Unravelling the structure of glycosylated and deglycosylated immunoglobulin G antibodies

Thesis presented for the degree of
Doctor of Philosophy

Valentina Arancia Spiteri

Department of Structural and Molecular Biology
University College London

March 2021
I, Valentina Arancia Spiteri, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Abstract

Immunoglobulin G (IgG) is composed of four IgG subclasses, IgG1, IgG2, IgG3 and IgG4, which although differ in function and structure, owing to variability in hinge length, all have a conserved N-linked glycan attached in the Fc region. However, the role of this glycan on the structure, stability and function of these IgG molecules is not fully understood. The focus of this thesis is to investigate the role of the Fc-glycan in respect to IgG1, IgG3 and IgG4 by using a multidisciplinary approach to study both their glycosylated and deglycosylated forms. Primarily by probing the full-length solution structure using small angle X-ray and neutron scattering, as well as analytical ultracentrifugation. Following this extensive computational modelling methods and analysis were used to extract the theoretical models which best fit the solution structure data in order to unpick the role of this glycan.

Several studies have investigated the role of the IgG1 Fc-glycan using different structural methods, however, most of these studies investigated the Fc region of IgG rather than the full-length antibody, composed of Fab, hinge and Fc. In this thesis all experiments are conducted on full-length IgGs, presenting a complete understanding on the effect of the Fc-glycan on the entirety of the IgG structure. Studies of IgG1 and IgG4 indicate that the Fc-glycan plays a role in restricting the flexibility of the Fc. This restriction is less obvious in IgG3; this may be owing to the molecule’s elongated hinge. Of the three antibodies studied, IgG3 has the longest hinge region, composed of 62 amino acids. This long hinge has historically made it difficult to study using other structural techniques, such as X-ray crystallography and NMR, therefore the study of the IgG3 solution structure presented herein is the most complete to date and provides insight into the dynamics of the hinge region.
Impact Statement

The focus of this thesis is to unravel the role of the Fc glycan on the structure and stability of IgG1, IgG4 (Chapters 4, 5 and 6) and to determine the solution structure of IgG3 (Chapter 7). To achieve this the antibodies were deglycosylated and studied using an approach which combined analytical ultracentrifugation, small angle X-ray and neutron scattering and atomistic modelling. For IgG1 and IgG4 differences in structural dynamics were seen which demonstrated that the glycosylated forms had a more conformationally restricted Fc domain. This increased conformational lability may explain why binding to FcyRs is abrogated following glycan removal. In comparison, IgG3 had fewer observable changes in structural dynamics following glycan removal.

IgGs are relatively large flexible molecules, with molecular masses greater than 140 kDa and maximum lengths of 15 to 28 nm. This renders studying the full-length IgG structure with other structural methods, such as X-ray crystallography and nuclear magnetic resonance, very difficult. The size of IgGs also makes all-atom molecular dynamics simulations very computationally expensive. To date most studies of the Fc glycan have been limited to investigations of the Fc region alone, no doubt due to the aforementioned technical hurdles. The methods employed in this thesis have proven to be effective at unveiling conformational dynamics of the full-length IgGs and are able to unravel the role of the Fc glycan. This approach will be of use to researchers in academia and in the biotechnology industry. Especially given that the market for therapeutic monoclonal antibodies, predominantly consisting of IgG1 and IgG4, is worth tens of billions of dollars.

The full-length solution atomistic model of IgG3 presented in this thesis is the first to be described. IgG3 is still very poorly understood, due to its elongated hinge consisting of 62 amino acid residues, which makes it challenging to study. The study presented in this thesis demonstrated that the Fc is more conformationally restricted than in the other subclasses. This IgG3 solution structure also completes a series of similar studies of all members of the IgG class produced by the Perkins group, therefore a comparative analysis of these studies can be used to form a rationale for differences in receptor binding affinities.

