. A subsequent SEC step using a HiPrep Superdex 75 16/60 was carried out in an effort to remove residual EndoH. The result of this final purification step can be seen in Figure 6.6. The purified *Mm*IgG1Fc and *Mm*IgG2aFc proteins were used for crystallisation trials as detailed in Section 2.7.

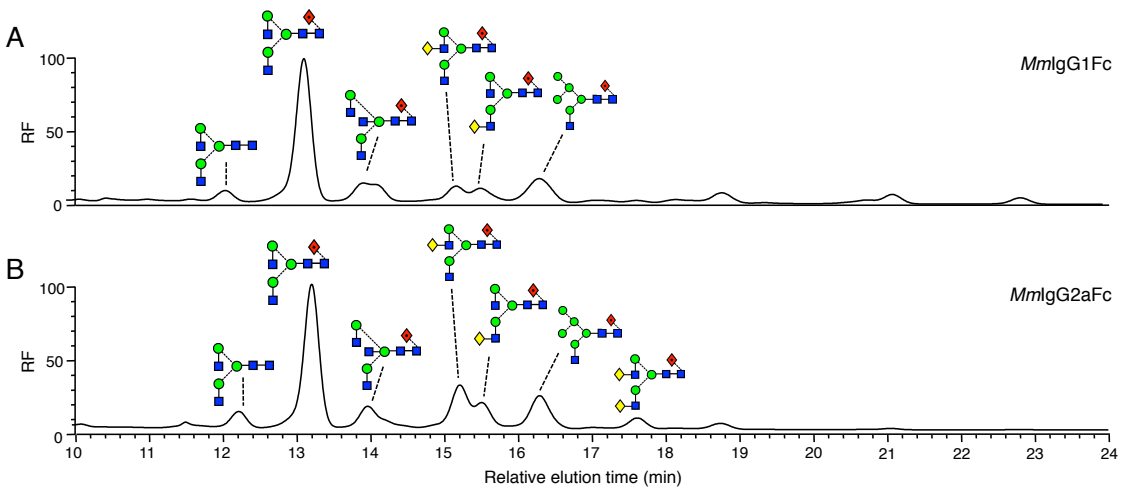


Figure 6.5: Glycosylation profiles of *Mm*IgG1Fc and *Mm*IgG2aFc. HPLC analysis is shown for fluorescently labelled glycans released from recombinantly expressed (A) *Mm*IgG1Fc and (B) *Mm*IgG2aFc. Peaks were assigned using MALDI-TOF MS. Glycans are labelled following Oxford glycan nomenclature [83] as described in Chapter 5, Figure 5.1. Fluorescence is reported as relative fluorescence units (RF).

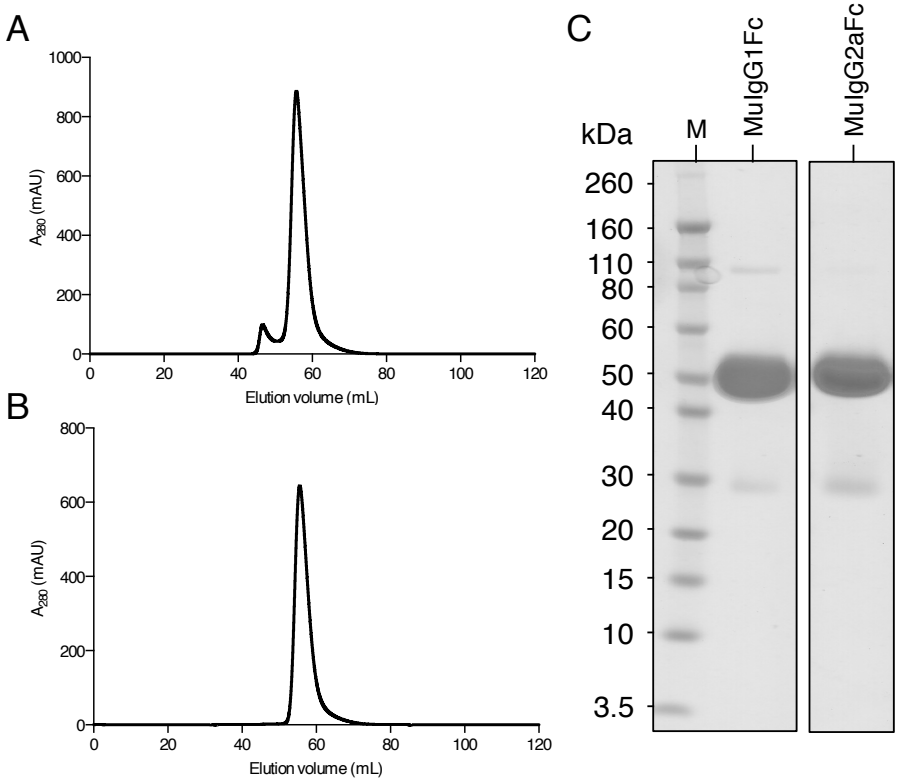


Figure 6.6: Purification of deglycosylated *Mm*IgG1Fc and *Mm*IgG2aFc. Chromatograms for the final SEC step of the purification protocol using a HiPrep Superdex 75 16/60 column are shown for (A) *Mm*IgG1Fc and (B) *Mm*IgG2aFc. In both cases the protein of interest elutes at approximately 55 mL. *Mm*IgG1Fc has an additional peak at 48 mL corresponding to residual EndoS from the deglycosylation reaction. (C) SDS-PAGE analysis of pooled A_{280} peak fractions shows high purity of the samples. The minor contaminants are identifiable as EndoS (110 kDa) and EndoH (30 kDa). M; molecular weight marker.

Crystallisation

In both cases, protein crystals were obtained using the vapour diffusion method where crystals were grown in a sitting-drop format in 96-well plates. The crystallisation drops were formed of 100 nL protein solution and 100 nL of precipitant equilibrated against 50 μL of mother liquor.

Crystals of *Mm*IgG1Fc were grown in 50 mM $MgCl_2 \cdot 6H_2O$, 100 mM HEPES pH 7.0 and 30% w/v PEG 550 MME (Fig. 6.7A). The *Mm*IgG2a Fc crystals were grown from 200 mM $MgCl_2$, 100 mM sodium acetate pH 5.5 and 20% w/v PEG 6000 (Fig. 6.8A). For both *Mm*IgG1Fc and *Mm*IgG2aFc, crystals were observed within 24 hours of setting up crystallisation experiments. The crystals were left to grow for additional 14 days at which point they were harvested for X-ray diffraction studies.

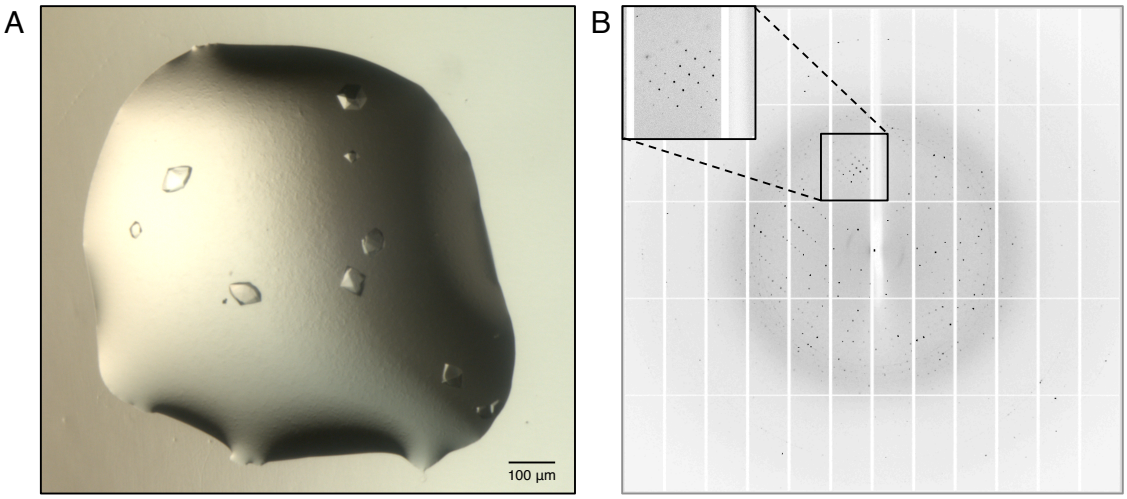


Figure 6.7: Crystallisation of *Mm*IgG1Fc. (A) The hexagonal crystals of *Mm*IgG1Fc grown from 50 mM $MgCl_2 \cdot 6H_2O$, 100 mM HEPES pH 7.0 and 30% w/v PEG 550 MME and were pictured using a light microscope. (B) An X-ray diffraction pattern from *Mm*IgG1Fc crystals reveals reflections from a 0.2° oscillation. The inset panel shows a magnification of the reflections.

Diffraction data collection and processing

Upon crystal harvesting, both *Mm*IgG1Fc and *Mm*IgG2aFc crystals were immersed into mother liquor containing an additional 25% (v/v) ethylene glycol. This was followed by rapid transfer to liquid nitrogen. This cryocooling process was carried out to reduce potential damage caused by synchrotron radiation during diffraction data collection (as reviewed in [270]). Data collection at DLS was kindly carried out by Dr Edward Lowe

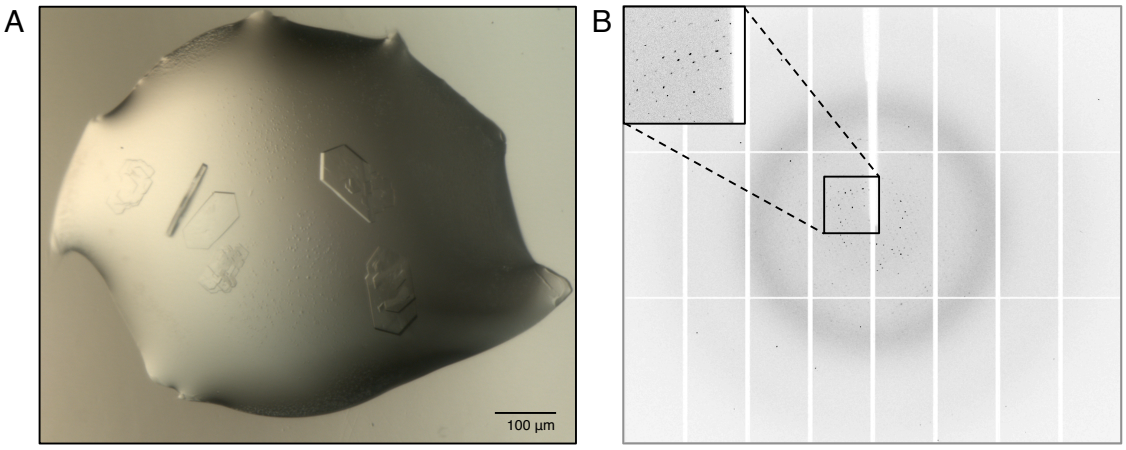


Figure 6.8: Crystallisation of *Mm*IgG2aFc. The plate-like crystals of *Mm*IgG2aFc (A) grown from 200 mM $MgCl_2$, 100 mM sodium acetate pH 5.5 and 20% w/v PEG 6000 and were pictured using a light microscope. An X-ray diffraction pattern from *Mm*IgG2aFc crystals reveals reflections from a 0.2° oscillation (B). The inset panel shows a magnification of the reflections.

(Department of Biochemistry, University of Oxford). Data collection statistics are given in Table 6.3 and example diffraction images are shown for both *Mm*IgG1Fc and *Mm*IgG2aFc in Figures 6.7B and 6.8B, respectively. Further details of the beamlines and data collection strategies are described earlier in Chapter 8, Section 2.8.1.

X-ray diffraction data collected from crystals of *Mm*IgG1Fc were integrated and scaled using the programmes XDS [288] and XSCALE, which were executed using xia2 [226]. For *Mm*IgG2aFc, X-ray diffraction data integration and scaling were carried out using the programmes XDS [288] and CCP4-Aimless [289], run using the data processing suite autoPROC[225]. Table 6.3 shows the crystallographic data collection and processing statistics.

Data quality was assessed using the measures $CC_{1/2}$ and $\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$ as described by Karplus and Diederichs [293] and as implemented by CCP4-Aimless [289]. Resolution limits were chosen using values of $CC_{1/2} > 0.5$ and $\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle \geq 2.0$ as a guide. Crystals of *Mm*IgG1Fc were of hexagonal symmetry in space group $P6_122$. The resolution of the resultant X-ray diffraction data was cut at 2.23 Å. *Mm*IgG2aFc crystals exhibited monoclinic symmetry in space group $P2_1$. The diffraction images from crystals of *Mm*IgG2aFc were processed to give a 3.27 Å resolution dataset. The low value of $\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$ (1.9) for the outer resolution range for data from crystals of *Mm*IgG2aFc can be attributed to poor signal to noise in the k direction where $\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$ reaches 2.0 at 3.40 Å resolution.

Given this, the *Mm*IgG2aFc diffraction images were also processed to the lower resolution cut off of 3.47 Å. This lower resolution dataset was used in initial model building and refinement stages as detailed below.

Table 6.2: *Mm*IgG1Fc and *Mm*IgG2a data collection statistics.

	*Mm*IgG1Fc	*Mm*IgG2aFc
Beamline	DLS I02	DLS I04-1
Wavelength (Å)	0.97860	0.92000
Resolution range (Å)	61.30–2.23 (2.49–2.23)	57.90–3.27 (3.36–3.27)
Spacegroup	$P6_122$	$P2_1$
Cell dimensions (Å)		
a, b, c (Å)	96.62, 96.62, 90.06	71.58, 71.65, 100.3
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 101.6
Total reflections	117731 (32004)	50640 (3361)
Unique reflections	12594 (3458)	15160 (1075)
Multiplicity	9.3 (9.3)	3.3 (3.1)
Mosaicity (°)	0.10	0.00
Completeness (%)	99.6 (98.5)	97.4 (96.6)
$\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$	25.0 (3.5)	9.2 (1.9)
$R_{\text{merge}}{}^{\dagger}$	0.065 (0.622)	0.114 (0.699)
$R_{\text{meas}}{}^{\ddagger}$	0.072 (0.697)	0.158 (0.965)
$R_{\text{p.i.m.}}{}^{*}$	0.032 (0.309)	0.109 (0.663)
$CC_{1/2}$	0.999 (0.870)	0.987 (0.639)

Values in parentheses refer to the outer resolution range.
$^{\dagger}R_{\text{merge}} = \sum_{\mathbf{h}}\sum_i \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}}\rangle \mid / \sum_{\mathbf{h}}\sum_i I_{\mathbf{h}}$, where $I_{\mathbf{h}i}$ is the scaled intensity of an individual measurement and $\langle I_{\mathbf{h}}\rangle$ is the average intensity from multiple observations of the unique reflection $\boldsymbol{h}$.
$^{\ddagger}R_{\text{meas}} = \sum_{\mathbf{h}}\sum_i \sqrt{(n_{\mathbf{h}}/n_{\mathbf{h}}-1)} \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}}\rangle \mid / \sum_{\mathbf{h}}\sum_i I_{\mathbf{h}}$
$^{*}R_{\text{p.i.m.}} = \sum_{\mathbf{h}}\sum_i \sqrt{(1/n-1)} \mid I_{\mathbf{h}i} - \langle I_{\mathbf{h}}\rangle \mid / \sum_{\mathbf{h}}\sum_i I_{\mathbf{h}}$

Phasing

To identify appropriate search models for use in molecular replacement, the amino acid sequences of both proteins were submitted to the FFAS server [295], searching against sequences in the PDB as of January 2013. In both cases, the hit with the highest sequence homology was used as a search model.

For *Mm*IgG1Fc, the best search model was the aglycosylated structure of IgG1 Fc from *M. musculus* (PDB ID: 3HKF). This showed 98% sequence identity to *Mm*IgG1Fc. The crystals used in structure determination of 3HKF exhibited the same cell dimensions and symmetry as the *Mm*IgG1Fc crystals, rendering molecular replacement unnecessary: as the point group was 622, there is no indexing ambiguity, and to obtain the initial phases

it was sufficient to run a single round of rigid body refinement with one rigid body per chain, starting from the 3HKF coordinates, with the computer programme autoBUSTER [229]. This corresponded to a solvent content of 52.56% (V_m = 2.59 Å^3 Da^{-1} [296]).

Molecular replacement using the programme Phaser [227] was used to determine the structure of *Mm*IgG2aFc. The search model used for molecular replacement of *Mm*IgG2aFc was the glycosylated Fc structure of a monoclonal IgG2b antibody from *M. musculus* (PDB ID: 2RGS). This protein has 76% sequence identity to *Mm*IgG2aFc. A total of four monomers were identified in the asymmetric unit giving a solvent content of 54.94% (V_m = 2.73 Å^3 Da^{-1} [296]). Initially the CH3 dimer (residues 345–443) of 2RGS was used to find the four copies of the *Mm*IgG2aFc CH3 in the electron density. The CH2 monomer (residues 237–344) from 2RGS chain A was then used to locate the four CH2 domains of *Mm*IgG2aFc with four successive rounds of molecular replacement in presence of the already placed CH3 domains.

Structure building and refinement

Iterative rounds of manual model building in Coot [228] and refinement with BUSTER (Global Phasing Ltd.) [229] were carried out for each structure. A full description of refinement and model statistics is shown in Table 6.3. For *Mm*IgG1Fc, no additional restraints were applied to atom positions or B factors during refinement.

For *Mm*IgG2aFc, all initial refinement rounds were done against the lower resolution data truncated to 3.47 Å. Additional TLS [326] and NCS restraints were also applied during refinement of *Mm*IgG2aFc. TLS bodies were determined by the TLS server [302] and NCS restraints were determined automatically by BUSTER [303]. In an effort to maintain secondary structure elements in the model, target restraints were also applied using the original molecular replacement solution for *Mm*IgG2a as a target structure. In the later rounds of refinement, hydrogen atoms were added to the model using the programme REDUCE via the MolProbity server [305]. Refinement with hydrogens was carried out with an aim to improve the R factors of the model [306]. A direct comparison between refinement of models with and without hydrogens showed that exclusion of hydrogen yielded higher R factors. For the final round of refinement, the model was refined against the data cut to 3.27 Å resolution as this modest increase in resolution did not dramatically

increase the R factors nor did it widen the gap between R_{work} and R_{free}.

The geometry and validity of the resultant models was assessed using MolProbity [307]. The Ramachandran plots generated for these models are shown in Figure 6.9. In both cases, there were no Ramachandran outliers and $\geq 98\%$ of residues were in Ramachandran favoured regions.

Table 6.3: *Mm*IgG1Fc and *Mm*IgG2a refinement statistics.

	*Mm*IgG1Fc	*Mm*IgG2aFc
Resolution range (Å)	61.30–2.23 (2.44–2.23)	57.90–3.27 (3.50–3.27)
Spacegroup	$P6_122$	$P2_1$
Observations	12552 (2883)	15150 (2684)
Observations in free set	612 (150)	761 (131)
R	0.1877 (0.2154)	0.2621 (0.3057)
R_{work}[†]	0.1859 (0.2132)	0.2610 (0.3029)
R_{free}[‡]	0.2252 (0.2535)	0.2830 (0.3636)
RMSD bond lengths (Å)	0.010	0.0098
RMSD bond angles (°)	1.055	1.120
Ramachandran outliers	0	0
Ramachandran favoured	98%	99.4%
Residues modelled (chain)	238–444 (A)	237–443 (A) 237–443 (B) 237–265, 268–444 (C) 235–237, 239–265, 268–443 (D)
Waters modelled	101	66
Average B, protein (Å^2)	50.221	99.408
Average B, water (Å^2)	50.433	35.300
Non-protein molecules	8 Ethylene glycol	0

Values in parentheses refer to the outer resolution range.
[†]$R_{work} = \sum_{hkl} | \, | F_o | - | F_c | \, | \, / \sum_{hkl} | F_o |$.
[‡]R_{free} was calculated in an identical manner to R_{work} using 5% of the data which were excluded prior to refinement.

6.2.2 Models of *Mm*IgG1Fc and *Mm*IgG2aFc

The structure of the *Mm*IgG1Fc monomer shows the expected fold for an IgG Fc. The CH2 and CH3 domains each comprise an immunoglobulin fold with the three- and four-stranded β-sheets connected by a disulphide bond. The monomer of the asymmetric unit forms the physiological dimer across one of the two-fold rotational symmetry axes as shown in Figure 6.10. The dimeric structure visible through this symmetry operation forms the canonical horse-shoe shape of the homodimeric IgG Fc protein.

Electron density for the C^{α} residues of the D/E, B/C and F/G loops were visible with the

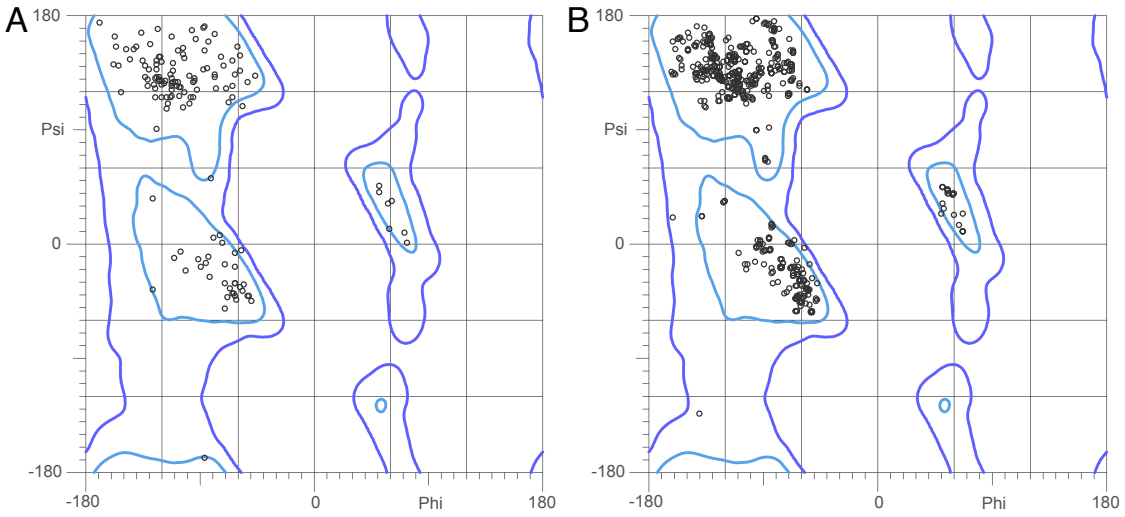


Figure 6.9: Ramachandran analysis of *Mm*IgG1Fc (A) and *Mm*IgG2aFc (B). Plots were calculated using MolProbity [307] where black circles indicate each residue with respect to its psi angle (°) (y-axis) and phi angle (°) (x-axis). Favoured regions are outlined in light blue and allowed regions are outlined in dark blue.

exception of density for residue N297 (Fig. 6.14B). The electron density for the side-chains in these loops are not well defined in some cases. Density was not observed for the terminal GlcNAc or Fuc residues covalently linked to N297. This lack of well-defined density for the loop regions is echoed by the higher *B* factors in these regions of the protein (Fig. 6.13).

The inter-CH2 distance measured between the C^{α} of P329 across the dimer formed by the two-fold axis is 15.2 Å. This distance is shorter than that observed for any of the human IgG Fc structures (18.9–41.1 Å) [44] and is characteristic of a 'closed' conformation of the CH2 domains. This distance is larger however, than that of the 'super-closed' state of the aglycosylated murine IgG1 Fc structure (11.7 Å) used for molecular replacement of *Mm*IgG1Fc (PDB ID: 3HKF) [327].