The anticipated outputs of this work is four publications, three are currently in preparation of which one has recently been published (Spiteri et al., 2021b).
For my family.
Acknowledgements

First and foremost, I owe an enormous debt of gratitude to my supervisor, Professor Stephen Perkins, for dedicating his time and energy to my PhD and mentorship. I am so happy to have had the opportunity to learn about structural immunology and biophysics from him and could not imagine having done my PhD anywhere else. I will miss his bad jokes and random history lessons, which have been much appreciated over the years! Secondly, I would like to thank Professor Paul Dalby, for his supervision and support, and for always being the best person to take problems to, as he will undoubtedly see a solution! I would also like to thank him for allowing me to join his weekly lab meetings. The Dalby group is diverse set, with interests spanning from lasers to molecular dynamics and everything in between, and I am forever grateful for having had the opportunity to learn from them. I am incredibly grateful for the support and supervision of Dr James Doutch, especially over the COVID crisis when he insisted that we have weekly catch ups, which did plenty to keep morale and focus up! I would also like to thank Dr Robert Rambo for taking on this project and supporting it.

I would like to thank the people that I have had to pester for help over the years, notably I would like to thank Dr Joseph Curtis, despite being an ocean away, our email conversations have always been hugely informative, and I appreciate the time he takes to explain things as thoroughly as possible; he has most certainly gone above the call! Dr Jayesh Bhatt for providing computational expertise and support over the years, always armed with a piece of scrap paper to explain the theory behind molecular simulations! I would like to thank Dr Gar Kay Hui for taking the time get me settled in the lab when I first joined. Jayesh Gor for his help with analytical ultracentrifugation and Dr Kersti Karu and Dr Malgorzata Puchnarewicz for their help with mass spectrometry.

I have been so lucky to be part of two fantastic groups, with amazing scientists, who I have had the chance to learn from daily. I would like to thank members of the Perkins group, past and present, but in particular Dr Amy Osborne, Dr Henna Zahid, Bela Pereira, and Veronica Boron for their care and support over the years, but most especially over the last year. I want to thank Hina Iqbal and Xin Gao, without their support and help the last leg of this PhD would have been significantly harder. From the Dalby group I would like to thank
Dr Cheng Zhang for always being incredibly approachable and helpful. Dr Thomas McDonnell for his support and friendship over this past year and for giving me the opportunity to work on Beta-2-glycoprotein which I have enjoyed so much. Dr Nuria Codina, for her wisdom, friendship and support. Christophe Lalaurie for many conversations that I have learnt from!

I am thankful to Dr Margaret Goodall who was so gracious to provide me with IgG3 samples, that went on to be an entire chapter of this thesis. There is a wider community of physicists and physical chemists in the scattering and computational modelling world that I have had the privilege of interacting with over my PhD, I am grateful for the friendships I have formed within this community and the opportunity to learn from so many exceptional scientists. I was fortunate to serve as Faculty of Life Science student representative for two years, the experience gave me significant exposure to how a university is run, I was invited to many regular meetings with members of staff from across UCL and am so thankful for the opportunity to learn so much about leadership from them.

It is not lost on me that I was incredibly lucky to have access to two amazing departments at UCL; Structural and Molecular biology and Biochemical Engineering, interacting with so much amazing research has no doubt enhanced my perspective and will be a great foundation going forward. There are many people across Biochemical engineering and the ISMB that have enriched my experience at UCL. Notably, I would like to thank Aisha Ben-Younis for being like a sister to me over my PhD, our daily phone calls have kept me on an even keel! Dr Andrew Potterton, who I had the honour of being sat next to for most of my PhD, he has taught me so much about the computational biology/chemistry world and continues to be a great friend. Rebecca Roddan for her friendship, I am so thankful for our coffees, walks and chilly Hampstead Heath pond swims! My gratitude goes to several others including, Dr Lucas Siemons, Dr Trishant Umrekar, Dr Charlie Eldrid, Matthew Sinnott, and Emma Kennedy.

I don’t know what my PhD would have been like, had I not met my partner Liam Martin, during a departmental health and safety induction of all things! Having gone through our PhDs together, he has understood without too much explanation the ups and downs of the process and has supported me throughout.