The structure of *Mm*IgG2aFc is formed of two physiological dimers (four monomers) in the asymmetric unit (Fig. 6.11). Each monomer has both a CH2 and CH3 domain and like *Mm*IgG1Fc, these domains have a seven-stranded immunoglobulin fold, each stabilised by a disulphide bond between the three-stranded and 4-stranded sheets. The two dimers in the asymmetric unit are related by non-crystallographic symmetry through a two-fold rotational axis. The *Mm*IgG2a dimers are packed in the crystals so that the CH2 of an adjacent asymmetric unit interlinks into the void between the CH2 domains. The

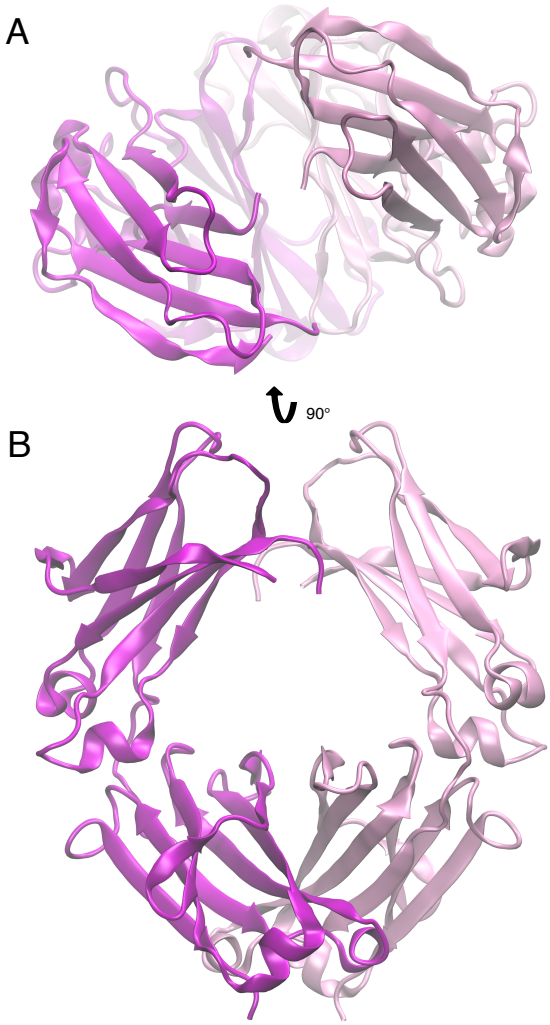


Figure 6.10: The crystal structure of *Mm*IgG1Fc. The *Mm*IgG1Fc monomer representative of the asymmetric unit is shown in magenta. An adjacent symmetry copy of the monomer related by a two-fold rotational axis is shown in light pink. Together, these monomers form the physiological Fc dimer shown from a top-down view (A) and a side view (B). Both monomers are shown in cartoon representation.

interlinked CH2 domains complement each other by taking up the space usually occupied by the *N*-linked glycan. Electron density for neither the GlcNAc nor the Fuc residues was observed for any of the monomers.

Superposition of dimer AD onto BC gives an RMSD of 0.35 Å. The electron density for the C^{α} atoms in the CH2 loops (D/E, B/C and F/G) are defined for both the A and B monomers. However, the B/C loop in monomer D and the D/E loop in monomers C and D had electron density that was either poor or not visible (Fig. 6.14C–F). Several residues were omitted from the model due to this lack of density as detailed in Table 6.3. The higher *B* factors of these loops reflect the poor electron density for these regions (Fig. 6.13B).

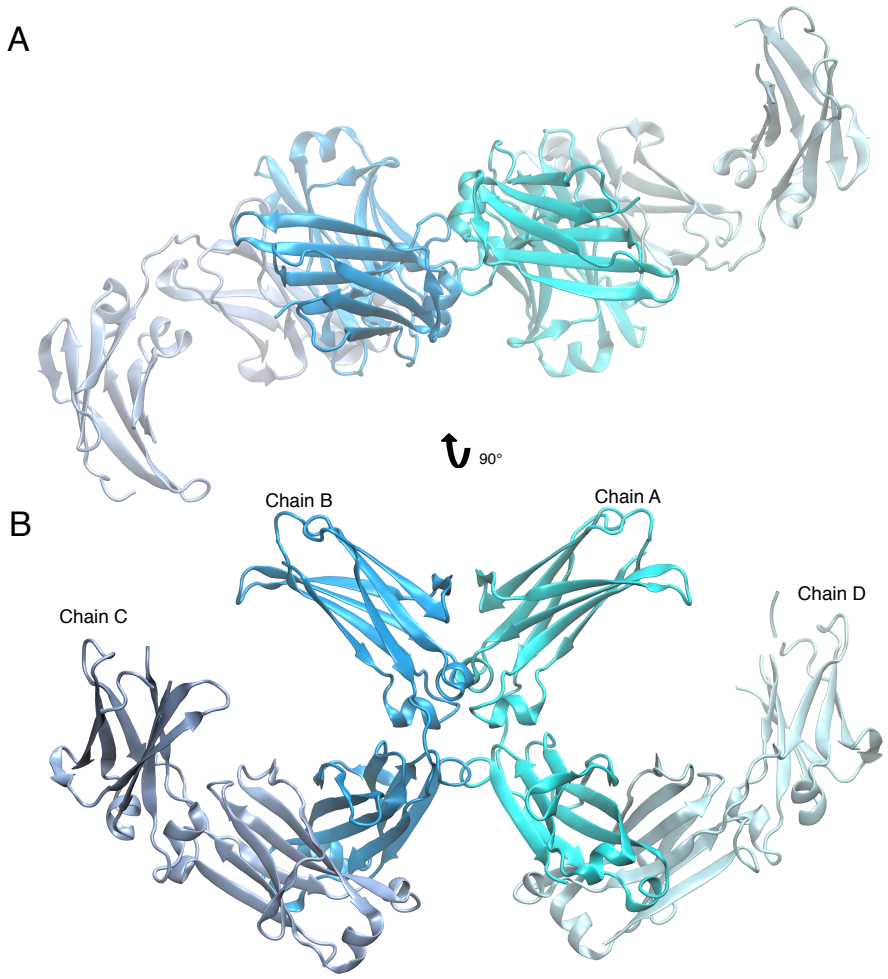


Figure 6.11: The crystal structure of *Mm*IgG2aFc. The two dimers of the asymmetric unit are shown in cartoon representation from above (A) and rotated 90° to give a side view (B). The A and D monomers are shown on the right in cyan and light cyan, respectively. The B and C monomers are shown on the left in turquoise and ice blue, respectively.

The inter-CH2 distance measured between the P329 residues of dimer BC is 43.7 Å. From the current pool of Fc structures, this is the longest inter-CH2 distance that has been observed. The inter-CH2 distance of the AD monomer is comparable (considering the RMSD of the overlaid dimers) but could not be measured directly by the P329–P329 distance as residue P329 is not built into the model of monomer D. This large distance is uncharacteristic of deglycosylated structures which usually have shorter CH2-CH2 distances and more 'closed' conformations [44, 41]. A top down view showing the apparent openness of each structure is shown in Figure 6.12. It must be stressed that the CH2 distance for the *Mm*IgG2aFc structure is heavily influenced by the crystal packing described above where the CH2 domains of adjacent dimers interlock with one another. This crystal packing inhibits the often observed 'closed' structure for deglycosylated *Mm*IgG2aFc as the CH2

domain from a dimer in an adjacent asymmetric unit effectively props the CH2 domains open.

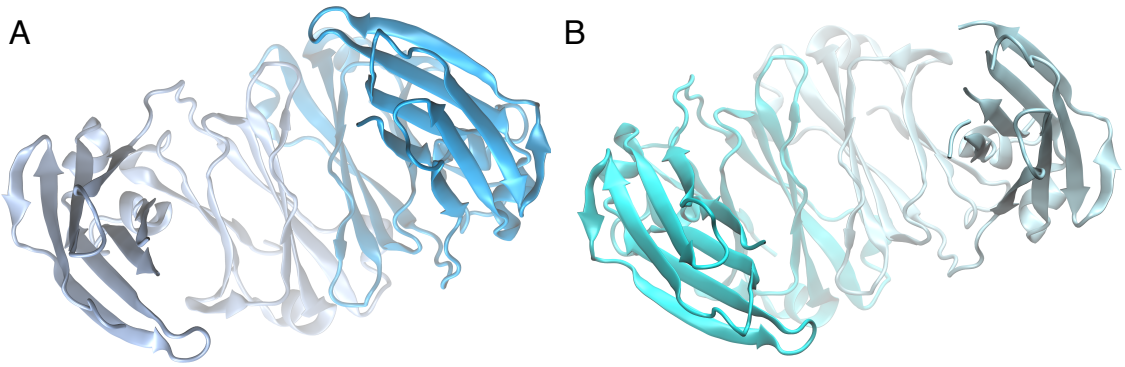


Figure 6.12: The open conformation of the *Mm*IgG2aFc structure. Each dimer of the asymmetric unit is shown from a top-down view to reveal the inter-CH2 spacing. The structures of *Mm*IgG2aFc monomers B and C (A) and monomers A and D (B) are shown in cartoon representation and are coloured according to Fig. 6.11.

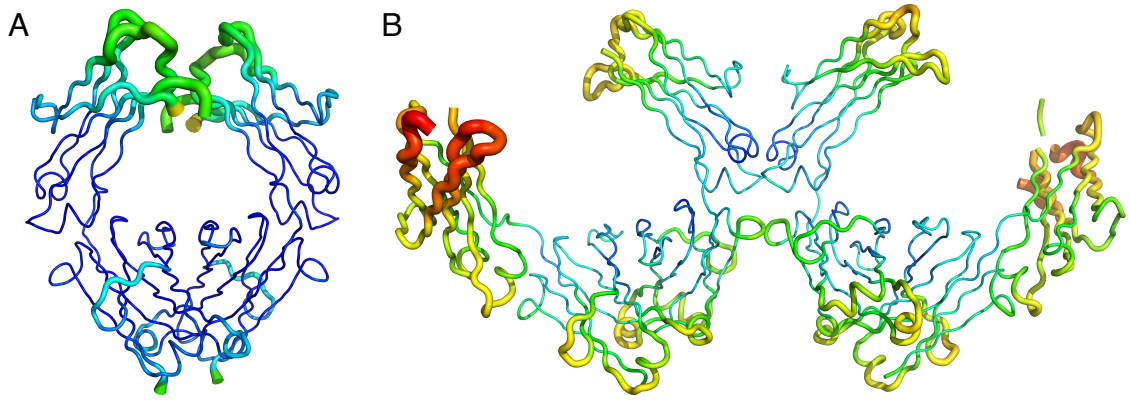


Figure 6.13: A representation of the *B* factors of *Mm*IgG1Fc (A) and *Mm*IgG2aFc (B) structures. Both structures are represented as tubes by drawing only the C^{α} of the polypeptide chain. Both the colour and thickness of the tubes is indicative of the *B* factor associated with the structure. Dark blue, ($\sim$ 20 Å^2); cyan, ($\sim$ 50 Å^2); green, ($\sim$ 80 Å^2); yellow, ($\sim$ 100 Å^2); red, ($\sim$ 130 Å^2). A thicker tube indicates a higher *B* factor.

In both structures, the B-factors are largest for the solvent facing loops in both the CH2 domains and CH3 domains, as is characteristic of IgG Fc structures. These loops are involved in binding to the Fc receptors, as is the lower-hinge region (residues ~224–237). The lower-hinge region is often not resolved in crystal structures of the Fc alone. This is usually attributed to its lack of assigned secondary structure and its potential to be disordered in solution. However, this region is resolved in crystal structures of Fc:FcγR complexes where it can be seen making multiple contacts with the FcγR (reviewed in [27, 328]). The lower-hinge region (residues 224–237) was included in both *Mm*IgG1Fc and *Mm*IgG2aFc constructs used for crystallisation however, neither dataset gave electron

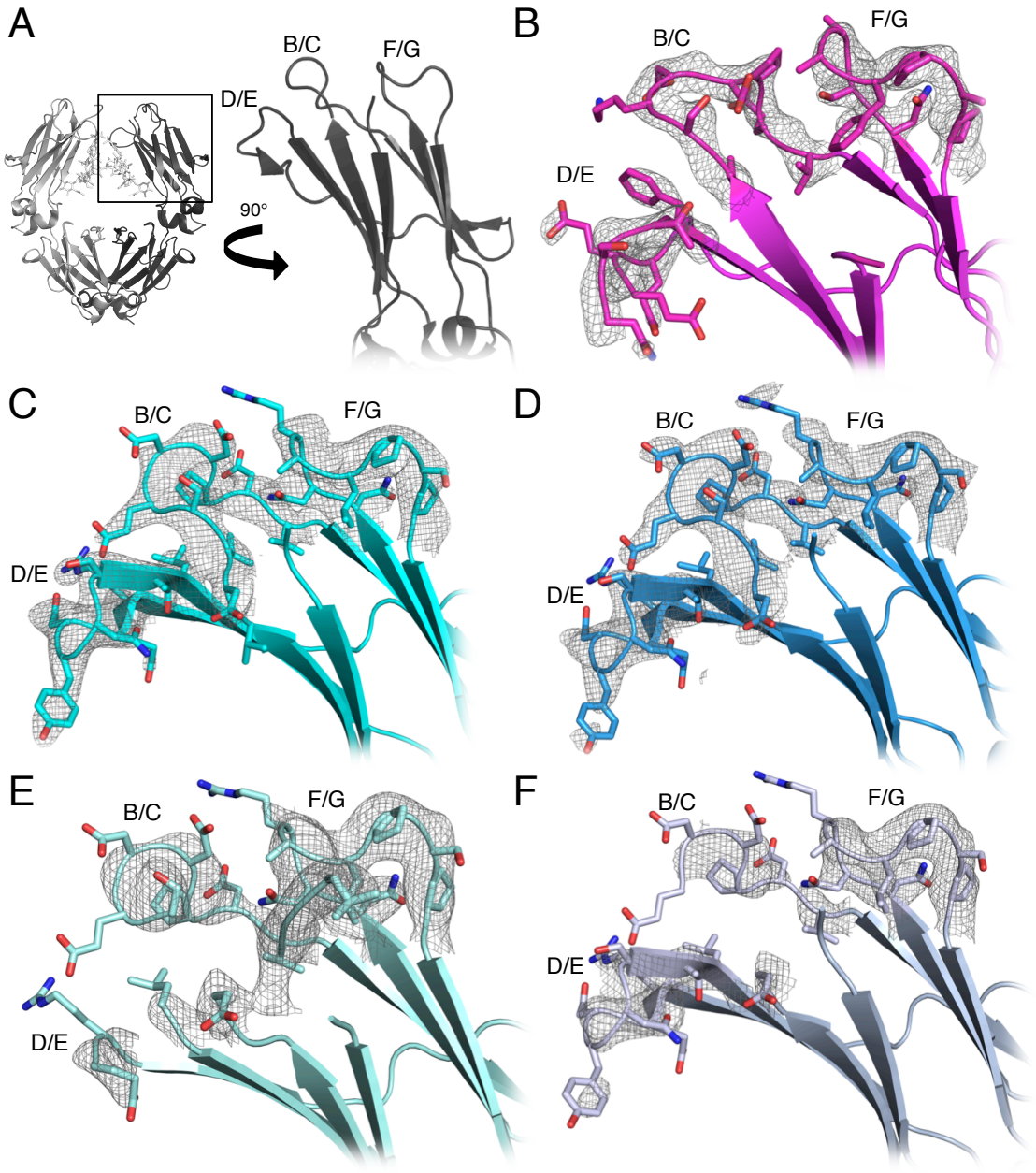


Figure 6.14: Electron density for the CH2 loops of *Mm*IgG1Fc and *Mm*IgG2aFc. The crystal structure for the glycosylated human IgG1 Fc (3AVE) is shown with a rotated and magnified view of the CH2 loops (A). The D/E, B/C, and F/G loops are labelled accordingly. The protein is shown in cartoon representation with glycans removed in the right panel for clarity. For the remaining figures, all proteins are represented as cartoons. Residues of the loop regions are shown as sticks with oxygens coloured red and nitrogens in blue. In this figure the loop residues were classed as such, D/E loop, 293–300; B/C loop, 266–273; F/G loop, 324–331. Hinge residues 235–236 are also represented as sticks when visible in the structure. A $2F_c$-F_o map is displayed around the loops and hinge and is contoured to 1σ. The panels show all monomers from the asymmetric unit of both structures and correspond to *Mm*IgG1 chain A (B), *Mm*IgG2a chain A (C), *Mm*IgG2a chain B (D), *Mm*IgG2a chain D (E) and *Mm*IgG2a chain C (F).

density for these residues. The exception to this was for the D monomer of *Mm*IgG2aFc where residues 235 and 236 were modelled into electron density.

Structural comparison to IgG Fc structures

The angle between the CH2 and CH3 domains of IgG Fc structures is flexible and is often characterised by the inter-CH2 distance. For human IgG1 Fcs alone this distance (listed here as the distance between P329 C^{α} atoms) ranges from 18.9–41.1 Å. These values come from Fcs crystallised with varying degrees of glycosylation both with and without bound receptor [44]. A summary of the crystal structures currently available for human IgG1 Fcs is given in Figure 6.15. The structures listed do not encompass Fcs which have been mutated and does not include the Fc portion of the full-length IgG1 Fc structure for the anti-HIV IgG1 B12 antibody (PDB ID: 1HZH). The structures for *Mm*IgG1Fc and *Mm*IgG2aFc were put into the context of this library of Fc structures in an attempt to cluster and identify structural effects of deglycosylation.

A top-down view of the aligned Fc structures (Fig. 6.15, top-panels) shows that the 'openness' of the CH2 domains with respect to each other varies between all unbound states, whether glycosylated or not. With respect to the position of the CH2 domains, there is overlap in the conformational space explored by both glycosylated and de- or algycosylated structures. In the receptor-bound structures, the inter-CH2 distance is well defined. The side view of the Fc dimers (Fig. 6.15, lower panels) shows that the conformation of the CH3 dimer is relatively constant across different glycosylation and receptor-binding states.

The Fc structures were aligned to the CH2 domain in an effort to determine whether the conformation of the D/E, B/C and F/G loops was dependent on either glycosylation state or on receptor-binding state (Fig. 6.16). Similar to the quaternary structure of the CH2 domains, the CH2 loop conformations varied within states where multiple loop conformations were evident for glycosylated and de- or aglycosylated Fcs. Some variation in CH2 loop conformations was also apparent for the receptor-bound Fc structures. The *Mm*IgG1Fc structure shows a higher deviation from the loop conformations seen in the human IgG1 Fc crystal structures in comparison to *Mm*IgG2aFc. *Mm*IgG2aFc closely follows the loop conformations of the published human IgG1 Fc structures.

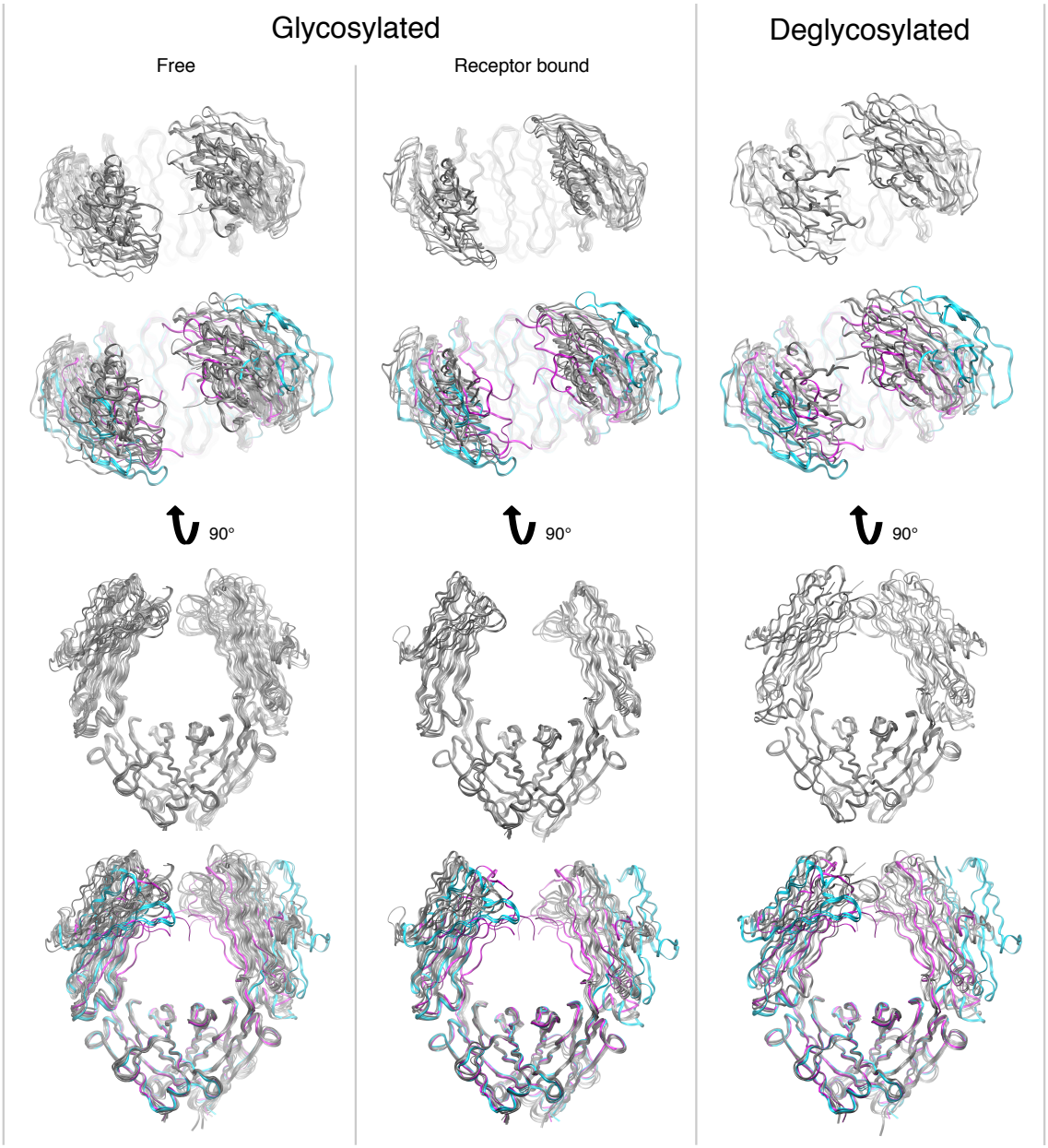


Figure 6.15: Structural alignment of human IgG Fc crystal structures with *Mm*IgG1Fc and *Mm*IgG2aFc. All PDBs were aligned using a least squared fit algorithm (LsqKab) [329] to the CH3 domain of the A chain of 3AVE. PDBs were classed based on glycosylation state. Glycosylated structures were classed as either free (1FC1, 1H3U, 1H3V, 1H3W, 1H3Y, 2DTS, 2WAH, 3DO3, 4B71, 4BYH, 4KU1, 4Q6Y, 4Q7D) or receptor bound (1E4K, 1T83, 1T89, 3AY4, 3SGJ, 3SGK). Deglycosylated structures included both aglycosylated (3S7G) and enzymatically-deglycosylated (4ACP, 3DNK) structures. All structures are shown in ribbon representation where both the glycans and hinge residues have been omitted for clarity. *Mm*IgG2a structures are shown in cyan, *Mm*IgG1 is shown in magenta and all remaining structures are shown in silver. Overlays of the PDB structures are shown with (bottom) and without (top) *Mm*IgG1 and *Mm*IgG2a structures where both a top-down view (upper panels) and a side view (lower panels) of the Fc regions are displayed.