I am fortunate to have a big Mediterranean family stocked full of uncles, aunties and cousins who are always there for me, checking in and offering
support. My grandparents, Nanna Mary, Nanna Vivi and Nannu Bertu for always encouraging and believing in me. I so wish Nannu Lino could have been here to see this, so much of the way that I think can be traced to him, and the time he invested in me growing up. I owe so much to my brother Andrea, for his support, good humour and letting me avail him of his superior coding skills in my time of need! Last but by no means the least, my mum and dad for their love and everything that they do for me, I could not have done it without you.
Table of Contents

Abstract .......................................................................................................................... 3
Impact Statement ........................................................................................................... 4
Acknowledgements ......................................................................................................... 6
Table of Contents .......................................................................................................... 9
List of Figures ................................................................................................................ 15
List of Tables ................................................................................................................ 19
Abbreviations ................................................................................................................ 21
Units ................................................................................................................................. 26
Amino Acid Abbreviations ............................................................................................ 27

Chapter 1 Immunoglobulins ......................................................................................... 28

1.1 The Immune system ................................................................................................. 29
  1.1.1 Immunoglobulins ............................................................................................... 32
  1.1.2 Immunoglobulin A ............................................................................................. 41
  1.1.3 Immunoglobulin D ............................................................................................. 42
  1.1.4 Immunoglobulin E ............................................................................................. 43
  1.1.5 Immunoglobulin M ............................................................................................ 43

1.2 Immunoglobulin G .................................................................................................... 44
  1.2.1 Immunoglobulin G1 .......................................................................................... 47
  1.2.2 Immunoglobulin G2 .......................................................................................... 47
  1.2.3 Immunoglobulin G3 .......................................................................................... 48
  1.2.4 Immunoglobulin G4 .......................................................................................... 49
  1.2.5 B lymphocyte immunoglobulin G production ................................................... 53
  1.2.6 Production of therapeutic antibodies ................................................................. 57
  1.2.7 Expression, purification and characterisation of monoclonal antibodies ............ 60

1.3 Receptors of Immunoglobulin G (FcγR) ................................................................. 62
  1.3.1 FcγRI (CD64) .................................................................................................... 69
  1.3.2 FcγRII (CD32) .................................................................................................. 71
  1.3.3 FcγRIII (CD16) ............................................................................................... 72
  1.3.4 FcRn .................................................................................................................. 74

1.4 The Complement System ......................................................................................... 75
Chapter 4 Solution structure of deglycosylated human IgG1 shows the role of C\textsubscript{H}2 glycans in its conformation ........................................ 149

4.1 Introduction ........................................................................ 150
4.2 Materials and Methods ...................................................... 155
4.2.1 Purification and composition of IgG1 ................................ 155
4.2.2 Sedimentation velocity data and analysis for IgG1 ............ 157
4.2.3 X-ray and neutron scattering data and analyses for IgG1 ...... 159
4.2.4 Atomistic modelling of IgG1 .......................................... 161
4.3 Results ............................................................................. 169
4.3.1 Purification and characterisation of glycosylated and deglycosylated IgG1 .... 169
4.3.2 Analytical ultracentrifugation of glycosylated and deglycosylated IgG1 .... 172
4.3.3 X-ray and neutron scattering of glycosylated and deglycosylated IgG1 .... 174
4.3.4 Atomistic modelling of glycosylated and deglycosylated IgG1 .......... 183
4.4 Discussion ....................................................................... 193

Chapter 5 Atomistic solution structures of glycosylated and deglycosylated human IgG4 reveal the stabilising role of its C\textsubscript{H}2 glycans ........................................................................ 198

5.1 Introduction ....................................................................... 199
5.2 Materials and Methods ...................................................... 204
5.2.1 Purification and composition of IgG4 ............................... 204
5.2.2 Sedimentation velocity data and analysis of IgG4 ............... 206
5.2.3 X-ray and neutron scattering data and analyses for IgG4 ...... 207
5.2.4 Atomistic modelling of IgG4 .......................................... 210
List of Figures

Figure 1.1 The cells of the immune system .......................................................... 30
Figure 1.2 Immunoglobulin structure ............................................................... 34
Figure 1.3 Structures of the five immunoglobulin classes ............................... 37
Figure 1.4 Antibody effector functions .............................................................. 39
Figure 1.5 Structures of the four isotypes of the immunoglobulin G .............. 46
Figure 1.6 Sequences of the immunoglobulin G heavy chain isotypes ........... 51
Figure 1.7 B lymphocyte activation and antibody production .......................... 54
Figure 1.8 Activating and inhibitory signalling pathways mediated by ITAMs and ITIMS ................................................................. 63
Figure 1.9 Extracellular domain structures of the FcγRs and FcRn .............. 66
Figure 1.10 Structures of the FcγRs and FcRn in complex with an IgG-Fc ..... 67
Figure 1.11 Schematic illustrating the classical, alternative and lectin complement pathways ................................................................. 77
Figure 2.1 Chair conformations of commonly occurring mammalian sugars .... 85
Figure 2.2 The three predominant N-inked oligosaccharide structures occurring in eukaryotes ................................................................. 87
Figure 2.3 Schematics illustrating sugar activation .............................................. 88
Figure 2.4 Diagram illustrating the biosynthesis of the N-glycan ...................... 90
Figure 2.5 Figure 2.5 Diagram of O-glycan core structures .............................. 93
Figure 2.6 IgG-Fc structures with C1r2 N-glycan ........................................... 95
Figure 3.1 Forces acting on a sedimenting molecule ......................................... 102
Figure 3.2 Absorbance and interference optical systems used in an analytical ultracentrifuge ................................................................. 107
Figure 3.3 AUC cell and sedimentation velocity experiment ......................... 110
Figure 3.4 Scattering from a particle .............................................................. 112
Figure 3.5 Beamline plan of Diamond Light Source ....................................... 118
Figure 3.6 Components of the B21 beamline ................................................. 123
Figure 3.7 Beamline plan at ISIS Neutron and Muon Source ......................... 124
Figure 3.8 SAS beamlines at ISIS Neutron and Muon Source ......................... 128
Figure 3.9 Analysis of IgG scattering curves using Guinier, distance distribution P(r) and dimensionless Kratky ................................................................. 129
Figure 5.6 Concentration dependence of the SAXS and SANS Guinier analyses. ................................................................................................................. 227
Figure 5.7 SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG4................................................................. 229
Figure 5.8 Atomistic modelling analyses for glycosylated and deglycosylated IgG4. ................................................................................................................. 232
Figure 5.9 Principal component analysis (PCA) of the best-fit glycosylated and deglycosylated IgG4 models. ................................................................. 233
Figure 5.10 Scattering curve fits to the experimental data for the best-fit model for each of the glycosylated and deglycosylated IgG4 samples ............... 235
Figure 5.11 Normalised Kratky plots for the experimental and best fit glycosylated and deglycosylated IgG4 scattering curves. ........................................... 237
Figure 5.12 Views of representative best-fit X-ray and neutron structures. .... 238
Figure 6.1 The human IgG1 and IgG4 domains and their glycosylation. ...... 246
Figure 6.2 Sequence alignment of human IgG1 and IgG4 ............................ 247
Figure 6.3 Purification, SDS-PAGE and mass spectrometry of human glycosylated and deglycosylated IgG1 and IgG4.................................................... 257
Figure 6.4 Thermal unfolding curves for glycosylated and deglycosylated IgG1 and IgG4. ................................................................................................................. 260
Figure 6.5 Thermostability parameters for IgG1 and IgG4. ......................... 262
Figure 6.6 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG1 and IgG4 at different temperatures. ..................... 267
Figure 6.7 Temperature dependence of the SAXS and SANS Guinier analyses. ................................................................................................................. 269
Figure 6.8 SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG1 and IgG4 across different temperatures. ................................................................. 275
Figure 6.9 Temperature of the peak maxima from the distance distribution analyses $P(r)$ ................................................................................................................. 276
Figure 6.10 Normalised Kratky plots for the experimental and best fit glycosylated and deglycosylated IgG1 and IgG4 scattering curves. ...................... 278
Figure 7.1 The human IgG3 domains and its glycosylation .......................... 289
Figure 7.2 Sequence alignment of IgG3 ......................................................... 292
Figure 7.3 Purification, SDS-PAGE and mass spectrometry of human glycosylated and deglycosylated IgG3................................................................. 309
Figure 7.4 Sedimentation velocity analyses of glycosylated and deglycosylated IgG3. ........................................................................................................ 310
Figure 7.5 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG3............................................................................ 313
Figure 7.6 Concentration dependence of the SAXS and SANS Guinier analyses. ................................................................................................................. 314
Figure 7.7 SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG3.................................................................. 316
Figure 7.8 Principal component analyses (PCA) and root-mean-square deviation (RMSD) of the hinge molecular dynamics simulations......................... 319
Figure 7.9 Five IgG3 start models used in torsion angle Monte Carlo (TAMC) simulations. ........................................................................................................... 321
Figure 7.10 Atomistic modelling analyses for glycosylated and deglycosylated IgG3. ........................................................................................................... 322
Figure 7.11 Principal component analyses (PCA) of the best-fit glycosylated and deglycosylated IgG3 models................................................................. 325
Figure 7.12 Scattering curve fits for the best-fit glycosylated and deglycosylated IgG3 models......................................................................................... 327
List of Tables