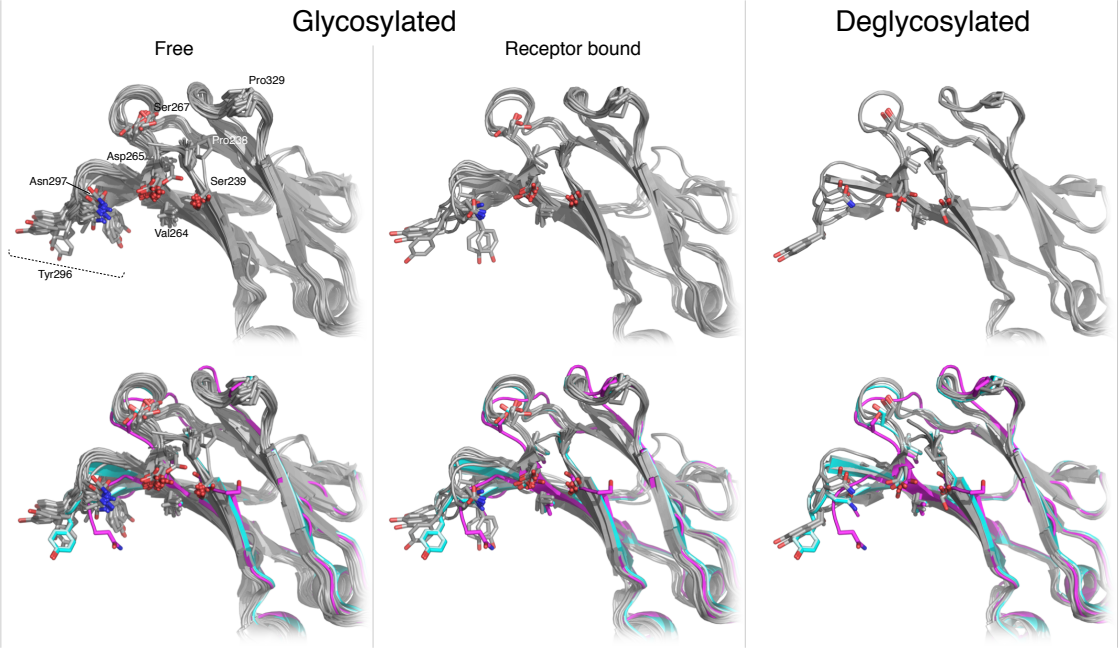


Figure 6.16: Structural alignment of the CH2 domain of IgG1 Fc crystal structures with *Mm*IgG1Fc and *Mm*IgG2aFc. All PDBs were aligned using a least squared fit algorithm (LsqKab) [329] to the CH2 domain of the A chain of 3AVE. PDBs were classed based on glycosylation state. Glycosylated structures were classed as either free (1FC1, 1H3U, 1H3V, 1H3W, 1H3Y, 2DTS, 2WAH, 3DO3, 4B71, 4BYH, 4KU1, 4Q6Y, 4Q7D) or receptor bound (1E4K, 1T83, 1T89, 3AY4, 3SGJ, 3SGK). Deglycosylated structures included both aglycosylated (3S7G) and enzymatically-deglycosylated (4ACP, 3DNK) structures. All structures are shown in cartoon representation where both the glycans and hinge residues have been omitted for clarity. Residues involved in FcγR binding are shown in stick representation. Tyr296 is also shown in stick representation to emphasise changes in the D/E loop. *Mm*IgG2a structures are shown in cyan, *Mm*IgG1 is shown in magenta and all remaining structures are shown in silver. Overlays of the PDB structures are shown with (bottom) and without (top) *Mm*IgG1 and *Mm*IgG2a structures.

In an effort to identify unique residues of *Mm*IgG2a which might be contributing to its ability to maintain FcγR binding upon deglycosylation with EndoS, a multiple sequence alignments of the protein sequences for the Fc region of all IgG subclasses from both human and mouse were carried out using Clustal Omega [29]. The sequences for *Mm*IgG1Fc and *Mm*IgG2aFc were also included in the alignment of mouse sequences. There is a higher sequence variability between the Fc sequences of different mouse IgG subclasses (Fig. 6.17) when compared to the variation in sequence across human IgG subclasses (Fig. 6.18). As with human IgG, the sequence and length of the lower-hinge regions of mouse IgGs vary significantly with subclass. Unfortunately, insufficient electron density for this portion of both the *Mm*IgG1Fc and *Mm*IgG2aFc structures meant that these residues could not be modelled and as such, could not be analysed for important structural differences.

Residues unique to the *Mm*IgG2a amino acid sequence are clustered in the CH3 domain with the exception of three residues in the CH2 domain. These residues are Met258, which is a lysine in all other subclasses and residues Leu300 and Arg326, which are conservative mutations where Ile and Lys, respectively, are present at those positions in other subclasses. A number of mutations also fall at the CH2-CH3 interface of *Mm*IgG2aFc. All of the identified residues are shown modelled onto the *Mm*IgG2aFc structure in Figure 6.19.

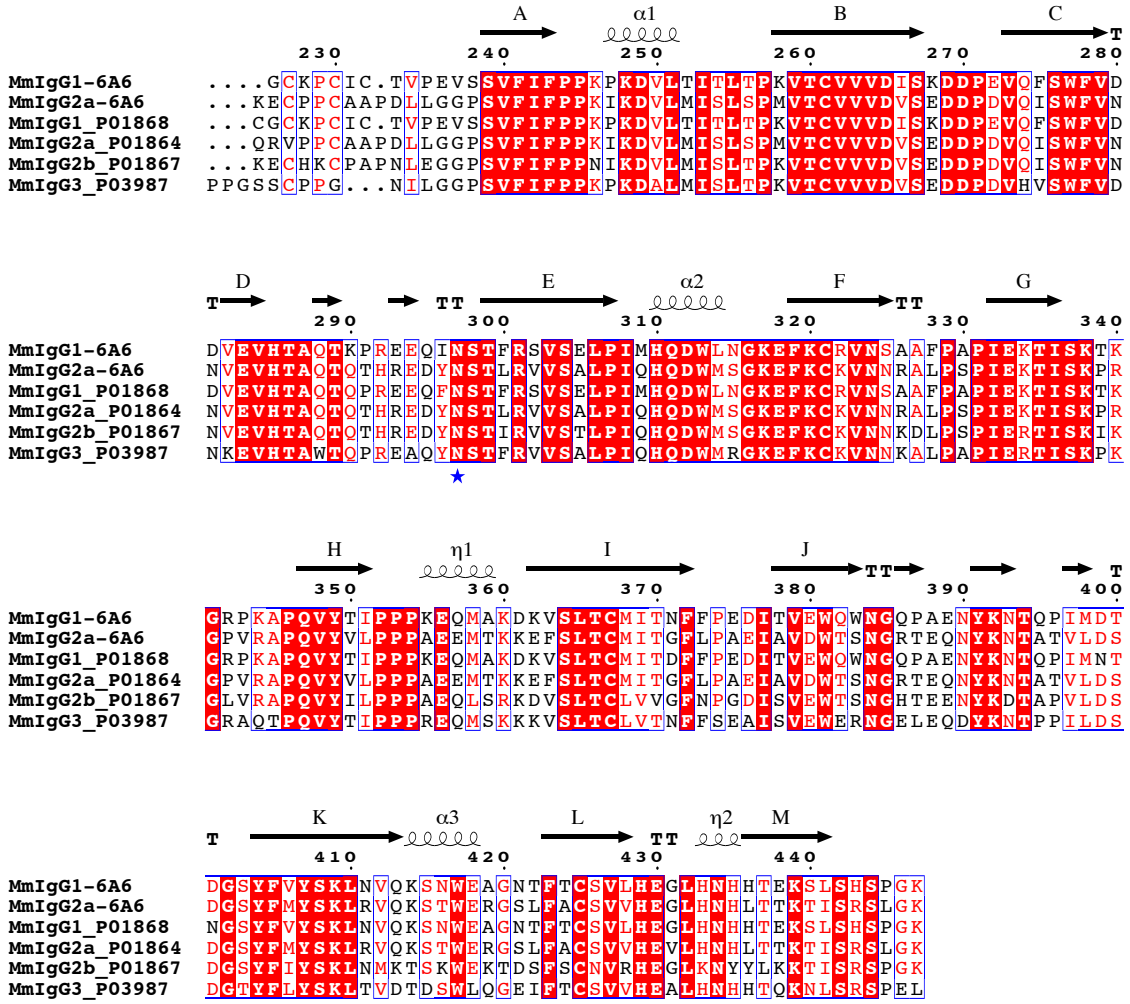


Figure 6.17: Multiple sequence alignment of the Fc domains from the IgG subclasses of *M. musculus*. The alignment was made using ESPript 3.0 [28]. The sequences of *Mm*IgG1Fc (*Mm*IgG1-6A6) and *Mm*IgG2a (*Mm*IgG2a-6A6) derive from the 6A6 antibodies described and used for crystallographic studies in this chapter. Generic sequences for the the Fc region of the major subclasses of IgG were taken from UniProt. The sequences for IgG1, IgG2a, IgG2b and IgG3 are listed with their respective UniProt number. Conserved residues are shown in white text and are highlighted in red. Non-conserved residues are shown in black text on a white background. Residues in red text indicate conservative changes or partial conservation across different subclasses. The position of the *N*-linked glycan is indicated by a blue star at position 297. The secondary structure elements listed above the alignment derive from the crystal structure of human IgG1 Fc (PDB ID: 3AVE). β-sheets are indicated as arrows, α-helices are shown as coils (α) as are the 3_{10}-helices (η). TT indicates a strict β turn.



Figure 6.18: Multiple sequence alignment of the Fc domains from the IgG subclasses of *Homo sapiens*. The alignment was made using ESPript 3.0 [28]. Generic sequences for the the Fc region of the four subclasses of IgG were taken from UniProt. Each subclass, IgG1, IgG2, IgG3 and IgG4 is listed with its UniProt number. Conserved residues are shown in white text and are highlighted in red. Non-conserved residues are shown in black text on a white background. Residues in red text indicate conservative changes or partial conservation across different subclasses. The position of the *N*-linked glycan is indicated by a blue star at position 297. The secondary structure elements listed above the alignment derive from the crystal structure of human IgG1 Fc (PDB ID: 3AVE). β-sheets are indicated as arrows, α-helices are shown as coils (α) as are the 3_{10}-helices (η). TT indicates a strict β turn.

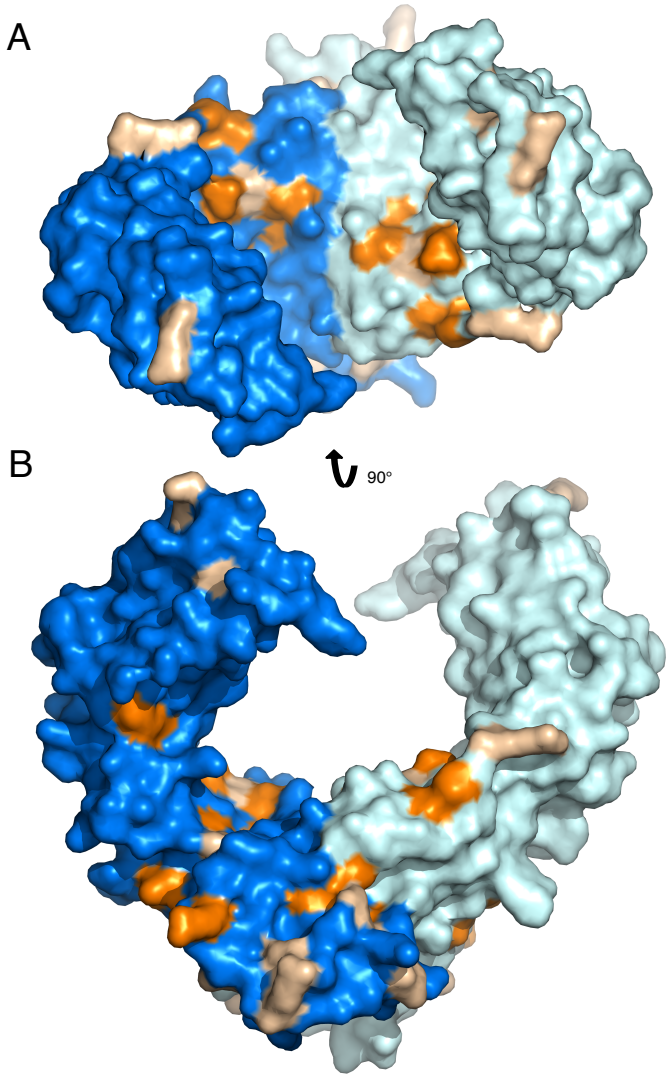


Figure 6.19: Mapping residues unique to the *Mm*IgG2aFc amino acid sequence. The structure of the *Mm*IgG2aFc BC dimer is shown in surface representation. The B chain is shown in blue and the C chain in light cyan. (A) A top-down view of the dimer and (B) a side-view of the dimer are shown. Residues unique to *Mm*IgG2aFc identified by a Clustal Omega sequence based alignment are shown in orange. Residues that are unique to *Mm*IgG2aFc, but are conservative mutations of residues found in other murine IgG subclasses are shown in pale orange.

6.3 Discussion

The aim of this work was to gain further insight into the structural basis for the loss of Fc:FcγR binding upon EndoS deglycosylation of Fcs. Understanding the structural changes responsible for Fc deactivation by EndoS is of use for the design of EndoS resistant antibodies to be used in Fc receptor-refocusing strategies, as described by Baruah *et al.* [142]. To understand changes caused by EndoS deglycosylation, the behaviour of Fc and Fc^{GlcNAc} in solution was analysed using SAXS. In addition, the X-ray crystal structures of *Mm*IgG1Fc and *Mm*IgG2aFc were solved with an aim of identifying key structural features of *Mm*IgG2aFc which allow it to maintain FcγR binding upon EndoS deglycosylation.

SAXS experiments which compared the behaviour of Fc and Fc^{GlcNAc} in solution identified several differences between these two glycosylation states. In general, Fc^{GlcNAc} gave parameters consistent with a protein with a smaller mass than Fc. These included the $I(0)$, the Porod volume and thus the calculated molecular weight. This is consistent with the loss of the glycan through EndoS hydrolysis. Despite a reduction in the calculated volume of Fc^{GlcNAc} compared to Fc (74,900 Å^3 vs. 86,100 Å^3), the R_g values for both proteins were similar. This suggests that despite the loss of the *N*-linked glycan, Fc^{GlcNAc} may still be maintaining the overall shape of the glycosylated Fc.

Although the R_g indicated that the overall shape of the particle may not be changing significantly upon deglycosylation, there was a difference in the shape of the $P(r)$ curves between Fc and Fc^{GlcNAc}. The loss of the glycan may be contributing to the changes in the $P(r)$ simply from the perspective that in Fc, the population of interatomic vectors which involved atoms of the glycans are no longer present in Fc^{GlcNAc}. Alternatively, or in conjunction with this, it may be that there are a subset of conformations for glycosylated Fc that are no longer explored by Fc^{GlcNAc}, or that there is a general increase in the flexibility of the Fc dimer. To put this into the context of receptor binding, it is possible that the now unexplored conformations are those which are more amenable to receptor binding or, that due to increased flexibility, receptor binding conformations are now less frequently explored. The dimensionless Kratky plot for Fc^{GlcNAc} showed differences at high q values between these two proteins. Much like the $P(r)$, the changes suggested that the two proteins differ in their behaviour and conformation in solution independent of their size and concentration.

The results of these experiments were largely consistent with the observations made by Borrok *et al.* [41] for the aglycosylated human IgG Fc protein. They, too, observed a similar change in the *P(r)* function between the glycosylated and aglycosylated structures, suggesting that the behaviour of EndoS-deglycosylated Fc in solution is comparable with that of aglycosylated Fc. In line with the conclusions presented by Borrok *et al.*, these results suggest that EndoS-mediated deglycosylation may be causing an increase in the flexibility of the Fc dimer. This conclusion is also consistent with the results of a study by Kelliher *et al.* [92] which utilised single molecule FRET to assess the effect of deglycosylation on Fc structure. Kelliher *et al.* showed that removing the glycan from Fc using PNGaseF resulted in changes to the energy transfer efficiency peaks that were consistent with an increase in the flexibility of Fc upon deglycosylation. The results of the SAXS analyses presented here suggest that the poor affinity of Fc^{GlcNAc} for FcγRs is potentially influenced by fewer favourable receptor-binding conformations being explored due to the increased flexibility induced by deglycosylation of the Fc. In addition, these results suggest that the overall quaternary structure of Fc in solution as well as local tertiary structure is important in establishing FcγR binding.

X-ray crystal structures for both *Mm*IgG1Fc and *Mm*IgG2aFc were determined and were used in an effort to determine the features of *Mm*IgG2aFc that allow it to maintain FcγR binding affinity even when treated with EndoS. These structures were put into the context of the current structural library of IgG Fc structures to try and identify unique tertiary and quaternary structural features of *Mm*IgG2aFc.

With regards to tertiary structure of the CH2 loops, significant structural variability amongst Fc structures was observed. Importantly, the conformations of the CH2 loops were not conserved across glycosylated Fcs, all of which have the propensity to bind FcγR receptors. Given the considerable structural variability in these loops of glycosylated Fc structures, all of which are capable of binding FcγRs (Fig 6.16 blue loops) it is proposed that assessing the deviations or differences of these loops is unlikely to offer any definitive reasons for the FcγR-binding properties of Fcs. Electron density for these loops is often poor and and the residues within these loops often have high *B* factors. This suggests that these loops are flexible or have multiple conformations in solution, further consolidating the proposal that strong conclusions about FcγR binding should not be placed on the

conformation of these loops observed in X-ray crystal structures.

The variation in quaternary structure of IgG Fc domains is summarised in Figure 6.15. The relative position of the CH3 domains is relatively constant and is likely to be a result of the the many protein–protein contacts which stabilise this dimer interface. This analysis also illustrated the low variability of the CH2-CH2 distance found in FcγR-bound structures. This distance is a result of the steric constraints caused by FcγR binding to crucial contacts in the CH2 loops and lower-hinge region (as reviewed in [27]). In contrast, the position of the CH2 domains in structures of unbound Fcs varies for both glycosylated and de- or aglycosylated structures, a point which is further discussed below.

It has previously been proposed that the collapsed conformations of the CH2 domains of de- and aglycosylated Fc structures are responsible for their inability to bind FcγR [41]. However, the inter-CH2 distance of Fc structures is liable to be influenced by crystal contacts and for the reasons described below, is not likely to be a good measure upon which to base conclusions about Fc functionality or binding capability. Previously, structures of de- or aglycosylated Fcs determined by X-ray crystallography have consistently been determined to reveal more 'closed' conformations than their glycosylated counterparts. However, the *Mm*IgG2aFc structure shows that deglycosylated Fcs can extend to this open conformation, albeit influenced by crystal contacts. This is further exemplified by another recent deglycosylated Fc structure of IgG4 where the the CH2 domains also form an 'open' conformation due to interlinked CH2 domains with neighbouring copies in the asymmetric unit (PDB ID: 4D2N, [330]). Finally, an open structure of a deglycosylated IgG1 Fc has also been determined (PDB ID: 3DNK) showing that these 'open' Fc conformations are observed for IgG1 Fc as well as for alternative IgG subclasses. In addition, inspection of a recent structure of human IgG1 Fc carrying a large, sialylated complex biantennary glycan, reveals that one of the dimers in the asymmetric unit forms a 'closed' structure within the crystal. This demonstrates that glycosylated Fcs can also form these collapsed conformations. The position of the CH2 domains is crucial in FγR binding, but the states captured in the crystal structures are just that. They are a reflection of possible states. Considering the structural variability as seen in Figure 6.15 for the glycosylated, free (non-receptor bound) structures, which all have the propensity to bind receptors, the inter-CH2 distance captured in crystal structures is unlikely to offer a metric of receptor-

binding capability or affinity.

It is, however, likely to be important which conformational states of the CH2 domain, with respect to the CH3 (*i.e.* the angle between the CH2 and CH3 domains), are likely to be explored by Fc. As discussed earlier, it may be that loss of the *N*-glycan at position 297 changes the conformational space the the Fc able to sample. This is likely a result of the loss of important stabilising interactions between the glycan-CH2 interface, as has been explored with previous biophysical experiments [89, 331]. There are likely to be many small changes to protein structure, including an increase in flexibility of the D/E loop, which steer deglycosylated Fcs away from favourable Fc binding conformations.

A multiple alignment of sequences from all four murine IgG subclasses was carried out to identify areas of sequence unique to *Mm*IgG2aFc which might be responsible for its FcγR binding behaviour. The main source of variation across IgG subclasses in both human and mice was identified in the lower-hinge region, which is directly involved in receptor binding interactions. Unfortunately, neither the *Mm*IgG1 or *Mm*IgG2a structure gave resolved electron density for the lower-hinge region so no comparisons of these regions could be carried out. Aside from the hinge region, the majority of mutations were isolated to the CH3 domain, with several at the CH2-CH3 domain interface. Mutations at the CH2-CH3 interface have previously been shown to influence receptor binding [62]. Notably, aglycosylated IgGs with specific mutations at this interface and in the CH3 domain are able to retain FcγR binding affinity [95, 96].

Jung *et al.* (2010) demonstrated that the amino acid substitutions E382V and M428I were able to restore binding of aglycosylated IgG1 Fc to FcγRI. Although the glycosylated form of this mutant was able to bind to all FcγRs, binding of the aglycosylated mutant to FcγRs other than FcγRI was not detected [95]. In a further publication by Jung *et al.*, receptor binding of aglycosylated IgG1 Fc was able to be restored for all FcγRs through mutation of 5–7 residues, however, the binding affinities to FcγRs were significantly different to that of wild-type IgG1 Fc where binding to FcγRIIa was notably 163-fold higher [96]. From this work, the authors suggested that restorative mutations may stabilise receptor-specific binding conformations. In line with this, the plethora of residues unique to *Mm*IgG2a may contribute to *Mm*IgG2aFc's maintenance of FcγR binding upon deglycosylation by helping *Mm*IgG2a maintain the ability to sample quaternary CH2 domain conformations

that are more conducive to receptor binding. Analysis of the IgG sequence alignments (Fig. 6.18 and Fig. 6.17) revealed that none of the specific mutations described above, which restore receptor binding in aglycosylated IgG, are found in *Mm*IgG2aFc. Therefore, the effect seen in *Mm*IgG2aFc is likely to occur via a different combination of amino acid substitutions.

For future studies, the structure of deglycosylated *Mm*IgG2aFc in complex with either of the activating receptors FcγRI or FcγRIIIa would be of use in ascertaining the crucial features of mouse IgG2a which permit binding in the absence of the *N*-glycan. Ultimately, the primary *Mm*IgG2aFc sequence must allow it to compensate for the loss of glycan-protein interactions that usually stabilise the Fc into conformations which favour receptor binding. This is likely to result from a global effect caused by many small interactions over the entire domain. It is apparent from this work that these many interactions are too convoluted to be identified from a single crystal structure, nevertheless, a general understanding of what factors might be influencing *Mm*IgG2aFc behaviour were still identified.