Table 1.1 Characteristics of the human immunoglobulin classes.......................... 40
Table 1.2 Characteristics of the isotypes of the IgG subclass................................. 50
Table 1.3 Summary table of the key characteristics of FcγRs and FcRn.............. 73
Table 3.1 Scattering lengths elements.................................................................. 115
Table 3.2 Scattering densities of biological macromolecules.............................. 116
Table 3.3 Characteristics of beamlines B21 (Diamond), Sans2d and Zoom (ISIS).................................................................................................................................................. 121
Table 4.1 Experimental data by X-ray and neutron scattering and analytical ultracentrifugation for glycosylated and deglycosylated IgG1....................... 158
Table 4.2 Modelling fits for the X-ray scattering and analytical ultracentrifugation data in light water................................................................................................................................. 165
Table 4.3 Summary of Principal component analysis for fits of X-ray scattering and analytical ultracentrifugation data in light water.............................. 166
Table 4.4 Modelling fits for the neutron scattering and analytical ultracentrifugation data in heavy water................................................................. 167
Table 4.5 Summary of Principal component analysis for fits of neutron scattering and analytical ultracentrifugation data in heavy water..................... 168
Table 5.1 Experimental data by X-ray and neutron scattering and analytical ultracentrifugation for glycosylated and deglycosylated IgG4...................... 209
Table 5.2 Modelling fits for the SAXS and AUC data in light water.................. 214
Table 5.3 Summary of Principal component analysis for fits of X-ray scattering and analytical ultracentrifugation data in light water.......................... 215
Table 5.4 Modelling fits for the neutron scattering and analytical ultracentrifugation data in heavy water................................................................. 216
Table 5.5 Summary of Principal component analysis for fits of neutron scattering and analytical ultracentrifugation data in heavy water..................... 217
Table 6.1 Experimental thermal unfolding data for glycosylated and deglycosylated IgG1................................................................. 263
Table 6.2 Experimental thermal unfolding data for glycosylated and deglycosylated IgG4................................................................. 263
Table 6.3 Experimental X-ray and neutron scattering data for glycosylated and deglycosylated IgG1................................................................. 279
Table 6.4 Experimental X-ray and neutron scattering data for glycosylated and deglycosylated IgG4................................................................. 280
Table 7.1 Experimental data by X-ray and neutron scattering and analytical ultracentrifugation for glycosylated and deglycosylated IgG3............................... 297
Table 7.2 Modelling fits for the X-ray scattering and analytical ultracentrifugation data in light water........................................................................................................... 304
Table 7.3 Principal component analysis of modelling fits for the X-ray scattering and analytical ultracentrifugation data in light water. ........................................ 305
Table 7.4 Modelling fits for the X-ray scattering and analytical ultracentrifugation data in heavy water. ................................................................................................. 306