In conclusion, the results presented here indicated that the entire quaternary structure of the Fc appears to be affected by EndoS deglycosylation. The loss of the the *N*-linked glycan results in a reduction in affinity to FcγRs which may, in part, be attributed to a reduced sampling of favourable receptor-binding conformations.

7

Structural characterisation of IdeS in complex with IgG1 Fc

IdeS is a cysteine protease produced by GAS which is specific for the hinge region of IgG. IdeS has shown potential as a treatment for autoimmune diseases and is currently in clinical trials as an immunosuppresant for organ transplant patients. Given the possible therapeutic applications of this enzyme, a better understanding of its interaction with and specificity for IgG was sought. The aim of these experiments was to produce a structural model of IdeS in complex with IgG Fc. Both X-ray crystallography and SAXS were pursued as methods to produce this model. X-ray crystallography experiments yielded crystals, however, these crystals were not of sufficient quality for structure determination using X-ray diffraction data. SAXS was used to derive an overall envelope of the IdeS:Fc complex and attempts to model the X-ray crystal structures of each component into this envelope are presented here. The experiments described here gave results consistent with

the formation of a 1:1 complex between IdeS and IgG1 Fc, however, further experiments are required to help determine crucial contacts and interface residues between these binding partners.

7.1 Preparation of the IdeS-Fc complex

X-ray crystallography has the potential to describe the atomic level detail of protein structure, therefore this method was employed to study the complex of IdeS and Fc. As described in Chapter 1, multiple X-ray crystal structures have been published for both IdeS and IgG Fc in isolation and although successful crystallisation of the individual components does not guarantee that the complex will be equally crystallisable, the ability of these proteins to form ordered crystals provided further rationale for using this method to study the complex.

7.1.1 Expression and purification of IdeS

X-ray crystallography often requires milligram amounts of pure protein for crystal growth, therefore, protocols were developed for the recombinant expression and purification of both IdeS and IgG1 Fc using *E. coli* cells and mammalian HEK 293T cells, respectively. The protocol for IgG1 Fc expression and purification is described in Chapter 3 so is not repeated here. Briefly, the Fc construct was designed to contain the lower hinge region including a single cysteine residue which allows formation of a disulphide bond between heavy chains. This lower hinge included the protease site for IdeS which falls between Gly236 and Gly237. Initial work on IdeS expression and purification was carried out by Mr Benjamin Krishna and Mrs Susan Tickner (Glycobiology Institute, Department of Biochemistry). In addition, Mrs Susan Tickner assisted with initial experiments for formation of the IdeS:Fc complex. This initial work utilised an IdeS protein from GAS serotype M59 which had been selected from a pool of IdeS sequences from UniProt. This construct also included the secretion signal sequence of IdeS.

All previous crystal structures of IdeS in isolation utilised the IdeS sequence from GAS serotype M1 [156, 155]. Given the successful crystallisation of the M1 IdeS protein, additional constructs were made using this M1 sequence. This was done considering that

the sequence of IdeS from serotypes M1 and M59 share only 85% identity. An analysis by Lei *et al.* showed that IdeS identified across various GAS serotypes can be classified into one of two classes (classes I and II) based on sequence identity [160]. IdeS from serotype M1 falls into class I and IdeS from M59 falls into class II. The sequences of M1 and M59 differ primarily in the middle third of the protein which can be seen in the alignment of the M1 and M59 sequences in Figure 7.1. This was expected based on the definition of this classification system [160].

Further constructs were made for both the M59 and M1 sequences where the first 38 residues of IdeS (which includes the signal sequence from residues 1–29) were truncated from the N-terminus to give constructs which matched the previously crystallised construct from Wenig *et al.* [156]. These shorter constructs are referred to as IdeS-SS and the full-length constructs are referred to as IdeSFL.

In order to produce a complex of IdeS and Fc whereby IdeS does not proteolyse the Fc hinge region, a catalytically inactive form of IdeS was required. The nucleophilic cysteine residue at position 94 of all constructs of IdeS was mutated to alanine (and serine in the case of the M1 constructs). Mutation of this residue has previously been shown to abrogate the proteolytic activity of IdeS [153]. Over-expression of the M59 and M1 proteins as both IdeSFL and IdeS-SS constructs was carried out using an *E. coli* expression system. The successful over-expression of these constructs as C-terminal hexahistidine fusions, as confirmed by western blot, is shown in Figure 7.2. In general, the IdeS-SS constructs showed higher levels of over-expression than the IdeSFL constructs.

All IdeS constructs used in this chapter were purified following an identical purification protocol which utilised an IMAC step followed by SEC. Example results of these purification steps are shown for the catalytically inactive mutants IdeSFL-M59^{C94A} and IdeS-SS-M59^{C94A} in Figure 7.3.

Similarly to the EndoS-Fc complex described in Chapter 3, analytical FPLC was used to determine whether a stable IdeS-Fc complex could be produced. This was initially carried out using a catalytically inactive form of the IdeSFL-M59 construct and was done in collaboration with Mrs Susan Tickner (Glycobiology Institute, University of Oxford). Purified IdeSFL-M59^{C94A} was applied to a Superdex200 10/300 GL at room temperature and was eluted with 10 mM HEPES pH 8.0, 150 mM NaCl. The same was repeated for

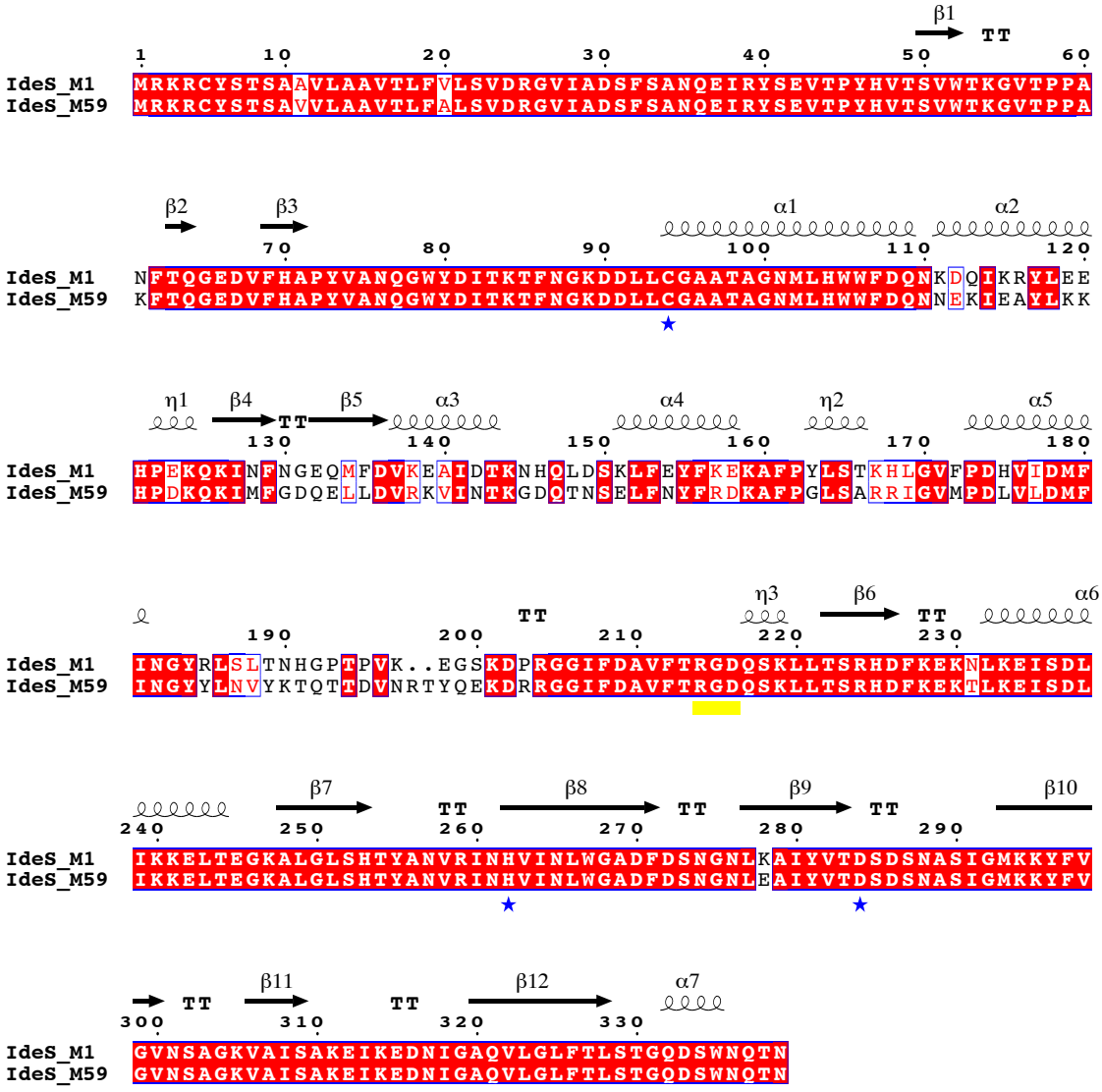


Figure 7.1: Multiple alignment of IdeS from GAS serotypes M1 and M59. The alignment of IdeS sequences from GAS serotype M1 (UniProt ID: Q7DAM2) and M59 (H8HA37) was made using ESPript 3.0 [28] and a multiple alignment file generated using Clustal Omega [29]. Conserved residues are shown in white text and are highlighted in red. Non-conserved residues are shown in black text on a white background. Residues in red text indicate conservative changes or partial conservation between the two sequences. The secondary structure described above the alignment is derived from PDB ID: 2AU1 [155]. The residues of the catalytic triad are indicated by blue stars (Cys94, His262, Asp284) and the position of the RGD motif is highlighted yellow.

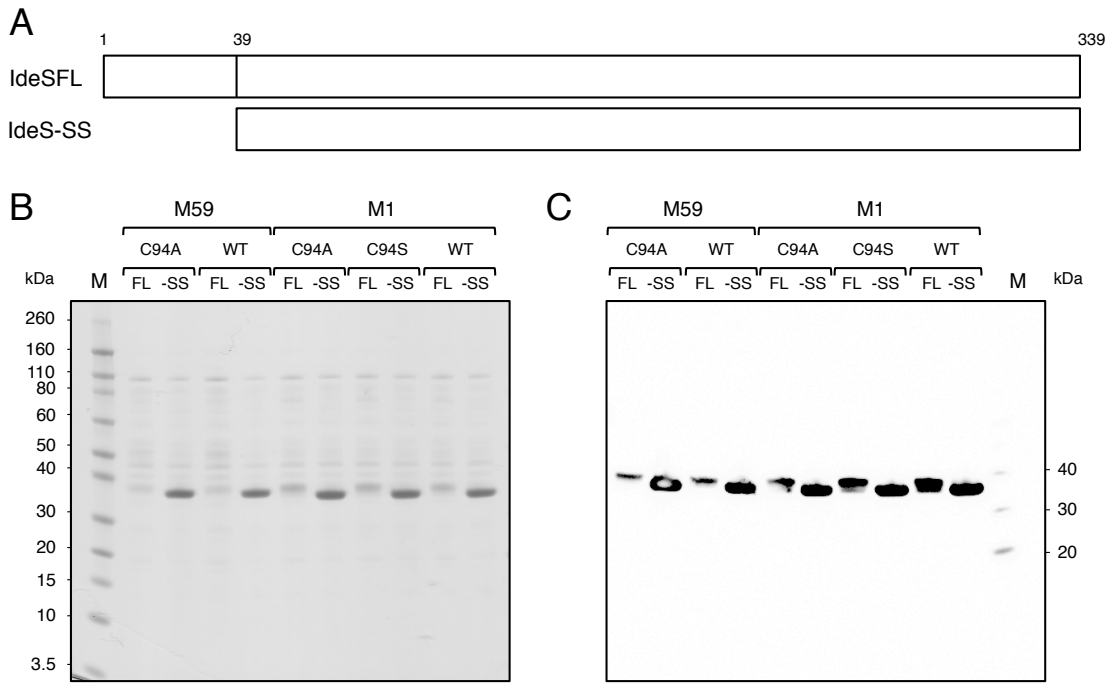


Figure 7.2: *E. coli* over-expression of IdeS. (A) A schematic of the expressed proteins shows the construct boundaries for IdeSFL and IdeS-SS. (B) An SDS-PAGE gel and (C) an anti-His western blot are shown for cell lysate fractions from *E. coli* cells capable of expressing the IdeS constructs. Lysates are grouped according to the serotype of IdeS expressed (M59 or M1). 'M', molecular weight marker; 'FL' full-length; '-SS', no signal sequence; 'WT', wild-type.

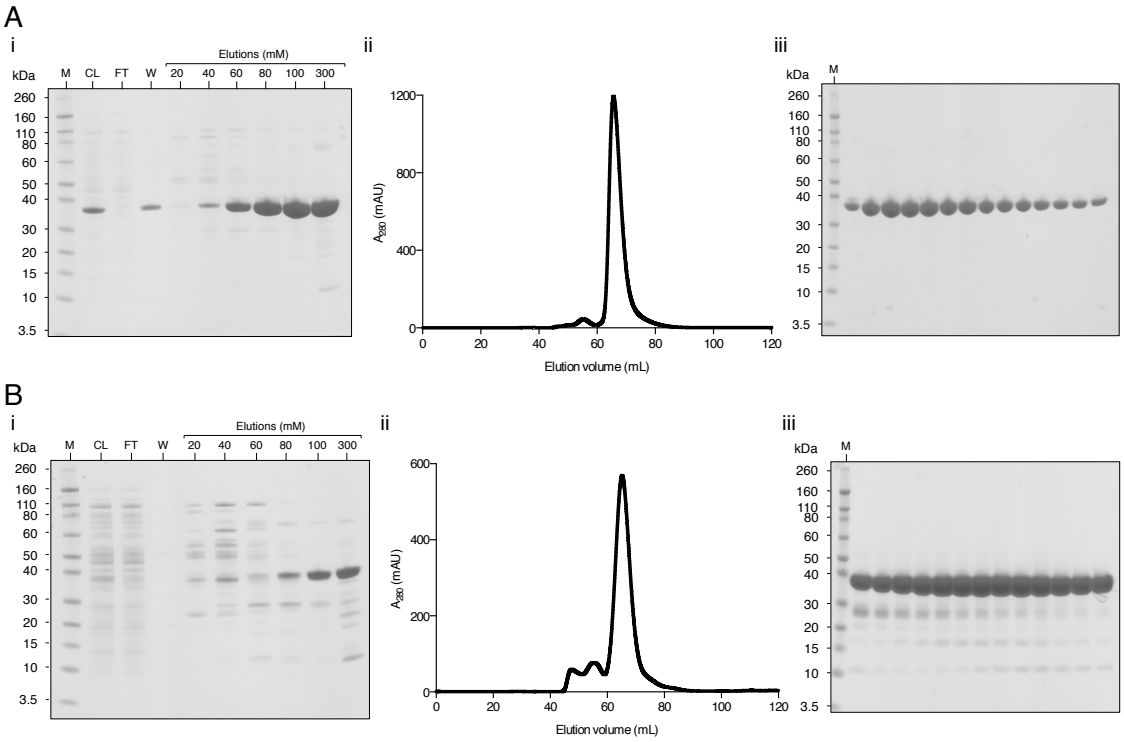


Figure 7.3: Purification of (A) IdeS-SS-M59 and (B) IdeSFL-M59 using IMAC and SEC. (i) SDS-PAGE analysis of fractions taken throughout IMAC purification show: 'CL', the total cell lysate from *E. coli* cells expressing a C-terminal hexahistidine tagged IdeS fusion; 'FT', the flow-through from the HisPur^{TM} Ni-NTA column; 'W', the wash fraction from the HisPur^{TM} Ni-NTA column; 'Elutions', the eluted fractions using PBS with increasing concentrations of imidazole in a stepwise gradient. Imidazole concentrations used for each step are denoted above each lane in mM. (ii) SEC was carried out using a HiPrep Superdex75 16/60 column where the major peaks correspond to the IdeS protein. (iii) SDS-PAGE analyses of the pooled fractions from the SEC peaks indicate the molecular weight and purity of the samples.

a sample of Fc and the elution volume corresponding to the inflection point of the A_{280} peak was noted for each protein. IdeSFL-M59^{C94A} eluted at 16.60 mL and Fc eluted at 15.13 mL. IdeSFL-M59^{C94A} and Fc were combined at a 1:1 molar ratio. This mixture was then loaded directly onto the Superdex200 10/300 GL column. The protein mixture eluted predominantly as one species where the inflection point of the major peak was seen at an elution volume of 14.35 mL. SDS-PAGE analysis of the major peak revealed that the peak contained both IdeSFL-M59^{C94A} and Fc. The elution volume of this peak was consistent with the formation of a 1:1 complex of IdeS:Fc.

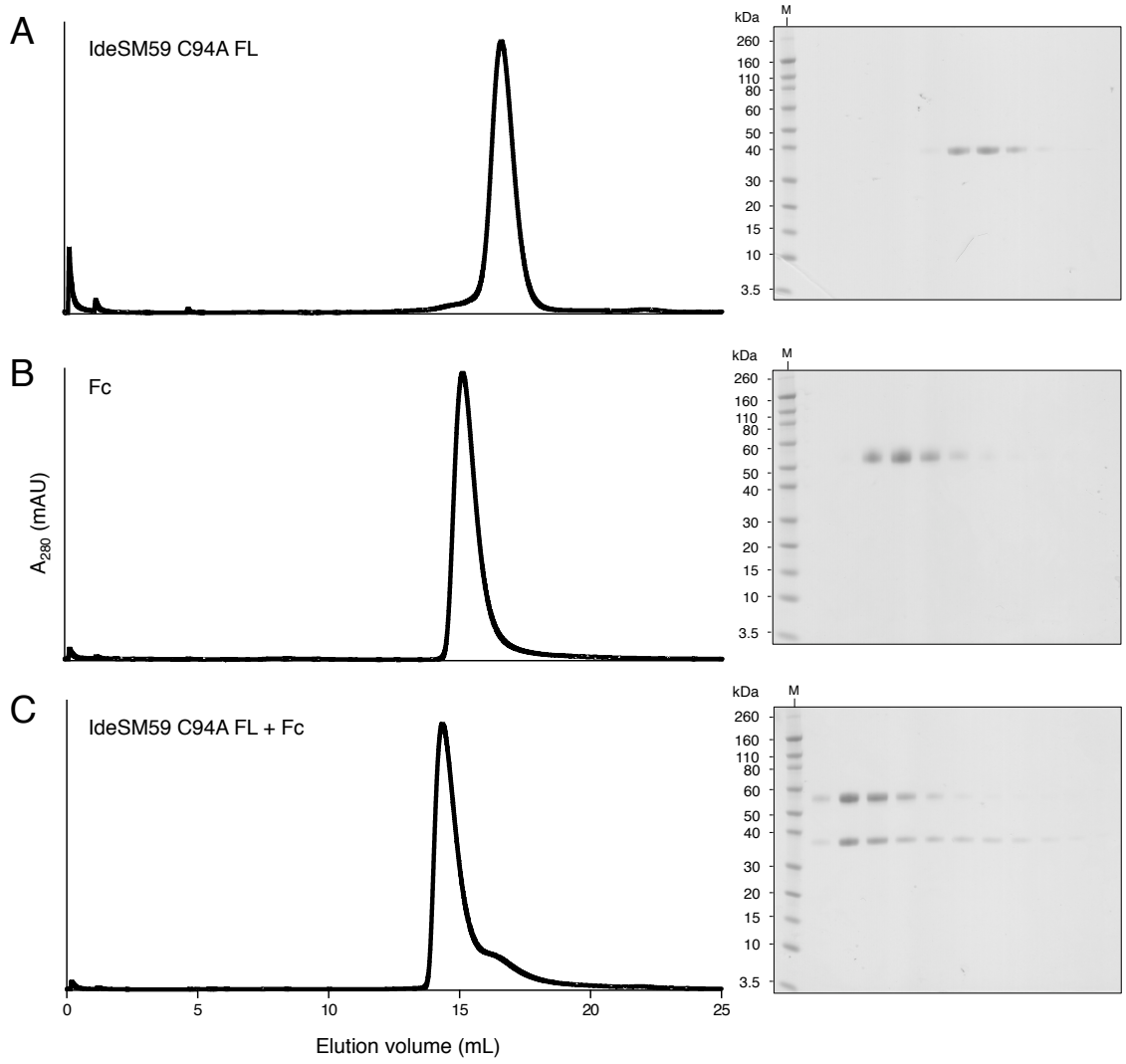


Figure 7.4: Analytical SEC of IdeSFL-M59 and Fc. The analytical SEC chromatograms and corresponding SDS-PAGE gels are shown for (A) IdeSFL-M59, (B) Fc and (C) a 1:1 molar ratio of IdeSFL-M59 and Fc. Each protein was subjected to SEC using a Superdex200 10/300 GL column and was eluted using 10 mM HEPES pH 8.0, 150 mM NaCl. SDS-PAGE gels were run under non-reducing conditions to preserve the integrity of the Fc homodimer.

7.2 X-ray crystallography

Large scale preparations of complexes of Fc with IdeSFL-M59, with IdeS-SS-M59 or with IdeS-SS-M1 were carried out as described above for use in crystallisation trials. SDS-PAGE analyses of the purified complexes are shown in Figure 7.5.

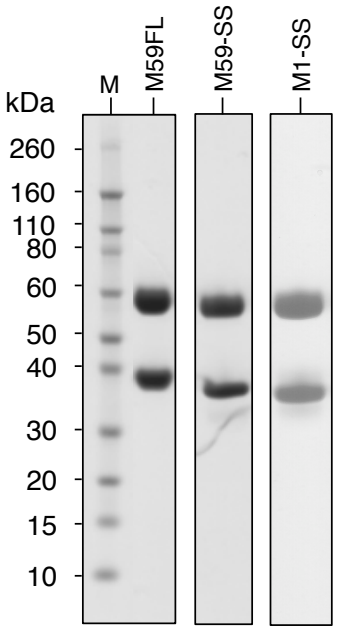


Figure 7.5: SDS-PAGE analysis of IdesFc complexes. Purified complexes of IdeSFL-M59:Fc, IdeS-SS-M59:Fc and IdeS-SS-M1:Fc were analysed by SDS-PAGE to check both the purity of the complex and that both components were present in the SEC elution.

7.2.1 Crystallisation

Crystallisation trials were carried out for each protein complex at both 4 °C and 18 °C. Protein concentration was also varied in an effort to identify suitable crystallisation conditions for the IdeS-Fc complexes (A_{280} = 5, 7.5, 8.5, 10, 10.5). Crystallisation trials of the IdeS-Fc complexes gave numerous crystal hits, however, hit conditions tended to show extensive nucleation with many small crystals in the crystallisation well.

Although larger crystals were also grown from the crystallisation trials, X-ray diffraction analysis of a number of these crystals revealed them to be crystals containing Fc only (data not shown). X-ray diffraction data collected from such crystals were integrated and scaled using the programmes XDS [288] and CCP4-Aimless [289], run using the data processing suite autoPROC[225]. For structure determination, molecular replacement

using the programme Phaser [227] was carried out using fragments of the glycosylated human IgG1 Fc as search models (PDB ID: 3AVE [46]). In such cases, attempts to place a copy of IdeS (PDB ID: 2AVW [155]) in the asymmetric unit were unsuccessful. Given that these crystals did not contain IdeS, these structures were not further refined and as such, are not described here.

7.2.2 Diffraction data collection and processing

Putative crystals of a complex of IdeS-Fc were grown from crystallisation plates containing the purified IdeSFL-M59-Fc complex. A shower of micro crystals was observed in a condition from the Proplex HT [332] screen containing 0.15 M ammonium sulphate, 15% w/v PEG 4000, 0.1 M MES pH 6.0 at 20 °C. Crystals appeared within 24 hours but did not show a significant increase in size upon further incubation.

Optimisation of this condition was pursued in an effort to produce larger crystals suitable for X-ray diffraction studies. Optimisation experiments using seeding techniques as well as additive screens were carried out, however, the most successful strategy used to grow larger crystals was to alter the concentration of the precipitant solutions and the pH. The original crystals and the optimised crystals from these experiments are shown in Figure 7.6. Prior to X-ray diffraction studies, crystals were cryocooled in an effort to reduce potential radiation damage caused by synchrotron radiation [270]. Using solutions of either PEG 400, glycerol or ethylene glycol crystals were cryoprotected before being cryocooled by immersion into liquid nitrogen.

The best dataset from these putative IdeS-Fc crystals was collected from a crystal which was cryoprotected with 25% v/v glycerol (Fig. 7.7). However, X-ray diffraction images taken from these crystals produced only low resolution reflections. Also, diffraction images for certain orientations of the crystal showed poor quality reflections which smeared into one another. Attempts to index and scale these images was carried out with extensive help from both Dr Pietro Roversi (Glycobiology Institute, Department of Biochemistry) and Dr Clemens Vonrhein (Global Phasing Ltd.). Indexing was not possible due to severe lattice-translocation disorder of the crystal which was apparent in the diffraction images (for a discussion of this crystal disorder phenomena, see [333]).

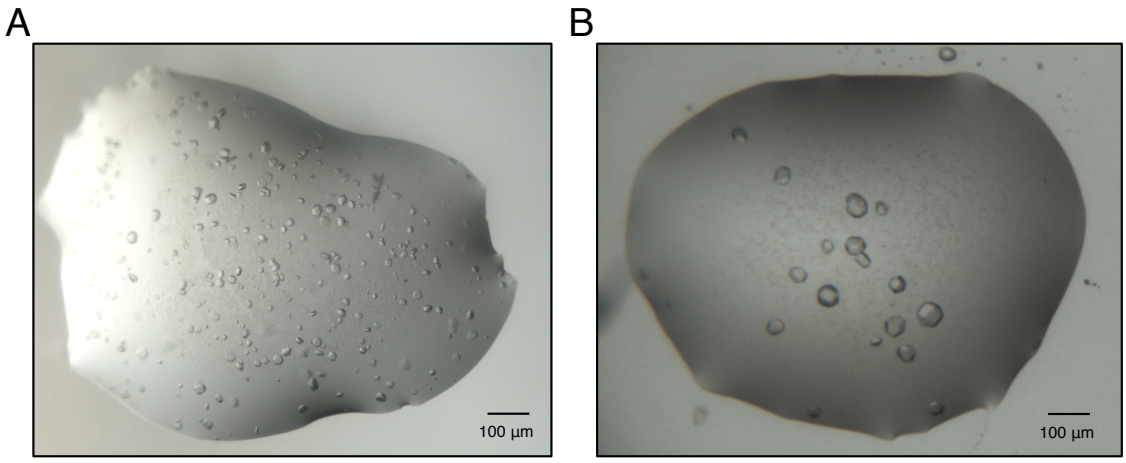


Figure 7.6: Putative crystals of an IdeSFL-M59:Fc complex. (A) An initial crystallisation hit was observed from Proplex HT [332] condition B10 (0.15 M ammonium sulphate, 15% w/v PEG 4000, 0.1 M MES pH 6.0). (B) Optimisation of the crystallisation condition resulted in growth of larger crystals suitable for X-ray diffraction studies. The optimised condition contained 0.1 M ammonium sulphate, 15% w/v PEG 4000, 0.1 M MES pH 5.8.

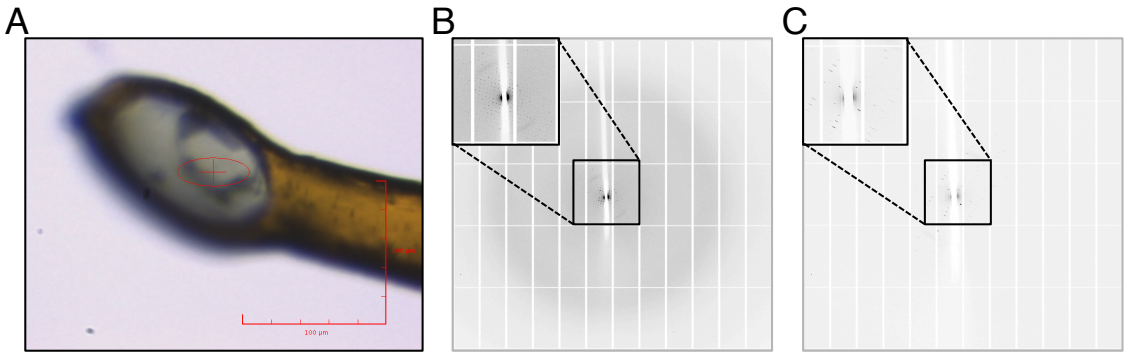


Figure 7.7: X-ray diffraction data from putative crystals of an IdeSFL-M59:Fc complex. (A) A single crystal from the optimised crystallisation condition shown in Figure 7.6B was looped using a and is pictured mounted on the goniometer at DLS beamline I03. (B) X-ray diffraction images for this crystal reveal low resolution reflections and (C) the smeared reflections observed for particular orientations of the crystal.

7.3 SAXS

In addition to X-ray crystallography studies, SAXS was used to analyse complexes of IdeS and Fc. In contrast to X-ray crystallography, SAXS does not require crystals of the protein of interest and can be carried out on the protein in solution. Given that crystals suitable for X-ray diffraction could not be produced, this method provided an alternative for generating structural information about the complex of IdeS and Fc.

7.3.1 Data collection and processing

SAXS data were collected for the IdeSFL-M59 protein and for the IdeSFL-M59-Fc complex at beamline P12, DESY. Data for Fc was collected previously at beamline BM29, ESRF and is described in Chapter 6. For the IdeS and IdeS-Fc complex samples, all measurements were recorded at 20 °C. In preparation for SAXS measurements, both samples were concentrated to 5 mg mL $^{-1}$ and then a 2 fold dilution series of both IdeS and the IdeS-Fc complex was made. SAXS images were then collected for each dilution in the series from short (0.05 s) exposures as the sample was flowed through a capillary perpendicular to the X-ray beam. For each dilution of the protein, 20 exposures were collected. Identical measurements of buffer only were taken immediately prior and after the protein measurement and were used for buffer subtraction purposes.

Automatic processing of SAXS images was carried out at the beamline with the programme PRIMUS [215]. For each dilution of protein in the concentration series, a curve of log scattering intensity, $I(q)$ versus q was produced from an average of all the exposures (see Section 2.6.4 for a more detailed description). The same was carried out for buffer measurements such that averaged buffer curves could then be subtracted from the protein curves to account for the scattering contribution of the solvent.

Using the programme SCÅTTER, the intensity at zero concentration ($I(0)$) was calculated by extrapolation of the measured scattering intensities. The paired-distance distribution function, *P(r)* and the real space values for R_g and $I(0)$ were also calculated using SCÅTTER. The programmes DAMMIF [271] and DAMMIN [272] were used to generate envelopes of the measured protein where the programme GNOM [273] was used to produce the input file for these methods. For a more thorough description of these programmes,

see Chapter 3, Section 3.7.1. Conversion of the DAMMIF and DAMMIN bead models into volumetric density maps was carried out using the programmes SITUS [274] and SCULPTOR [275].

A summary of the results from the SAXS analyses of IdeSFL-M59, Fc and the IdeSFL-M59-Fc complex are presented in Table 7.1 and Figure 7.8. The Guinier approximated values for $I(0)$ and R_g were similar to the real space values in all cases suggesting that these data did not show signs of radiation damage or sample aggregation. The R_g for IdeS was 24.14 Å and gave a *P(r)* distribution consistent with a small globular protein. The R_g for Fc was 29.17 Å consistent with Fc being a larger protein than IdeS. The R_g of the complex was greater still at 40.55 Å. The *P(r)* function of the IdeS-Fc complex revealed a larger proportion of interatomic vectors between 80–120 Å compared to the corresponding *P(r)* functions for both of the individual components. The *P(r)* function for IdeS-Fc was smooth and featureless consistent with a globular complex.

The DAMMIN refined models gave an indication of how both the components and the IdeS-Fc complex look in solution. The envelope for IdeS consisted of a large globular domain with a protrusion at one end. The crystal structure of IdeS was fitted to the globular part of the envelope such that the protrusion is either an experimental artefact or is a result of the scattering of the N-terminal residues present in the construct measured here that are not modelled in the crystal structure. The envelope calculated for the IdeS-Fc complex forms an elongated globular shape with minor surface features. In all cases, the theoretical curves calculated for these models correlated well with the experimental data.

Table 7.1: SAXS analysis of putative IdeSFL-M59:Fc complex.

	IdeSFL-M59	Fc	IdeSFL-M59:Fc
$I(0)^{\dagger}$	7040 (6370)	55.4 (53.0)	17700 (16300)
R_g (Å)†	24.14 (22.92)	29.17 (28.74)	40.55 (36.93)
M_w (kDa)*	33.6	58.1	84.7
D_{max} (Å)	76	101	127
Porod volume (Å^3)‡	60175	86100	168758
V_c (Å^2)‡	296.3	425.0	641.5

†Values were calculated using the Guinier approximation [217].Values in parentheses are real space values calculated using SCÅTTER [277].
*Approximated using the $I(0)$ and the $I(0)_{BSA}$.
‡The Porod volume and the Volume-of-Correlation (V_c) were calculated using SCÅTTER as described in [214].

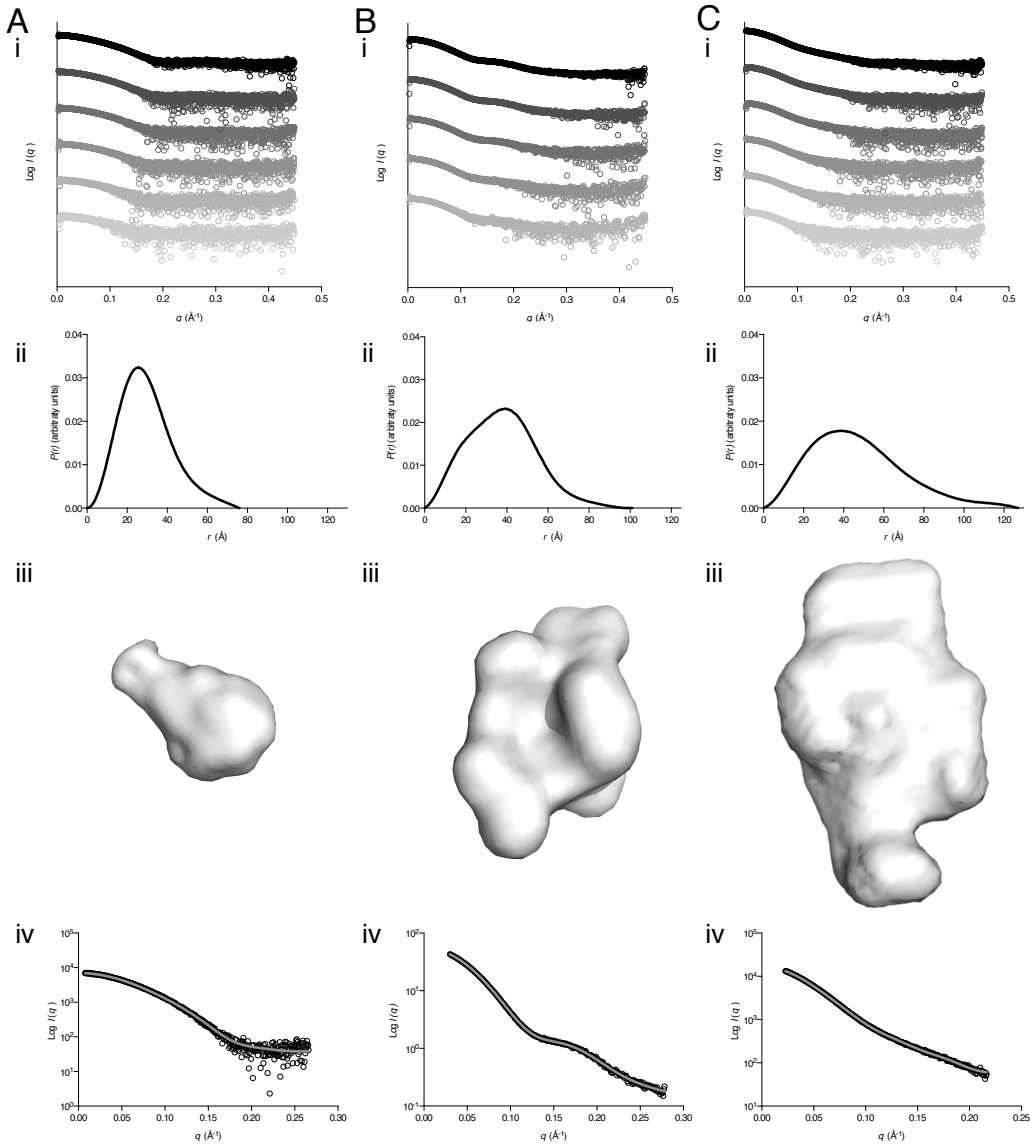


Figure 7.8: DAMMIN models of the IdeS-Fc complex. Data are depicted for (A) IdeSFL-M59^{C94A}, (B) Fc and (C) IdeSFL-M59^{C94A}-Fc. (i) The solvent corrected scattered intensities are shown for the 2-fold dilution series of each sample. The data collected from the highest protein concentration (5 mg mL^{-1}) are shown in black with the lower concentrations shown in increasingly lighter shades of grey. Each dilution is offset along the y axis by a factor of 10^3 so that all curves can be visualised on the same axes. (ii) The *P(r)* functions for the components and the IdeS-Fc complex represent a frequency distribution of interatomic vectors of length r. The curves were calculated using SCÅTTER and were normalised to the area under the curve and are presented on a common scale. (iii) The final DAMMIN refined envelopes for each sample reveal the overall shape of the components and the complex. The 23 DAMMIF envelopes used to produce the DAMMIN models gave mean NSD values of 0.714 ± 0.062, 0.696 ± 0.039 and 0.653 ± 0.062 for IdeSFL-M59^{C94A}, Fc and IdeSFL-M59^{C94A}-Fc, respectively (iv) A theoretical scattering curve was calculated for each of the models (grey) and is plotted against the scattering data (black circles).

7.3.2 Model of IdeS-Fc complex in solution

The X-ray crystal structures of IdeS (PDB ID: 2AVW [155]) and Fc (PDB ID: 3AVE [46]) were used to make various models of the encounter complex of IdeS and Fc using the programme ClusPro [282]. A description of this programme is given in Chapter 3, Section 3.7.4 where it was used to predict models of EndoS in complex with Fc. In contrast to ClusPro modelling of the EndoS-Fc complex, the IdeS-Fc complex was modelled such that Fc was treated as the rigid receptor molecule and IdeS was unconstrained. The ClusPro models were filtered using the SAXS data so that only models which were consistent with the experimental data were produced.

The results of the ClusPro docking experiment are shown in Figure 7.9. Here, the structure of Fc is shown with spheres marking the center of mass of the clusters of the docked IdeS structures. Analysis of these clusters showed that they did not colocalise to a common region of the Fc structure. The IdeS structures were found docked to the CH2-CH3 interface of both monomers, the N-terminal regions of the CH2 domains and to the C-terminal regions of the CH3 domains. The docked IdeS structure corresponding to the largest cluster (Fig. 7.9, black sphere) was bound to the outer face of one of the CH2 domains.

Previous investigations of the IdeS-Fc binding interaction revealed that FcγRs were able to block binding of IdeS to Fc, suggesting that IdeS and FcγRs share a common binding site [163]. As described in Chapter 1, Section 1.3.7, FcγRs bind to the CH2 and lower hinge regions of Fc. Therefore, the ClusPro models of IdeS in complex with Fc which also show binding to this region and would likely compete with FcγR for binding were investigated. More importantly, these models were investigated considering that the active site of IdeS must access the lower hinge residues of Fc in order to specifically cleave between residues 236 and 237.

The top ClusPro model that localised to this region is shown fitted to the DAMMIN refined envelope of the IdeS-Fc complex in Figure 7.10. The active site of IdeS is oriented towards the lower hinge region of Fc, but does not make specific contacts with Gly236 and Gly237. Some of the density at the top of the IdeS-Fc envelope is unaccounted for by the ClusPro model, but the overall shape of the envelope resembles that of the ClusPro

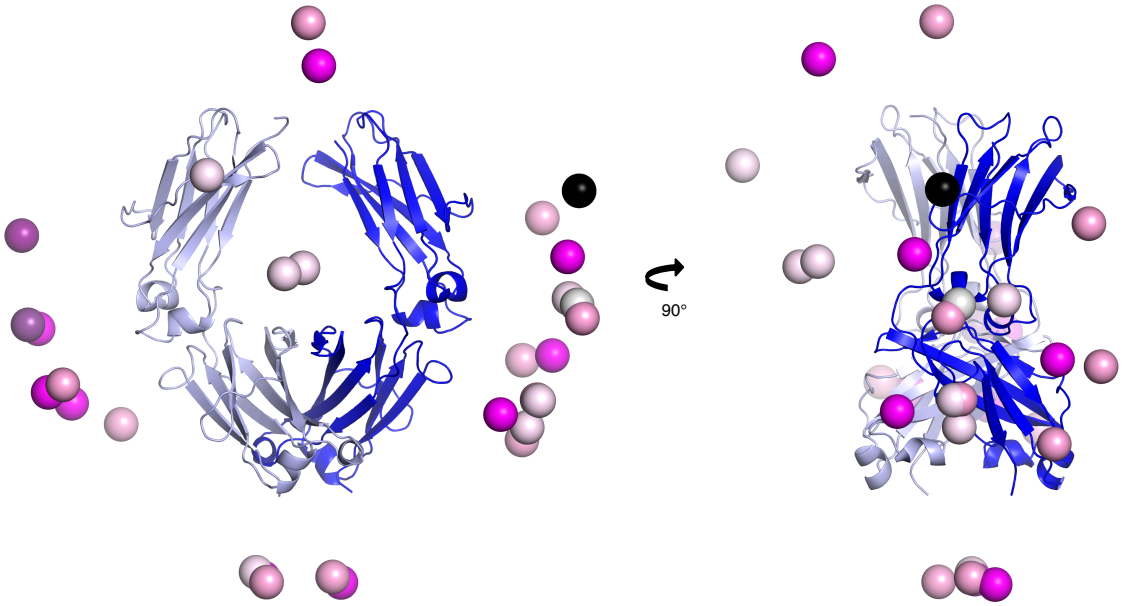


Figure 7.9: Prediction of IdeSFc complex structures using ClusPro. The crystal structure of Fc (PDB ID: 3AVE [46]) used to create these models is shown in blue in cartoon representation. The position of the centre of mass of the IdeS monomer from each ClusPro cluster is represented as a sphere. The spheres are coloured from dark purple to light pink based on the size of the cluster they represent. The largest cluster is represented by a black sphere.

complex.

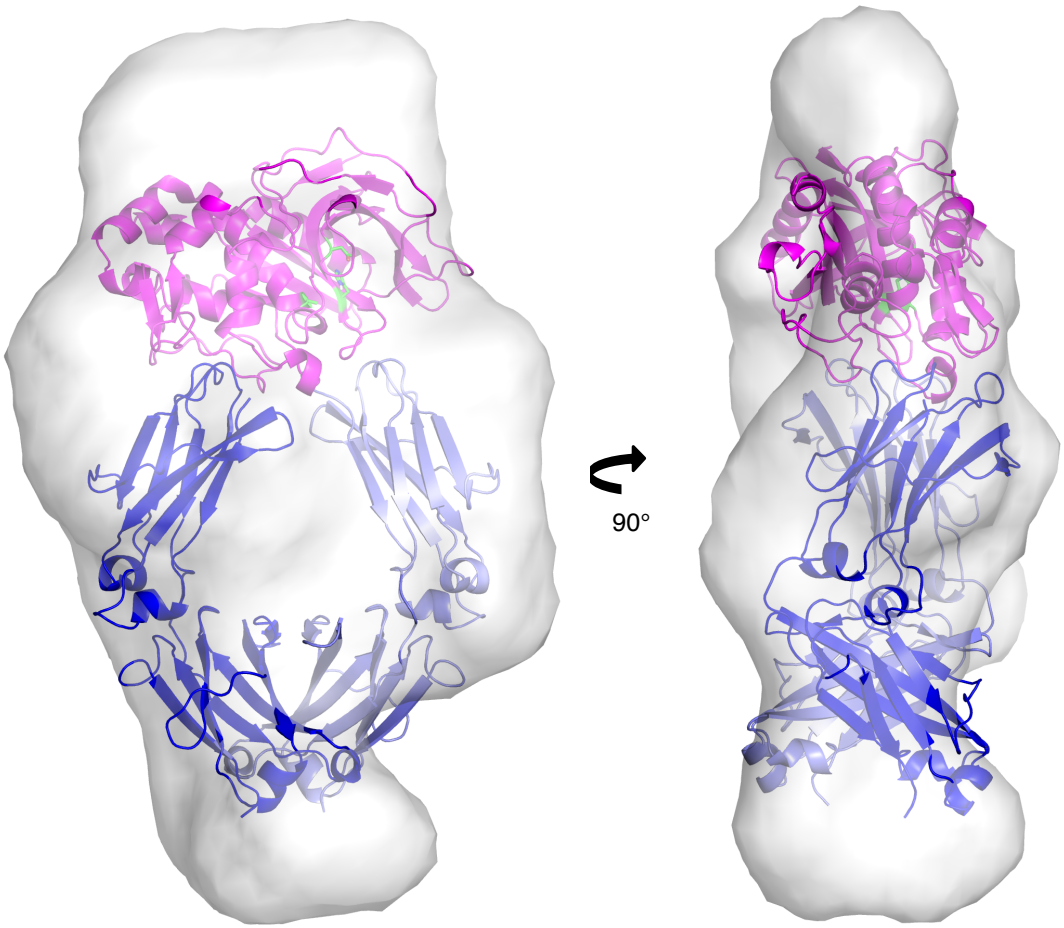


Figure 7.10: Preliminary model of the IdeS-Fc complex. One of the ClusPro-derived models of IdeS in complex with Fc was fitted to the DAMMIN refined envelope for the complex using the programme SUPCOMB. IdeS is shown in magenta with the residues of the catalytic triad shown as sticks in green. The Fc dimer is shown with one monomer coloured in blue and in the other in slate.

7.4 Discussion

IdeS is a cysteine protease which specifically cleaves the lower hinge region of IgG. Cleavage of IgG by IdeS abolishes FcγR binding capabilities and also affects binding of C1q [153]. Given its ability to alter IgG effector functions, IdeS has been investigated as a novel treatment for IgG-mediated autoimmune diseases and is currently in clinical trials as an immunosuppressant drug for organ transplant patients [193]. Considering the therapeutic potential of IdeS, further investigations into its interaction with IgG were carried out with an aim to identify key interface residues and contacts which mediate IdeS specificity to

IgG.

The data shown here are consistent with the formation of a 1:1 IdeS:Fc complex. This further supports the argument that IdeS does not function as a dimer, but is catalytically active as a monomer against the hinge region of IgG. Furthermore, considering the position of the two identical cleavage sites on the IgG lower hinge (one from each heavy chain), the proximity of these chains would likely make simultaneous binding of two monomers sterically unfavourable. IdeS cleaves the two heavy chains of IgG sequentially, one hinge chain at a time to give scIgG after the first cleavage and then $F(ab)_2$ with two Fc monomers after the second cleavage (Chapter 1, Fig. 1.14) [153]. Investigations into the kinetics of this reaction indicated that the first cleavage event was more efficient than the second and that the pool of IgG was converted almost entirely to scIgG before further cleavage into $F(ab)_2$ and Fc monomers [159]. Altogether, this information suggests that a single IdeS monomer is binding to the Fc dimer or to the scIgG fragment during catalysis. Moreover, the efficiency of this first cleavage step compared to the second suggests that IdeS may have greater affinity for the Fc dimer present in IgG compared to the asymmetrical scIgG. Based on competition assays with FcγRs, IdeS is predicted to bind in similar location to these Fc receptors [163]. This is consistent with the fact that the proteolytic site for IdeS is in the lower hinge region of IgG and that IdeS is likely to have to contact the CH2 domains in order to access this site. It may be that IdeS exhibits a similar binding mode to the FcγRs which bind asymmetrically to both CH2 domains of the Fc dimer. This could potentially explain the more efficient activity of IdeS observed against IgG compared to scIgG.

Attempts to crystallise the IdeS-Fc complex were made difficult by the selective crystallisation of Fc from crystallisation trials containing IdeS-Fc complexes. Initially it was hoped that the crystallisability of Fc would encourage crystallisation of complex of IdeS-Fc. However, this was not the case and multiple crystals which contained only Fc were grown. This selective crystallisation of Fc was more prominent in trials which utilised the IdeS-SS constructs in comparison to the longer IdeSFL-M59 construct. Alternative Fc-based substrate partners could be explored for future crystallisation trials to try and alleviate this problem. For example, the CH2 dimer containing the lower hinge region of IgG which was expressed and purified for functional studies of EndoS as described in

Chapter 5 may be of use. However, this approach may also be unsuccessful as by removing the CH3 domains of Fc, stabilising contacts between the Fc monomers would be lost. The destabilisation of the Fc domains through loss of these CH3 contacts may increase the entropic barrier to crystallisation. Cross-linking experiments could be pursued in order to lock the IdeS and Fc components together before crystallisation, however, this technique risks formation of non-specific complexes which may interfere with successful crystallisation.

The Kd of IdeS for IgG as measured by surface plasmon resonance was reported to be 2.6 μM and the Kd for Fc alone was reported as 5.6 μM. This affinity is comparable to the low affinity FcγRIIB receptor [52] which binds IgG1 with a Kd of 8.3 μM. Successful crystallisation and structure determination of this complex required mutagenesis of interface residues to increase the affinity of this interaction [61]. However, crystallisation of the FcγRIIIB with IgG1 Fc was successful using the wild-type receptor and Fc proteins [57]. The affinity of this interaction is approximately 5 μM [52] which is comparable to the published affinity of IdeS to Fc, suggesting that cocrystallisation of these complexes is feasible. Nevertheless, another approach towards crystallising the complex and avoiding selective crystallisation of Fc would be to explore mutations to Fc (or to IdeS) that confer tighter binding of the complex.

Although the models presented here provide an example of what the IdeS-Fc complex may look like, these models need to be put into context with further experimental data. To further probe the validity of these models, HDX-MS could be pursued as a method to map important interface residues between IgG Fc and IdeS (reviewed in [286]). Additional X-ray scattering experiments using contrast variation could be used to assign parts of the envelope of the complex to either Fc of IdeS. These contrast variation experiments using SAXS and small angle neutron scattering (SANS) to understand protein–protein complexes are described in a recent review by Jacques and Trewhella [212]. Binding studies to isolate residues that affect the binding affinity of IdeS to IgG using surface plasmon resonance are being pursued by a fellow student in the Glycobiology Institute, Miss Iona Easthope. Selective point mutations of the Fc CH2 and hinge regions have been made in an effort to essentially map the binding site of IdeS on IgG Fc using mutagenesis. These studies in concert with the SAXS data presented here may provide further insight

into the binding interface between Fc and IdeS such that a better understanding of IdeS specificity for IgG can be obtained.

Overall this chapter presents preliminary work towards the characterisation of IdeS in complex with Fc. Purification of IdeS-Fc complexes was made possible by using catalytically inactive IdeS proteins. Structural characterisation of these complexes using X-ray crystallography was hampered by selective crystallisation of Fc and poor quality crystals which did not give sufficient diffraction for further analysis. However, analysis of IdeS-Fc complexes using SAXS gave insight into the overall shape of these complexes. Further verification of these SAXS data using alternative methods will hopefully lead to a complete picture of how IdeS interacts with IgG Fc.

8

Conclusions

The aim of this thesis was to structurally characterise the interaction between IgG and the streptococcal enzymes EndoS and IdeS in an effort to understand the mechanism and specificity of these emerging therapeutic agents.

Work towards structural characterisation of EndoS was successful and utilised both bioinformatic and biophysical analyses to identify previously undescribed domains in the C-terminal region of EndoS. SAXS analysis gave an overall shape of EndoS in solution and revealed the multi-domain, L-shape structure of EndoS. The fact that these results were consistent with the recently published X-ray crystal structure of EndoS is testament to the power of these bioinformatic and biophysical approaches in the characterisation of proteins. Identification of a CBM in the C-terminal region of EndoS led to further exploration of this domain both structurally and functionally. Although no definitive carbohydrate binding capability was identified from the studies conducted here, this domain appears to have a functional role in the catalytic activity of EndoS against IgG glycans.

Upon publication of the X-ray crystal structure of EndoS, personal communication with the group responsible for the structure alerted us to the fact that they were actively pursuing the structure of the complex. Given this, efforts were dedicated elsewhere in the project. Despite this, preliminary characterisation of EndoS in complex with Fc had already been carried out using SAXS. The results of this analysis suggested that a conformationally stable complex of EndoS and Fc was not formed despite the coelution of these proteins during SEC. EndoS forms a structure resembling 'beads on a string' [149], with each bead or domain connected by a flexible linker. Given the potential for independent movement of these domains with respect to each other, it could be that conformational changes are induced or required for binding of Fc by EndoS. This may lead to a mixture of conformational states in the sample which are not able to be deconvoluted using SAXS.

A systematic characterisation of EndoS specificity was carried out during this project and revealed that EndoS was able to hydrolyse a wide variety of Fc glycans including oligomannose and hybrid type glycans and glycans with a bisecting GlcNAc. Contrary to the reported specificity of EndoS for IgG glycans only, these studies also identified the ability of EndoS to cleave glycans from alternative glycoprotein substrates. In a clinical context however, the non-specific activity of EndoS is unlikely to be significant. Nevertheless, appreciation of the off-target effects of EndoS is crucial if it is to be pursued further as a therapeutic agent.

Structural changes to Fc induced by EndoS deglycosylation were consistent with those observed when Fc is aglycosylated, compounding the notion that EndoS deglycosylation leads to a loss of FcγR and C1q binding by disrupting the structural integrity of the Fc domain. Investigations into *Mm*IgG2aFc were carried out in order to understand its ability to maintain binding to FcγRs even when deglycosylated. Although structural information was obtained for both *Mm*IgG2aFc and the *Mm*IgG1Fc, further analysis of these structures alone and in the context of the pool of published Fc structures is required in order to appreciate the factors which contribute to this unique behaviour of *Mm*IgG2a.

Work on the IdeS-Fc complex is still in preliminary stages. A stable complex of IdeS-Fc was able to be produced using a catalytically inactive mutant complexed with IgG1 Fc. The complex produced was consistent with a 1:1 binding stoichiometry between

IdeS and Fc and was used for further structural studies. Here, a combination of SAXS and *in situ* protein–protein docking experiments were used to create a model of IdeS in complex with Fc. X-ray crystallography of the IdeS-Fc complex was hampered by selective crystallisation of Fc, suggesting that alternative approaches may be required if this method is to be pursued. Overall, further work using additional complementary methods is required to fully characterise the binding interface and interaction between IdeS and Fc.

In conclusion, the work described here has contributed to the understanding of these important and useful enzymes, which is of worth as these enzymes are taken forward as therapeutic agents.

Bibliography

[1] Lancefield, R. C. (1933) A serological differentiation of human and other groups of hemolytic streptococci. *The Journal of Experimental Medicine.* **57**:4, pp. 571–595.

[2] Steer, A. C. *et al.* (2009) Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. *The Lancet. Infectious diseases.* **9**:10, pp. 611–616.

[3] Harvey, R. A., Champe, P. C. and Fisher, B. D. (2007) *Microbiology.* Ed. by R. A. Harvey and P. C. Champe. 2nd ed. Lippincott Williams & Wilkins.

[4] Severin, A. *et al.* (2007) Proteomic Analysis and Identification of *Streptococcus pyogenes* Surface-Associated Proteins. *Journal of bacteriology.* **189**:5, pp. 1514–1522.

[5] WHO (2005) *The current evidence for the burden of group A streptococcal diseases.* Discussion papers on child health. WHO.

[6] Cunningham, M. W. (2000) Pathogenesis of Group A Streptococcal Infections. *Clinical Microbiology Reviews.* **13**:3, pp. 470–511.

[7] Breiman, R. F. *et al.* (1993) Defining the Group A Streptococcal Toxic Shock Syndrome. *JAMA.* **269**:3, pp. 390–391.

[8] Lancefield, R. C. (1962) Current Knowledge of Type-Specific M Antigens of Group a Streptococci. *Journal of Immunology.* **89**:3, pp. 307–313.

[9] Guilherme, L., Kalil, J. and Cunningham, M. (2006) Molecular mimicry in the autoimmune pathogenesis of rheumatic heart disease. *Autoimmunity.* **39**:1, pp. 31–39.

[10] Massell, B. F., Honikman, L. H. and Amezcua, J (1969) Rheumatic fever following streptococcal vaccination. Report of three cases. *JAMA.* **207**:6, pp. 1115–1119.

[11] Steer, A. C., Dale, J. B. and Carapetis, J. R. (2013) Progress Toward a Global Group A Streptococcal Vaccine. *The Pediatric Infectious Disease Journal.* **32**:2, pp. 180–182.

[12] Pichichero, M. E. and Casey, J. R. (2007) Systematic review of factors contributing to penicillin treatment failure in *Streptococcus pyogenes* pharyngitis. *Otolaryngology–Head and Neck Surgery.* **137**:6, pp. 851–857.

[13] Shah, S. S. *et al.* (2009) Intravenous immunoglobulin in children with streptococcal toxic shock syndrome. *Clinical Infectious Diseases.* **49**:9, pp. 1369–1376.

[14] Kaul, R. *et al.* (1999) Intravenous immunoglobulin therapy for streptococcal toxic shock syndrome–a comparative observational study. The Canadian Streptococcal Study Group. *Clinical Infectious Diseases.* **28**:4, pp. 800–807.

[15] Darenberg, J. *et al.* (2003) Intravenous immunoglobulin G therapy in streptococcal toxic shock syndrome: a European randomized, double-blind, placebo-controlled trial. *Clinical Infectious Diseases.* **37**:3, pp. 333–340.

[16] Rohde, M. and Chhatwal, G. S. (2013) Adherence and Invasion of Streptococci to Eukaryotic Cells and Their Role in Disease Pathogenesis. In: *Host-Pathogen Interactions in Streptococcal Diseases*. Ed. by G. S. Chhatwal. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 83–110.

[17] LaPenta, D. *et al.* (1994) Group A streptococci efficiently invade human respiratory epithelial cells. *Proceedings of the National Academy of Sciences.* **91**:25, pp. 12115–12119.

[18] Kovarik, P. and Fieber, C. (2014) Responses of innate immune cells to Group A Streptococcus. *Frontiers in Cellular and Infection Microbiology.* **4**, pp. 1–7.

[19] Senavsky, C., Craft, N. and Miller, L. S. (2008) Bacterial Infections. In: *Clinical and basic immunodermatology*. Ed. by A. Gaspari and S. K. Tyring. Springer-Verlag London Limited, pp. 335–361.

[20] Nguyen, L. T., Haney, E. F. and Vogel, H. J. (2011) The expanding scope of antimicrobial peptide structures and their modes of action. *Trends in Biotechnology.* **29**:9, pp. 464–472.

[21] Fuchs, T. A. *et al.* (2007) Novel cell death program leads to neutrophil extracellular traps. *The Journal of Cell Biology.* **176**:2, pp. 231–241.

[22] Brinkmann, V. *et al.* (2004) Neutrophil extracellular traps kill bacteria. *Science.* **303**:5663, pp. 1532–1535.

[23] Murphy, K., Travers, P. and Walport, M. (2008) *Janeway's Immunobiology.* 7th ed. Garland Science.

[24] Amzel, M. L. and Poljak, R. J. (1979) Three-Dimensional Structure of Immunoglobulins. *Annual Review of Biochemistry.* 48, pp. 961–997.

[25] Stavnezer, J., Guikema, J. E. J. and Schrader, C. E. (2008) Mechanism and regulation of class switch recombination. *Immunology.* **26**, pp. 261–292.

[26] Stoop, J. W. *et al.* (1969) Serum immunoglobulin levels in healthy children and adults. *Clinical and Experimental Immunology.* **4**:1, pp. 101–112.

[27] Jefferis, R. and Lund, J. (2002) Interaction sites on human IgG-Fc for Fc gamma R: current models. *Immunology Letters.* **82**:1-2, pp. 57–65.

[28] Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research.* **42**, pp. W320–W324.

[29] Sievers, F. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology.* **7**:539, pp. 1–6.

[30] Edelman, G. M. *et al.* (1969) The covalent structure of an entire gammaG immunoglobulin molecule. *Proceedings of the National Academy of Sciences.* **63**:1, pp. 78–85.

[31] Woof, J. M. and Russell, M. W. (2011) Structure and function relationships in IgA. *Mucosal Immunology.* **4**:6, pp. 590–597.

[32] Macpherson, A. J. *et al.* (2001) IgA responses in the intestinal mucosa against pathogenic and non-pathogenic microorganisms. *Microbes and Infection.* **3**:12, pp. 1021–1035.

[33] Chen, K. and Cerutti, A. (2010) New insights into the enigma of immunoglobulin D. *Immunological reviews.* **237**:1, pp. 160–179.

[34] Gould, H. J. *et al.* (2003) The biology of IgE and the basis of allergic disease. *Immunology.* **21**, pp. 579–628.

[35] Iio, A., Waldmann, T. A. and Strober, W. (1978) Metabolic Study of Human IgE: Evidence for an Extravascular Catabolic Pathway. *The Journal of Immunology.* **120**:5, pp. 1696–1701.

[36] Gould, H. J. and Sutton, B. J. (2008) IgE in allergy and asthma today. *Nature Reviews Immunology.* **8**:3, pp. 205–217.

[37] Coutinho, A., Kazatchkine, M. D. and Avrameas, S. (1995) Natural autoantibodies. *Current opinion in immunology.* **7**:6, pp. 812–818.

[38] Ehrenstein, M. R. and Notley, C. A. (2010) The importance of natural IgM: scavenger, protector and regulator. *Nature Reviews Immunology.* **10**:11, pp. 778–786.

[39] Nimmerjahn, F. and Ravetch, J. V. (2010) Antibody-mediated modulation of immune responses. *Immunological reviews.* **236**, pp. 265–275.

[40] Nose, M. and Wigzell, H. (1983) Biological significance of carbohydrate chains on monoclonal antibodies. *Proceedings of the National Academy of Sciences.* **80**:21, pp. 6632–6636.

[41] Borrok, M. J. *et al.* (2012) Revisiting the role of glycosylation in the structure of human IgG Fc. *ACS chemical biology.* **7**:9, pp. 1596–1602.

[42] Saphire, E. O. *et al.* (2001) Crystal Structure of a Neutralizing Human IgG Against HIV-1: A Template for Vaccine Design. *Science.* **293**:5532, pp. 1155–1159.

[43] Huber, R. *et al.* (1976) Crystallographic structure studies of an IgG molecule and an Fc fragment. *Nature.* **264**:5585, pp. 415–420.

[44] Ahmed, A. A. *et al.* (2014) Structural Characterization of Anti-Inflammatory Immunoglobulin G Fc Proteins. *Journal of Molecular Biology.* **426**:18, pp. 3166–3179.

[45] Halaby, D. M., Poupon, A. and Mornon, J.-P. (1999) The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Engineering.* **12**:7, pp. 563–571.

[46] Matsumiya, S. *et al.* (2011) Corrigendum to "Structural Comparison of Fucosylated and Nonfucosylated Fc Fragments of Human Immunoglobulin G1" [J. Mol. Biol.386/3 (2007) 767–779]. *Journal of Molecular Biology.* **408**:5, pp. 1001–1001.

[47] Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology.* **372**:3, pp. 774–797.

[48] Laskowski, R. A. (2007) Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. *Bioinformatics.* **23**:14, pp. 1824–1827.

[49] Ravetch, J. V. and Bolland, S. (2001) IgG Fc Receptors. *Immunology.* **19**:1, pp. 275–290.

[50] Duncan, A. R. and Winter, G. (1988) The binding site for C1q on IgG. *Nature.* **332**:6166, pp. 738–740.

[51] Rother, K, Till, G. and Hänsch, G. M. (1998) *The Complement System.* Springer Science & Business Media.

[52] Bruhns, P. (2012) Properties of mouse and human IgG receptors and their contribution to disease models. *Blood.* **119**:24, pp. 5640–5649.

[53] Nimmerjahn, F. and Ravetch, J. V. (2008) Fcγ receptors as regulators of immune responses. *Nature Reviews Immunology.* **8**:1, pp. 34–47.

[54] Ravetch, J. V. and Perussia, B. (1989) Alternative membrane forms of Fc gamma RIII(CD16) on human natural killer cells and neutrophils. Cell type-specific expression of two genes that differ in single nucleotide substitutions. *The Journal of Experimental Medicine.* **170**:2, pp. 481–497.

[55] Graf, S. W. *et al.* (2012) Low copy number of the FCGR3B gene and rheumatoid arthritis: a case-control study and meta-analysis. *Arthritis Research & Therapy.* **14**:R28, pp. 1–7.

[56] Fossati, G. *et al.* (2002) Differential role of neutrophil Fc gamma receptor IIIb (CD16) in phagocytosis, bacterial killing, and responses to immune complexes. *Arthritis & Rheumatism.* **46**:5, pp. 1351–1361.

[57] Sondermann, P. *et al.* (2000) The 3.2-angstrom crystal structure of the human IgG1 Fc fragment-Fc gamma RIII complex. *Nature.* **406**:6793, pp. 267–273.

[58] Radaev, S. *et al.* (2001) The structure of a human type III Fc gamma receptor in complex with Fc. *The Journal of Biological Chemistry.* **276**:19, pp. 16469–16477.

[59] Ferrara, C. *et al.* (2011) Unique carbohydrate–carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proceedings of the National Academy of Sciences.* **108**:31, pp. 12669–12674.

[60] Ramsland, P. A. *et al.* (2011) Structural basis for Fc gammaRIIa recognition of human IgG and formation of inflammatory signaling complexes. *Journal of Immunology.* **187**:6, pp. 3208–3217.

[61] Mimoto, F. *et al.* (2013) Engineered antibody Fc variant with selectively enhanced FcγRIIb binding over both FcγRIIaR131 and FcγRIIaH131. *Protein Engineering Design and Selection.* **26**:10, pp. 589–598.

[62] Shields, R. L. *et al.* (2000) High Resolution Mapping of the Binding Site on Human IgG1 for Fcgamma RI, Fcgamma RII, Fcgamma RIII, and FcRn and Design of IgG1 Variants with Improved Binding to the Fcgamma R. *Journal of Biological Chemistry.* **276**:9, pp. 6591–6604.

[63] Roopenian, D. C. and Akilesh, S. (2007) FcRn: the neonatal Fc receptor comes of age. *Nature Reviews Immunology.* **7**:9, pp. 715–725.

[64] Morgan, A. *et al.* (1995) The N-terminal end of the CH2 domain of chimeric human IgG1 anti-HLA-DR is necessary for C1q, Fc gamma RI and Fc gamma RIII binding. *Immunology.* **86**:2, pp. 319–324.

[65] McEwan, W. A. *et al.* (2013) Intracellular antibody-bound pathogens stimulate immune signaling via the Fc receptor TRIM21. *Nature Immunology.* **14**:4, pp. 327–336.

[66] James, L. C. *et al.* (2007) Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proceedings of the National Academy of Sciences.* **104**:15, pp. 6200–6205.

[67] Keeble, A. H. *et al.* (2008) TRIM21 is an IgG receptor that is structurally, thermodynamically, and kinetically conserved. *Proceedings of the National Academy of Sciences.* **105**:16, pp. 6045–6050.

[68] Sondermann, P. *et al.* (2013) General mechanism for modulating immunoglobulin effector function. *Proceedings of the National Academy of Sciences.* **110**:24, pp. 9868–9872.

[69] Anthony, R. M. *et al.* (2008) Identification of a receptor required for the anti-inflammatory activity of IVIG. *Proceedings of the National Academy of Sciences.* **105**:50, pp. 19571–19578.

[70] Yu, X. *et al.* (2013a) Dissecting the Molecular Mechanism of IVIg Therapy: The Interaction between Serum IgG and DC-SIGN is Independent of Antibody Glycoform or Fc Domain. *Journal of Molecular Biology.* **425**:8, pp. 1253–1258.

[71] Deisenhofer, J. (1981) Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution. *Biochemistry.* **20**:9, pp. 2361–2370.

[72] Sauer-Eriksson, A. E. *et al.* (1995) Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure.* **3**:3, pp. 265–278.

[73] Falugi, F. *et al.* (2013) Role of Protein A in the Evasion of Host Adaptive Immune Responses by Staphylococcus aureus. *mBio.* **4**:5, pp. e00575.

[74] Nitsche-Schmitz, D. P. *et al.* (2007) Group G Streptococcal IgG Binding Molecules FOG and Protein G Have Different Impacts on Opsonization by C1q. *Journal of Biological Chemistry.* **282**:24, pp. 17530–17536.

[75] Wollein Waldetoft, K. *et al.* (2012) Streptococcal Surface Proteins Activate the Contact System and Control Its Antibacterial Activity. *Journal of Biological Chemistry.* **287**:30, pp. 25010–25018.

[76] Anumula, K. R. (2012) Quantitative glycan profiling of normal human plasma derived immunoglobulin and its fragments Fab and Fc. *Journal of Immunological Methods.* **382**:1-2, pp. 167–176.

[77] Holland, M. *et al.* (2006) Differential glycosylation of polyclonal IgG, IgG-Fc and IgG-Fab isolated from the sera of patients with ANCA-associated systemic vasculitis. *Biochimica Et Biophysica Acta-General Subjects.* **1760**:4, pp. 669–677.

[78] Jefferis, R. (2005) Glycosylation of recombinant antibody therapeutics. *Biotechnology Progress.* **21**:1, pp. 11–16.

[79] Arnold, J. N. *et al.* (2007) The impact of glycosylation on the biological function and structure of human immunoglobulins. *Immunology.* **25**, pp. 21–50.

[80] Weerapana, E. and Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology.* **16**:6, pp. 91R–101R.

[81] Kim, P.-J., Lee, D.-Y. and Jeong, H. (2009) Centralized Modularity of N-Linked Glycosylation Pathways in Mammalian Cells. *PLoS ONE.* **4**:10, pp. e7317.

[82] Reeves, P. J., Kim, J. M. and Khorana, H. G. (2002) Structure and function in rhodopsin: A tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proceedings of the National Academy of Sciences.* **99**:21, pp. 13413–13418.

[83] Harvey, D. J. *et al.* (2009) Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics.* **9**:15, pp. 3796–3801.

[84] Vliegenthart, J. F. G., Dorland, L. and van Halbeek, H. (1983) High-Resolution, 1H-Nuclear Magnetic Resonance Spectroscopy as a Tool in the Structural Analysis of Carbohydrates Related to Glycoproteins. *Advances in Carbohydrate Chemistry and Biochemistry.* **41**, pp. 209–374.

[85] Guile, G. R. *et al.* (1996) A Rapid High-Resolution High-Performance Liquid Chromatographic Method for Separating Glycan Mixtures and Analyzing Oligosaccharide Profiles. *Analytical Biochemistry.* **240**:2, pp. 210–226.

[86] Lund, J. *et al.* (1996) Multiple interactions of IgG with its core oligosaccharide can modulate recognition by complement and human Fc gamma receptor I and influence the synthesis of its oligosaccharide chains. *Journal of Immunology.* **157**:11, pp. 4963–4969.

[87] Yu, X. *et al.* (2013b) Engineering hydrophobic protein-carbohydrate interactions to fine-tune monoclonal antibodies. *J. Am. Chem. Soc.* **135**:26, pp. 9723–9732.

[88] Crispin, M., Yu, X. and Bowden, T. A. (2013) Crystal structure of sialylated IgG Fc: Implications for the mechanism of intravenous immunoglobulin therapy. *Proceedings of the National Academy of Sciences.* **110**:38, pp. E3544–E3546.

[89] Mimura, Y. *et al.* (2000) The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. *Molecular Immunology.* **37**:12-13, pp. 697–706.

[90] Krapp, S. *et al.* (2003) Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *Journal of Molecular Biology.* **325**:5, pp. 979–989.

[91] Bowden, T. A. *et al.* (2012) Chemical and structural analysis of an antibody folding intermediate trapped during glycan biosynthesis. *J. Am. Chem. Soc.* **134**:42, pp. 17554–17563.

[92] Kelliher, M. T. *et al.* (2014) The effect of sugar removal on the structure of the Fc region of an IgG antibody as observed with single molecule Förster Resonance Energy Transfer. *Molecular Immunology.* **60**:2, pp. 103–108.

[93] Walker, M. R. *et al.* (1989) Aglycosylation of human IgG1 and IgG3 monoclonal antibodies can eliminate recognition by human cells expressing Fc gamma RI and/or Fc gamma RII receptors. *Biochemical Journal.* **259**:2, pp. 347.

[94] Sazinsky, S. L. *et al.* (2008) Aglycosylated immunoglobulin G1 variants productively engage activating Fc receptors. *Proceedings of the National Academy of Sciences.* **105**:51, pp. 20167–20172.

[95] Jung, S. T. *et al.* (2010) Aglycosylated IgG variants expressed in bacteria that selectively bind Fc gamma RI potentiate tumor cell killing by monocyte-dendritic cells. *Proceedings of the National Academy of Sciences.* **107**:2, pp. 604–609.

[96] Jung, S. T. *et al.* (2013) Effective Phagocytosis of Low Her2 Tumor Cell Lines with Engineered, Aglycosylated IgG Displaying High Fc gamma RIIa Affinity and Selectivity. *ACS chemical biology.* **8**:2, pp. 368–375.

[97] Ackerman, M. and Nimmerjahn, F. (2013) *Antibody Fc.* Linking Adaptive and Innate Immunity. Academic Press.

[98] Parekh, R. B. *et al.* (1985) Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature.* **316**:6027, pp. 452–457.

[99] Frick, I.-M. *et al.* (2003) SIC, a Secreted Protein of *Streptococcus pyogenes* that Inactivates Antibacterial Peptides. *Journal of Biological Chemistry.* **278**:19, pp. 16561–16566.

[100] Fernie-King, B. A., Seilly, D. J. and Lachmann, P. J. (2004) The interaction of streptococcal inhibitor of complement (SIC) and its proteolytic fragments with the human beta defensins. *Immunology.* **111**:4, pp. 444–452.

[101] Fernie-King, B. A. *et al.* (2002) Streptococcal Inhibitor of Complement Inhibits Two Additional Components of the Mucosal Innate Immune System: Secretory Leukocyte Proteinase Inhibitor and Lysozyme. *Infection and immunity.* **70**:9, pp. 4908–4916.

[102] Nyberg, P., Rasmussen, M. and Björck, L. (2004) α2-Macroglobulin-Proteinase Complexes Protect *Streptococcus pyogenes* from Killing by the Antimicrobial Peptide LL-37. *Journal of Biological Chemistry.* **279**:51, pp. 52820–52823.

[103] Vindebro, R. (2014) "Studies on secreted cysteine proteases of *Streptococcus pyogenes* : IdeS and SpeB". PhD thesis. Umeå University.

[104] Collin, M. and Olsén, A. (2003) Extracellular enzymes with immunomodulating activities: variations on a theme in *Streptococcus pyogenes*. *Infection and immunity.* **71**:6, pp. 2983–2992.

[105] Schmidtchen, A., Frick, I.-M. and Björck, L. (2001) Dermatan sulphate is released by proteinases of common pathogenic bacteria and inactivates antibacterial alpha-defensin. *Molecular Microbiology.* **39**:3, pp. 708–713.

[106] Jarva, H. *et al.* (2003) Complement resistance mechanisms of streptococci. *Molecular Immunology.* **40**:2–4, pp. 95–107.

[107] Horstmann, R. D. *et al.* (1988) Antiphagocytic Activity of Streptococcal-M Protein - Selective Binding of Complement Control Protein Factor-H. *Proceedings of the National Academy of Sciences.* **85**:5, pp. 1657–1661.

[108] Gustafsson, M. C. U. *et al.* (2013) Factor H binds to the hypervariable region of many *Streptococcus pyogenes* M proteins but does not promote phagocytosis resistance or acute virulence. *PLoS Pathogens.* **9**:4, pp. e1003323.

[109] Pandiripally, V., Gregory, E. and Cue, D. (2002) Acquisition of regulators of complement activation by *Streptococcus pyogenes* serotype M1. *Infection and immunity.* **70**:11, pp. 6206–6214.

[110] Kihlberg, B.-M. *et al.* (1999) Protein H, an Antiphagocytic Surface Protein in *Streptococcus pyogenes*. *Infection and immunity.* **67**:4, pp. 1708–1714.

[111] Horstmann, R. D. *et al.* (1992) Role of Fibrinogen in Complement Inhibition by Streptococcal M-Protein. *Infection and immunity.* **60**:12, pp. 5036–5041.

[112] Carlsson, F., Sandin, C. and Lindahl, G. (2005) Human fibrinogen bound to *Streptococcus pyogenes* M protein inhibits complement deposition via the classical pathway. *Molecular Microbiology.* **56**:1, pp. 28–39.

[113] Thern, A. *et al.* (1995) Ig-binding surface proteins of *Streptococcus pyogenes* also bind human C4b-binding protein (C4BP), a regulatory component of the complement system. *The Journal of Immunology.* **154**:1, pp. 375–386.

[114] Berggard, K *et al.* (2001) Binding of human C4BP to the hypervariable region of M protein: a molecular mechanism of phagocytosis resistance in *Streptococcus pyogenes*. *Molecular Microbiology.* **42**:2, pp. 539–551.

[115] Hidalgo-Grass, C. *et al.* (2006) A streptococcal protease that degrades CXC chemokines and impairs bacterial clearance from infected tissues. *The EMBO Journal.* **25**:19, pp. 4628–4637.

[116] Joiner, K. A., Brown, E. J. and Frank, M. M. (1984) Complement and Bacteria: Chemistry and Biology in Host Defense. *Immunology.* **2**:1, pp. 461–492.

[117] Berends, E. T. M. *et al.* (2013) Distinct localization of the complement C5b-9 complex on Gram-positive bacteria. *Cellular Microbiology.* **15**:12, pp. 1955–1968.

[118] Åkesson, P., Sjoholm, A. G. and Björck, L. (1996) Protein SIC, a novel extracellular protein of *Streptococcus pyogenes* interfering with complement function. *The Journal of Biological Chemistry.* **271**:2, pp. 1081–1088.

[119] Fernie-King, B. A. *et al.* (2001) Streptococcal inhibitor of complement (SIC) inhibits the membrane attack complex by preventing uptake of C567 onto cell membranes. *Immunology.* **103**:3, pp. 390–398.

[120] Lin, A. *et al.* (2009) Streptolysin S Inhibits Neutrophil Recruitment during the Early Stages of *Streptococcus pyogenes* Infection. *Infection and immunity.* **77**:11, pp. 5190–5201.

[121] Sumby, P. *et al.* (2005) Extracellular deoxyribonuclease made by group A Streptococcus assists pathogenesis by enhancing evasion of the innate immune response. *Proceedings of the National Academy of Sciences.* **102**:5, pp. 1679–1684.

[122] Chang, A. *et al.* (2011) Functional analysis of *Streptococcus pyogenes* nuclease A (SpnA), a novel group A streptococcal virulence factor. *Molecular Microbiology.* **79**:6, pp. 1629–1642.

[123] Buchanan, J. T. *et al.* (2006) DNase expression allows the pathogen group A streptococcus to escape killing in neutrophil extracellular traps. *Current Biology.* **16**:4, pp. 396–400.

[124] Necas, J. *et al.* (2008) Hyaluronic acid (hyaluronan): a review. *Veterinarni medicina.* **53**:8, pp. 397–411.

[125] Wessels, M. R. and Bronze, M. S. (1994) Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice. *Proceedings of the National Academy of Sciences.* **91**:25, pp. 12238–12242.

[126] Dinkla, K. *et al.* (2003) *Streptococcus pyogenes* recruits collagen via surface-bound fibronectin: a novel colonization and immune evasion mechanism. *Molecular Microbiology.* **47**:3, pp. 861–869.

[127] Sierig, G. *et al.* (2003) Cytotoxic Effects of Streptolysin O and Streptolysin S Enhance the Virulence of Poorly Encapsulated Group A Streptococci. *Infection and immunity.* **71**:1, pp. 446–455.

[128] Goldmann, O. *et al.* (2009) *Streptococcus pyogenes* induces oncosis in macrophages through the activation of an inflammatory programmed cell death pathway. *Cellular Microbiology.* **11**:1, pp. 138–155.

[129] Timmer, A. M. *et al.* (2009) Streptolysin O Promotes Group A Streptococcus Immune Evasion by Accelerated Macrophage Apoptosis. *The Journal of Biological Chemistry.* **284**:2, pp. 862–871.

[130] Hakansson, A. *et al.* (2005) Cytolysin-dependent evasion of lysosomal killing. *Proceedings of the National Academy of Sciences.* **102**:14, pp. 5192–5197.

[131] Hertzen, E. *et al.* (2010) M1 Protein-Dependent Intracellular Trafficking Promotes Persistence and Replication of *Streptococcus pyogenes* in Macrophages. *Journal of Innate Immunity.* **2**:6, pp. 534–545.

[132] Staali, L. *et al.* (2003) *Streptococcus pyogenes* expressing M and M-like surface proteins are phagocytosed but survive inside human neutrophils. *Cellular Microbiology.* **5**:4, pp. 253–265.

[133] Thulin, P. *et al.* (2006) Viable Group A Streptococci in Macrophages during Acute Soft Tissue Infection. *PLOS Medicine.* **3**:3, pp. e53.

[134] Gomi, H. *et al.* (1990) The Gene Sequence and Some Properties of Protein-H - a Novel IgG-Binding Protein. *Journal of Immunology.* **144**:10, pp. 4046–4052.

[135] Lindahl, G. and Stenberg, L. (1990) Binding of IgA and/or IgG is a common property among clinical isolates of group A streptococci. *Epidemiology and Infection.* **105**:1, pp. 87–93.

[136] Stenberg, L., O'Toole, P. and Lindahl, G. (1992) Many Group A Streptococcal Strains Express 2 Different Immunoglobulin-Binding Proteins, Encoded by Closely Linked Genes - Characterization of the Proteins Expressed by 4 Strains of Different M-Type. *Molecular Microbiology.* **6**:9, pp. 1185–1194.

[137] Cleary, P. and Retnoningrum, D. Group A streptococcal immunoglobulin-binding proteins: adhesins, molecular mimicry or sensory proteins? *Trends in Microbiology.* **2**:4, pp. 131–136.

[138] Berge, A. *et al.* (1997) Streptococcal protein H forms soluble complement-activating complexes with IgG, but inhibits complement activation by IgG-coated targets. *The Journal of Biological Chemistry.* **272**:33, pp. 20774–20781.

[139] Collin, M. and Olsén, A. (2001a) Effect of SpeB and EndoS from *Streptococcus pyogenes* on human immunoglobulins. *Infection and immunity.* **69**:11, pp. 7187–7189.

[140] Collin, M. *et al.* (2002) EndoS and SpeB from *Streptococcus pyogenes* inhibit immunoglobulin-mediated opsonophagocytosis. *Infection and immunity.* **70**:12, pp. 6646–6651.

[141] Collin, M. and Olsén, A. (2001b) EndoS, a novel secreted protein from *Streptococcus pyogenes* with endoglycosidase activity on human IgG. *The EMBO Journal.* **20**:12, pp. 3046–3055.

[142] Baruah, K. *et al.* (2012) Selective deactivation of serum IgG: a general strategy for the enhancement of monoclonal antibody receptor interactions. *Journal of Molecular Biology.* **420**:1-2, pp. 1–7.

[143] Goodfellow, J. J. *et al.* (2012) An Endoglycosidase with Alternative Glycan Specificity Allows Broadened Glycoprotein Remodelling. *J. Am. Chem. Soc.* **134**:19, pp. 8030–8033.

[144] Albert, H. *et al.* (2008) In vivo enzymatic modulation of IgG glycosylation inhibits autoimmune disease in an IgG subclass-dependent manner. *Proceedings of the National Academy of Sciences.* **105**:39, pp. 15005–15009.

[145] Allhorn, M. *et al.* (2008) Human IgG/FcγR Interactions Are Modulated by Streptococcal IgG Glycan Hydrolysis. *PLoS ONE.* **3**:1, pp. e1413.

[146] Sjögren, J. J. *et al.* (2011) Study of the IgG endoglycosidase EndoS in group A streptococcal phagocyte resistance and virulence. *BMC microbiology.* **11**, pp. 120–120.

[147] Allhorn, M., Olsén, A. and Collin, M. (2008) EndoS from *Streptococcus pyogenes* is hydrolyzed by the cysteine proteinase SpeB and requires glutamic acid 235 and tryptophans for IgG glycan-hydrolyzing activity. *BMC microbiology.* **8**:3, pp. 1–10.

[148] Kobe, B. B. and Kajava, A. V. A. (2001) The leucine-rich repeat as a protein recognition motif. *Current Opinion in Structural Biology.* **11**:6, pp. 725–732.

[149] Trastoy, B. *et al.* (2014) Crystal structure of *Streptococcus pyogenes* EndoS, an immunomodulatory endoglycosidase specific for human IgG antibodies. *Proceedings of the National Academy of Sciences.* **111**:18, pp. 6714–6719.

[150] Sjögren, J. *et al.* (2013) EndoS2 is a unique and conserved enzyme of serotype M49 group A Streptococcus that hydrolyses N-linked glycans on IgG and α1-acid glycoprotein. *Biochem. J.* **455**:1, pp. 107–118.

[151] Lei, B. *et al.* (2000) Identification and immunogenicity of group A Streptococcus culture supernatant proteins. *Infection and immunity.* **68**:12, pp. 6807–6818.

[152] Lei, B. *et al.* (2001) Evasion of human innate and acquired immunity by a bacterial homolog of CD11b that inhibits opsonophagocytosis. *Nature Medicine.* **7**:12, pp. 1298–1305.

[153] Von Pawel-Rammingen, U., Johansson, B. P. and Björck, L. (2002) IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *The EMBO Journal.* **21**:7, pp. 1607–1615.

[154] Kawabata, S. *et al.* (2002) A novel, anchorless streptococcal surface protein that binds to human immunoglobulins. *Biochemical and biophysical research communications.* **296**:5, pp. 1329–1333.

[155] Agniswamy, J. *et al.* (2006) Crystal Structure of Group A Streptococcus Mac-1: Insight into Dimer-Mediated Specificity for Recognition of Human IgG. *Structure.* **14**:2, pp. 225–235.

[156] Wenig, K. *et al.* (2004) Structure of the streptococcal endopeptidase IdeS, a cysteine proteinase with strict specificity for IgG. *Proceedings of the National Academy of Sciences.* **101**:50, pp. 17371–17376.

[157] Lei, B. *et al.* (2003) Histidine and aspartic acid residues important for immunoglobulin G endopeptidase activity of the group A Streptococcus opsonophagocytosis-inhibiting Mac protein. *Infection and immunity.* **71**:5, pp. 2881–2884.

[158] Vincents, B. *et al.* (2004) Enzymatic characterization of the streptococcal endopeptidase, IdeS, reveals that it is a cysteine protease with strict specificity for IgG cleavage due to exosite binding. *Biochemistry.* **43**:49, pp. 15540–15549.

[159] Vindebro, R., Spoerry, C. and von Pawel-Rammingen, U. (2013) Rapid IgG heavy chain cleavage by the streptococcal IgG endopeptidase IdeS is mediated by IdeS monomers and is not due to enzyme dimerization. *FEBS Letters.* **587**:12, pp. 1818–1822.

[160] Lei, B. *et al.* (2002) Opsonophagocytosis-inhibiting mac protein of group a streptococcus: identification and characteristics of two genetic complexes. *Infection and immunity.* **70**:12, pp. 6880–6890.

[161] Soderberg, J. J., Engstrom, P. and von Pawel-Rammingen, U. (2008) The intrinsic immunoglobulin G endopeptidase activity of streptococcal Mac-2 proteins implies a unique role for the enzymatically impaired Mac-2 protein of M28 serotype strains. *Infection and immunity.* **76**:5, pp. 2183–2188.

[162] Ryan, M. H. *et al.* (2008) Proteolysis of purified IgGs by human and bacterial enzymes in vitro and the detection of specific proteolytic fragments of endogenous IgG in rheumatoid synovial fluid. *Molecular Immunology.* **45**:7, pp. 1837–1846.

[163] Agniswamy, J. *et al.* (2004) Insight of Host Immune Evasion Mediated by Two Variants of Group A Streptococcus Mac Protein. *Journal of Biological Chemistry.* **279**:50, pp. 52789–52796.

[164] Brezski, R. J. *et al.* (2009) Tumor-associated and microbial proteases compromise host IgG effector functions by a single cleavage proximal to the hinge. *Proceedings of the National Academy of Sciences.* **106**:42, pp. 17864–17869.

[165] Von Pawel-Rammingen, U. *et al.* (2002) *Streptococcus pyogenes* and phagocytic killing. *Nature Medicine.* **8**:11, pp. 1329–1329.

[166] D'Souza, S. E., Ginsberg, M. H. and Plow, E. F. (1991) Arginyl-Glycyl-Aspartic Acid (RGD) - a Cell-Adhesion Motif. *Trends in Biochemical Sciences.* **16**:7, pp. 246–250.

[167] Yamada, T. *et al.* (2002) Bacterial redox protein azurin, tumor suppressor protein p53, and regression of cancer. *Proceedings of the National Academy of Sciences.* **99**:22, pp. 14098–14103.

[168] Punj, V. *et al.* Bacterial cupredoxin azurin as an inducer of apoptosis and regression in human breast cancer. *Oncogene.* **23**:13, pp. 2367–2378.

[169] Chaudhari, A. *et al.* (2006) Azurin, *Plasmodium falciparum* malaria and HIV/AIDS: inhibition of parasitic and viral growth by Azurin. *Cell cycle.* **5**:15, pp. 1642–1648.

[170] Alcami, A. and Saraiva, M. (2009) Chemokine Binding Proteins Encoded by Pathogens. *Pathogen-Derived Immunomodulatory Molecules.* **666**, pp. 167–179.

[171] Donaldson, D. S. and Williams, N. A. (2009) Bacterial Toxins as Immunomodulators. *Pathogen-Derived Immunomodulatory Molecules.* **666**, pp. 1–18.

[172] Elliott, D. E. and Weinstock, J. V. (2009) Helminthic Therapy: Using Worms to Treat Immune-Mediated Disease. *Pathogen-Derived Immunomodulatory Molecules.* **666**, pp. 157–166.

[173] Fallon, P. G. and Alcami, A. (2006) Pathogen-derived immunomodulatory molecules: future immunotherapeutics? *Trends in immunology.* **27**:10, pp. 470–476.

[174] Rueter, C. and Hardwidge, P. R. (2014) 'Drugs from Bugs': bacterial effector proteins as promising biological (immune-) therapeutics. *FEMS Microbiology Letters.* **351**:2, pp. 126–132.

[175] Lim, P.-L. and Zouali, M. (2006) Pathogenic autoantibodies: emerging insights into tissue injury. *Immunology Letters.* **103**:1, pp. 17–26.

[176] Bambauer, R. *et al.* (2013) Therapeutic apheresis in autoimmune diseases. *Open Access Rheumatology: Research and Reviews.* **5**, pp. 93–103.

[177] Edwards, J. C. W. and Cambridge, G. (2006) B-cell targeting in rheumatoid arthritis and other autoimmune diseases. *Nature Reviews Immunology.* **6**:5, pp. 394–403.

[178] Kojouri, K. *et al.* (2004) Splenectomy for adult patients with idiopathic thrombocytopenic purpura: a systematic review to assess long-term platelet count responses, prediction of response, and surgical complications. *Blood.* **104**:9, pp. 2623–2634.

[179] Nandakumar, K. S. and Holmdahl, R. (2008) Therapeutic cleavage of IgG: new avenues for treating inflammation. *Trends in immunology.* **29**:4, pp. 173–178.

[180] Collin, M. (2012) EndoS, a "drug from a bug" as a novel treatment for autoimmunity of the blood. *ISBT Science Series.* **7**:1, pp. 142–145.

[181] Johansson, B. P., Shannon, O. and Björck, L. (2008) IdeS: a bacterial proteolytic enzyme with therapeutic potential. *PLoS ONE.* **3**:2, pp. e1692.

[182] Nandakumar, K. S. *et al.* (2007b) Endoglycosidase treatment abrogates IgG arthritogenicity: Importance of IgG glycosylation in arthritis. *European Journal of Immunology.* **37**:10, pp. 2973–2982.

[183] Collin, M., Shannon, O. and Björck, L. (2008) IgG glycan hydrolysis by a bacterial enzyme as a therapy against autoimmune conditions. *Proceedings of the National Academy of Sciences.* **105**:11, pp. 4265–4270.

[184] Allhorn, M. *et al.* (2010) The IgG-specific endoglycosidase EndoS inhibits both cellular and complement-mediated autoimmune hemolysis. *Blood.* **115**:24, pp. 5080–5088.

[185] Yang, R. *et al.* (2010) Successful treatment of experimental glomerulonephritis with IdeS and EndoS, IgG-degrading streptococcal enzymes. *Nephrology Dialysis Transplantation.* **25**:8, pp. 2479–2486.

[186] Hirose, M. *et al.* (2012) Enzymatic autoantibody glycan hydrolysis alleviates autoimmunity against type VII collagen. *Journal of Autoimmunity.* **39**:4, pp. 304–314.

[187] Tradtrantip, L. *et al.* (2013) Enzymatic deglycosylation converts pathogenic neuromyelitis optica anti-aquaporin-4 immunoglobulin G into therapeutic antibody. *Annals of Neurology.* **73**:1, pp. 77–85.

[188] Lood, C. *et al.* (2012) IgG glycan hydrolysis by endoglycosidase S diminishes the proinflammatory properties of immune complexes from patients with systemic lupus

erythematosus: A possible new treatment? *Arthritis & Rheumatism.* **64**:8, pp. 2698–2706.

[189] Cines, D. B. and Blanchette, V. S. (2002) Immune thrombocytopenic purpura. *The New England Journal of Medicine.* **346**:13, pp. 995–1008.

[190] Nandakumar, K. S. *et al.* (2007a) Blocking of experimental arthritis by cleavage of IgG antibodies in vivo. *Arthritis & Rheumatism.* **56**:10, pp. 3253–3260.

[191] Tradtrantip, L., Asavapanumas, N. and Verkman, A. S. (2013) Therapeutic cleavage of anti-aquaporin-4 autoantibody in neuromyelitis optica by an IgG-selective proteinase. *Molecular pharmacology.* **83**:6, pp. 1268–1275.

[192] Okamoto, S. *et al.* (2005) Systemic immunization with streptococcal immunoglobulin-binding protein Sib35 induces protective immunity against group A Streptococcus challenge in mice. *Vaccine.* **23**:40, pp. 4852–4859.

[193] Hansa Medical AB (2014) Phase II Study to Evaluate Safety and Efficacy of Intravenous Immunoglobulin G Degrading Enzyme of *Streptococcus pyogenes* (IdeS) in Chronic Kidney Disease. *http://www.clinicaltrials.gov/show/NCT02224820,* accessed October 2014.

[194] Fan, S. Q., Huang, W and Wang, L. X. (2012) Remarkable Transglycosylation Activity of Glycosynthase Mutants of Endo-D, an Endo- -N-acetylglucosaminidase from Streptococcus pneumoniae. *Journal of Biological Chemistry.* **287**:14, pp. 11272–11281.

[195] Shi, J., Blundell, T. L. and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of Molecular Biology.* **310**:1, pp. 243–257.

[196] Yang, Z. R. *et al.* (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics.* **21**:16, pp. 3369–3376.

[197] Thomson, R. *et al.* (2003) Characterizing proteolytic cleavage site activity using bio-basis function neural networks. *Bioinformatics.* **19**:14, pp. 1741–1747.

[198] Yang, Z. R. and Thomson, R. (2005) Bio-basis function neural network for prediction of protease cleavage sites in proteins. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council.* **16**:1, pp. 263–274.

[199] Gibson, D. G. *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods.* **6**:5, pp. 343–345.

[200] Miller, J. H. (1972) *Experiments in molecular genetics*. 3rd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

[201] Nimmerjahn, F. *et al.* (2005) FcγRIV: A Novel FcR with Distinct IgG Subclass Specificity. *Immunity.* **23**:1, pp. 41–51.

[202] Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification.* **41**:1, pp. 207–234.

[203] Aricescu, A. R., Lu, W. and Jones, E. Y. (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D.* **62**:10, pp. 1243–1250.

[204] Chang, V. T. *et al.* (2007) Glycoprotein structural genomics: solving the glycosylation problem. *Structure.* **15**:3, pp. 267–273.

[205] Smith, P. K. *et al.* (1985) Measurement of protein using bicinchoninic acid. *Analytical Biochemistry.* **150**:1, pp. 76–85.

[206] Gasteiger, E. *et al.* (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607.

[207] Pace, C. N. *et al.* (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Science.* **4**:11, pp. 2411–2423.

[208] Edelhoch, H. (1967) Spectroscopic Determination of Tryptophan and Tyrosine in Proteins*. *Biochemistry.* **6**:7, pp. 1948–1954.

[209] Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry.* **182**:2, pp. 319–326.

[210] Savitzky, A. and Golay, M. J. E. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry.* **36**:8, pp. 1627–1639.

[211] Whitmore, L. and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research.* **32**:Web Server issue, pp. W668–W673.

[212] Jacques, D. A. and Trewhella, J. (2010) Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein Science.* **19**:4, pp. 642–657.

[213] Feigin, L. A. and Svergun, D. I. (1987) *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Ed. by G. W. Taylor. New York: Plenium Press.

[214] Rambo, R. P. and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers.* **95**:8, pp. 559–571.

[215] Konarev, P. V., Volkov, V. V. and Sokolova, A. V. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography.* **36**, pp. 1277–1282.

[216] Rambo, R. P. and Tainer, J. A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* **496**:7446, pp. 477–481.

[217] Guinier, A. (1939) "La diffraction des rayons X aux trés petits angles : application a l'étude de phénomènes ultramicroscopiques". PhD thesis. Faculté des sciences de Paris.

[218] Glatter, O. and Kratky, O., eds. (1982) *Small Angle X-ray Scattering.* New York, U.S.A.: London Academic Press Inc.

[219] Petoukhov, M. V., Konarev, P. V. and Kikhney, A. G. (2007) ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Journal of Applied Crystallography.* 40, pp. s232–s240.

[220] Schrödinger, L. (2009) The PyMOL Molecular Graphics System, Version 1.2r1.

[221] Delaglio, F. F. *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR.* **6**:3, pp. 277–293.

[222] Vranken, W. F. *et al.* (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics.* **59**:4, pp. 687–696.

[223] McPherson, A. J. (1976) The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. *Methods of biochemical analysis.* **23**, pp. 249–345.

[224] Gorrec, F. (2009) The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography.* **42**, pp. 1035–1042.

[225] Vonrhein, C. *et al.* (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D.* **67**:Pt 4, pp. 293–302.

[226] Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography.* **43**:1, pp. 186–190.

[227] McCoy, A. J. *et al.* (2007) Phaser crystallographic software. *Journal of Applied Crystallography.* **40**:Pt 4, pp. 658–674.

[228] Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D.* **60**:12, pp. 2126–2132.

[229] Bricogne, G. *et al.* (2011) BUSTER. Cambridge, United Kingdom: Global Phasing Ltd.

[230] Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics.* **14**:1, pp. 33–38.

[231] Bornsen, K. O., Mohr, M. D. and Widmer, H. M. (1995) Ion-Exchange and Purification of Carbohydrates on a Nafion® Membrane as a New Sample Pretreatment for Matrix-Assisted Laser-Desorption Ionization Mass-Spectrometry. *Rapid communications in mass spectrometry.* **9**:11, pp. 1031–1034.

[232] Harvey, D. J. *et al.* (2008) Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Analytical Biochemistry.* **376**:1, pp. 44–60.

[233] Harvey, D. J. *et al.* (2013) Travelling wave ion mobility and negative ion fragmentation for the structural determination of N-linked glycans. *Electrophoresis.* **34**:16, pp. 2368–2378.

[234] Domon, B. and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate Journal.* **5**, pp. 397–409.

[235] Huang, W. *et al.* (2012) Chemoenzymatic glycoengineering of intact IgG antibodies for gain of functions. *J. Am. Chem. Soc.* **134**:29, pp. 12308–12318.

[236] Ferguson, K. M. (2008) Structure-based view of epidermal growth factor receptor regulation. *Annual Review of Biophysics.* **37**, pp. 353–373.

[237] Romero, P. *et al.* (2001) Sequence complexity of disordered protein. *Proteins: Structure, Function, and Bioinformatics.* **42**:1, pp. 38–48.

[238] Romero, P., Obradovic, Z. and Dunker, A. K. (1997) Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome informatics. Workshop on Genome Informatics.* **8**, pp. 110–124.

[239] Li, X. *et al.* (1999) Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics. Workshop on Genome Informatics.* **10**, pp. 30–40.

[240] Perrakis, A. *et al.* (1994) Crystal structure of a bacterial chitinase at 2.3 Åresolution. *Structure.* **2**:12, pp. 1169–1180.

[241] Hollis, T. *et al.* (2000) The X-ray structure of a chitinase from the pathogenic fungus *Coccidioides immitis*. *Protein Science.* **9**:3, pp. 544–551.

[242] Van Aalten, D. M. *et al.* (2000) Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9 Åresolution. *Proceedings of the National Academy of Sciences.* **97**:11, pp. 5842–5847.

[243] Sun, Y. J. *et al.* (2001) The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site. *The Journal of Biological Chemistry.* **276**:20, pp. 17507–17514.

[244] Meng, G. *et al.* (2010) Structure of human stabilin-1 interacting chitinase-like protein (SI-CLP) reveals a saccharide-binding cleft with lower sugar-binding selectivity. *The Journal of Biological Chemistry.* **285**:51, pp. 39898–39904.

[245] Marino, M. *et al.* (2002) GW domains of the *Listeria monocytogenes* invasion protein InlB are SH3-like and mediate binding to host ligands. *The EMBO Journal.* **21**:21, pp. 5623–5634.

[246] Ebbes, M. *et al.* (2011) Fold and function of the InlB B-repeat. *The Journal of Biological Chemistry.* **286**:17, pp. 15496–15506.

[247] Huizinga, E. G. *et al.* (2002) Structures of glycoprotein Ibα and its complex with von Willebrand factor A1 domain. *Science.* **297**:5584, pp. 1176–1179.

[248] Price, S. R., Evans, P. R. and Nagai, K (1998) Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature.* **394**:6694, pp. 645–650.

[249] Baxter, R. H. G. *et al.* (2010) A heterodimeric complex of the LRR proteins LRIM1 and APL1C regulates complement-like immunity in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* **107**:39, pp. 16817–16822.

[250] Gaskell, A., Crennell, S. and Taylor, G. (1995) The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure.* **3**:11, pp. 1197–1205.

[251] Ito, N. *et al.* (1991) Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature.* **350**:6313, pp. 87–90.

[252] Abbott, D. W. and Boraston, A. (2011) Structural analysis of a putative family 32 carbohydrate-binding module from the Streptococcus pneumoniae enzyme EndoD. *Acta Crystallographica Section F.* **67**:4, pp. 429–433.

[253] Ficko-Blean, E. *et al.* (2012) Carbohydrate Recognition by an Architecturally Complex α-*N*-Acetylglucosaminidase from *Clostridium perfringens*. *PLoS ONE.* **7**:3, pp. e33524.

[254] Montanier, C. Y. *et al.* (2011) A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. *Journal of Biological Chemistry.* **286**:25, pp. 22499–22509.

[255] Lee, W. C. *et al.* (2006) Crystal structure and desulfurization mechanism of 2'-hydroxybiphenyl-2-sulfinic acid desulfinase. *The Journal of Biological Chemistry.* **281**:43, pp. 32534–32539.

[256] Zeng, Z. *et al.* (2011) Structural basis for Tetrahymena telomerase processivity factor Teb1 binding to single-stranded telomeric-repeat DNA. *Proceedings of the National Academy of Sciences.* **108**:51, pp. 20357–20361.

[257] Murase, K. *et al.* (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature.* **456**:7221, pp. 459–463.

[258] Esser, L. *et al.* (1998) Synapsin I is structurally similar to ATP-utilizing enzymes. *The EMBO Journal.* **17**:4, pp. 977–984.

[259] Dym, O. *et al.* (2008) Crystal structure of the Agrobacterium virulence complex VirE1-VirE2 reveals a flexible protein that can accommodate different partners. *Proceedings of the National Academy of Sciences.* **105**:32, pp. 11170–11175.

[260] Boraston, A. B. *et al.* (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**:3, pp. 769–781.

[261] Scanlan, C. N. *et al.* (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1→2 mannose residues on the outer face of gp120. *Journal of Virology.* **76**:14, pp. 7306–7321.

[262] Trkola, A. *et al.* (1996) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *Journal of Virology.* **70**:2, pp. 1100–1108.

[263] Calarese, D. A. *et al.* (2003) Antibody Domain Exchange Is an Immunological Solution to Carbohydrate Cluster Recognition. *Science.* **300**:5628, pp. 2065–2071.

[264] Funkhouser, J. D. and Aronson, N. N. (2007) Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC evolutionary biology.* **7**:96, pp. 1–16.

[265] Kelly, S. M., Jess, T. J. and Price, N. C. (2005) How to study proteins by circular dichroism. *Biochimica et Biophysica Acta.* **1751**:2, pp. 119–139.

[266] Compton, L. A. and Johnson, W. C. J. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Analytical Biochemistry.* **155**:1, pp. 155–167.

[267] Manavalan, P. and Johnson, W. C. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Analytical Biochemistry.* **167**:1, pp. 76–85.

[268] Sreerama, N. and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical Biochemistry.* **287**:2, pp. 252–260.

[269] Page, R. *et al.* (2003) Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome. *Acta Crystallographica Section D.* **59**:Pt 6, pp. 1028–1037.

[270] Garman, E. F. and Schneider, T. R. (1997) Macromolecular Cryocrystallography. *Journal of Applied Crystallography.* **30**:3, pp. 211–237.

[271] Franke, D. and Svergun, D. I. (2010) *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. *Journal of Applied Crystallography.* **42**:2, pp. 342–346.

[272] Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal.* **76**:6, pp. 2879–2886.

[273] — (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography.* **25**, pp. 495–503.

[274] Wriggers, W. (2010) Using Situs for the integration of multi-resolution structures. *Biophysical reviews.* **2**:1, pp. 21–27.

[275] Birmanns, S., Rusu, M. and Wriggers, W. (2011) Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. *Journal of Structural Biology.* **173**:3, pp. 428–435.

[276] Matthews, B. W. (1968) Solvent content of protein crystals. *Journal of Molecular Biology.* **33**:2, pp. 491–497.

[277] Rambo, R. P. and Tainer, J. A. (2010) Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Current Opinion in Structural Biology.* **20**:1, pp. 128–137.

[278] Volkov, V. V. and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *Journal of Applied Crystallography.* **36**, pp. 860–864.

[279] Trastoy, B. *et al.* (2013) Liquid-liquid diffusion crystallization improves the X-ray diffraction of EndoS, an endo-β-*N*-acetylglucosaminidase from *Streptococcus pyogenes* with activity on human IgG. *Acta Crystallographica Section F.* **69**:Pt 12, pp. 1405–1410.

[280] Kozin, M. B. and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography.* **34**:1, pp. 33–41.

[281] Bernado, P. *et al.* (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**:17, pp. 5656–5664.

[282] Comeau, S. R. *et al.* (2004) ClusPro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics.* **20**:1, pp. 45–50.

[283] Ficko-Blean, E. and Boraston, A. B. (2012) Insights into the recognition of the human glycome by microbial carbohydrate-binding modules. *Current Opinion in Structural Biology.* **22**:5, pp. 570–577.

[284] Newstead, S. L. *et al.* (2005) Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. *Acta Crystallographica Section D.* **61**:11, pp. 1483–1491.

[285] Abbott, D. W. *et al.* (2009) Streptococcus pneumoniae Endohexosaminidase D, Structural and Mechanistic Insight into Substrate-assisted Catalysis in Family 85 Glycoside Hydrolases. *Journal of Biological Chemistry.* **284**:17, pp. 11676–11689.

[286] Chalmers, M. J. *et al.* (2011) Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert review of proteomics.* **8**:1, pp. 43–59.

[287] Flot, D. *et al.* (2010) The ID23-2 structural biology microfocus beamline at the ESRF. *Journal of Synchrotron Radiation.* **17**:1, pp. 107–118.

[288] Kabsch, W. (2010) XDS. *Acta Crystallographica Section D.* **66**:Pt 2, pp. 125–132.

[289] Evans, P. R. and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallographica Section D.* **69**:7, pp. 1204–1214.

[290] Rossmann, M. G. and Blow, D. M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica.* **15**:1, pp. 24–31.

[291] Diederichs, K. and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology.* **4**:4, pp. 269–275.

[292] Weiss, M. S. and Hilgenfeld, R. (1997) On the use of the merging R factor as a quality indicator for X-ray data. *Journal of Applied Crystallography.* **30**:2, pp. 203–205.

[293] Karplus, P. A. and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science.* **336**:6084, pp. 1030–1033.

[294] Keegan, R. M. and Winn, M. D. (2007) MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallographica Section D.* **64**:1, pp. 119–124.

[295] Xu, D. *et al.* (2014) FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics.* **30**:5, pp. 660–667.

[296] Kantardjieff, K. A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science.* **12**:9, pp. 1865–1871.

[297] Evans, G. and Pettifer, R. F. (2001) CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal of Applied Crystallography.* **34**, pp. 82–86.

[298] Roversi, P., Johnson, S. and Lea, S. M. (2010) With phases: how two wrongs can sometimes make a right. *Acta Crystallographica Section D.* **66**:4, pp. 420–425.

[299] Bricogne, G. *et al.* (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallographica Section D.* **59**:Pt 11, pp. 2023–2030.

[300] Blow, D. (2002) *Outline of Crystallography for Biologists.* Oxford University Press.

[301] Abrahams, J. P. and Leslie, A. G. W. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica Section D.* **52**:1, pp. 30–42.

[302] Painter, J. and Merritt, E. A. (2006) TLSMDweb server for the generation of multi-group TLS models. *Journal of Applied Crystallography.* **39**:1, pp. 109–111.

[303] Smart, O. S. *et al.* (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallographica Section D.* **68**:Pt 4, pp. 368–380.

[304] Hendrickson, W. A. and Lattman, E. E. (1970) Representation of phase probability distributions for simplified combination of independent phase information. *Acta Crystallographica Section B.* **26**:2, pp. 136–143.

[305] Word, J. M. *et al.* (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of Molecular Biology.* **285**:4, pp. 1735–1747.

[306] Afonine, P. V. and Adams, P. D. (2012) On the contribution of hydrogen atoms to X-ray scattering. *Computational Crystallography Newsletter.* **3**, pp. 18–21.

[307] Chen, V. B. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D.* **66**:Pt 1, pp. 12–21.

[308] Theobald, D. L. and Wuttke, D. S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics.* **22**:17, pp. 2171–2172.

[309] Holm, L. and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research.* **38**, pp. W545–W549.

[310] Bhogaraju, S. *et al.* (2011) Crystal structure of the intraflagellar transport complex 25/27. *The EMBO Journal.* **30**:10, pp. 1907–1918.

[311] Wang, W. *et al.* (2012) Structural insights into SUN-KASH complexes across the nuclear envelope. *Cell Research.* **22**:10, pp. 1440–1452.

[312] Xu, Q. *et al.* (2004) Crystal structure of an allantoicase (YIR029W) from *Saccharomyces cerevisiae* at 2.4 Å resolution. *Proteins: Structure, Function, and Bioinformatics.* **56**:3, pp. 619–624.

[313] Boraston, A. B., Ficko-Blean, E. E. and Healey, M. M. (2007) Carbohydrate recognition by a large sialidase toxin from Clostridium perfringens. *Biochemistry.* **46**:40, pp. 11352–11360.

[314] Wormald, M. R. *et al.* (1997) Variations in oligosaccharide-protein interactions in immunoglobulin G determine the site-specific glycosylation profiles and modulate the dynamic motion of the Fc oligosaccharides. *Biochemistry.* **36**:6, pp. 1370–1380.

[315] Bonomelli, C. *et al.* (2011) The Glycan Shield of HIV Is Predominantly Oligomannose Independently of Production System or Viral Clade. *PLoS ONE.* **6**:8, pp. e23521.

[316] Royle, L. *et al.* (2003) Secretory IgA N- and O-glycans provide a link between the innate and adaptive immune systems. *The Journal of Biological Chemistry.* **278**:22, pp. 20140–20153.

[317] Mattu, T. S. *et al.* (1998) The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fc alpha receptor interactions. *The Journal of Biological Chemistry.* **273**:4, pp. 2260–2272.

[318] Butler, M. *et al.* (2003) Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan processing step and provides an insight into pathogenesis. *Glycobiology.* **13**:9, pp. 601–622.

[319] Lin, Z. *et al.* (2012) An N-glycosylation Analysis of Human Alpha-2-Macroglobulin Using an Integrated Approach. *Journal of Proteomics & Bioinformatics.* **5**, pp. 127–134.

[320] Kolarich, D. *et al.* (2006) Comprehensive glyco-proteomic analysis of human α1-antitrypsin and its charge isoforms. *Proteomics.* **6**:11, pp. 3369–3380.

[321] Lund, J. *et al.* (2000) Expression and characterization of truncated forms of humanized L243 IgG1. Architectural features can influence synthesis of its oligosaccharide chains and affect superoxide production triggered through human Fcgamma receptor I. *FEBS Journal.* **267**:24, pp. 7246–7257.

[322] Von Pawel-Rammingen, U. and Björck, L. (2003) IdeS and SpeB: immunoglobulin-degrading cysteine proteinases of *Streptococcus pyogenes*. *Current Opinion in Microbiology.* **6**:1, pp. 50–55.

[323] Rüdiger, S., Buchberger, A. and Bukau, B. (1997) Interaction of Hsp70 chaperones with substrates. *Nature Structural & Molecular Biology.* **4**:5, pp. 342–349.

[324] Herr, A. B., Ballister, E. R. and Bjorkman, P. J. (2003) Insights into IgA-mediated immune responses from the crystal structures of human Fc—[alpha]—RI and its complex with IgA1-Fc. *Nature.* **423**:6940, pp. 614–620.

[325] Suckow, M. A., Stevens, K. A. and Wilson, R. P. (2012) *The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents*. Academic Press.

[326] Schomaker, V. and Trueblood, K. N. (1968) On the Rigid-Body Motion of Molecules in Crystals*. *Acta Crystallographica B.* **24**:1, pp. 63–76.

[327] Feige, M. J. *et al.* Structure of the Murine Unglycosylated IgG1 Fc Fragment. *Journal of Molecular Biology.* **391**:3, pp. 599–608.

[328] Sondermann, P., Kaiser, J. and Jacob, U. (2001) Molecular basis for immune complex recognition: A comparison of Fc-receptor structures. *Journal of Molecular Biology.* **309**:3, pp. 737–749.

[329] Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A.* **32**:5, pp. 922–923.

[330] Davies, A. M., Jefferis, R. and Sutton, B. J. (2014) Crystal structure of deglycosylated human IgG4-Fc. *Molecular Immunology.* **62**:1, pp. 46–53.

[331] Ghirlando, R. *et al.* (1999) Glycosylation of human IgG-Fc: influences on structure revealed by differential scanning micro-calorimetry. *Immunology Letters.* **68**:1, pp. 47–52.

[332] Grimm, C. *et al.* (2010) A crystallization screen based on alternative polymeric precipitants. *Acta Crystallographica Section D.* **66**:Pt 6, pp. 685–697.

[333] Tsai, Y., Sawaya, M. R. and Yeates, T. O. (2009) Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta Crystallographica Section D.* **65**:9, pp. 980–988.