Table 7.5 Principal component analysis of modelling fits for the X-ray scattering and analytical ultracentrifugation data in heavy water. ........................................ 307
<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation induced cytidine deaminase</td>
<td>AID</td>
</tr>
<tr>
<td>Aggregation temperature</td>
<td>$T_{agg}$</td>
</tr>
<tr>
<td>Analytical ultracentrifugation</td>
<td>AUC</td>
</tr>
<tr>
<td>Antibody-dependent cell-mediate cytotoxicity</td>
<td>ADCC</td>
</tr>
<tr>
<td>Antibody-dependent cellular phagocytosis</td>
<td>ADCP</td>
</tr>
<tr>
<td>Antigen presenting cell</td>
<td>APC</td>
</tr>
<tr>
<td>Apparent melting temperature</td>
<td>$T_{m,app}$</td>
</tr>
<tr>
<td>B cell receptor</td>
<td>BCR</td>
</tr>
<tr>
<td>Barycentric mean</td>
<td>BCM</td>
</tr>
<tr>
<td>Bruton’s tyrosine kinase</td>
<td>BTK</td>
</tr>
<tr>
<td>Buoyancy Force</td>
<td>$F_b$</td>
</tr>
<tr>
<td>c-Jun N-terminal kinase</td>
<td>JNK</td>
</tr>
<tr>
<td>Centrifugal force</td>
<td>$F_{sed}$</td>
</tr>
<tr>
<td>Charge-coupled device camera</td>
<td>CCD camera</td>
</tr>
<tr>
<td>Chinese hamster ovary</td>
<td>CHO</td>
</tr>
<tr>
<td>Circular dichroism</td>
<td>CD</td>
</tr>
<tr>
<td>Class switching recombination</td>
<td>CSR</td>
</tr>
<tr>
<td>Cluster of differentiation (16, 32, and 64)</td>
<td>CD (16, 32, and 64)</td>
</tr>
<tr>
<td>Collaborative Computational Project for Small-Angle Scattering</td>
<td>CCP-SAS</td>
</tr>
<tr>
<td>Complement receptor type</td>
<td>CR4</td>
</tr>
<tr>
<td>Complement receptor type 3</td>
<td>CR3</td>
</tr>
<tr>
<td>Complementarity determining regions</td>
<td>CDRs</td>
</tr>
<tr>
<td>Constant domain</td>
<td>C domain</td>
</tr>
<tr>
<td>Continuous distribution</td>
<td>$c(s)$</td>
</tr>
<tr>
<td>Cross-sectional radius</td>
<td>$R_{xs-1}$ and $R_{xs-2}$</td>
</tr>
<tr>
<td>D-Galactose</td>
<td>Gal</td>
</tr>
<tr>
<td>D-Mannose</td>
<td>Man</td>
</tr>
<tr>
<td>Data analysis workbench</td>
<td>DAWN</td>
</tr>
</tbody>
</table>
Density of solvent \( \rho \)
Deoxyribonucleic acid DNA
Deuterium \(^2\text{H}\)
Diacylglycerol DAG
Diamond Light Source Diamond
Differential scanning microcalorimetry DSM
Diffusion coefficient D
Dissociation constant \( K_D \)
Dolichol DoI
Endoplasmic reticulum ER
Enzyme-linked immunosorbent assay ELISA
European Bioinformatics Institute EBI
European Molecular Biology Laboratory EMBL
Fab arm exchange FAE
Fc \( \gamma \) receptor Fc\(\gamma\)R
Forward scattered intensity \( I(0) \)
Fragment crystallisable Fc
Fragment of antibody binding Fab
Frictional Force \( F_f \)
Genetic markers Gm
Glucose Glu
Glycosylphosphatidylinositol GPI
Graphical user interface GUI
Guanosine diphosphate GDP
Guanosine triphosphate GTP
Heavy chain H chain
High Performance Computer HPC
High pressure liquid chromatography HPLC
High-speed atomic force microscopy HS-AFM
Human embryonic kidney HEK
Hydrogen/deuterium exchange mass spectrometry HDX-MS
Immuno-receptor tyrosine activation motif ITAM
<table>
<thead>
<tr>
<th>Term</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immuno-receptor tyrosine inhibition motif</td>
<td>ITIM</td>
</tr>
<tr>
<td>Immunoglobulin</td>
<td>Ig</td>
</tr>
<tr>
<td>Immunoglobulin A</td>
<td>IgA</td>
</tr>
<tr>
<td>Immunoglobulin D</td>
<td>IgD</td>
</tr>
<tr>
<td>Immunoglobulin E</td>
<td>IgE</td>
</tr>
<tr>
<td>Immunoglobulin G</td>
<td>IgG</td>
</tr>
<tr>
<td>Incident Radiation</td>
<td>ki</td>
</tr>
<tr>
<td>Inositol polyphosphate 5’-phosphatase</td>
<td>SHIP</td>
</tr>
<tr>
<td>Inositol-14,5-triphosphate</td>
<td>InsP3</td>
</tr>
<tr>
<td>Intensity of scattering vector</td>
<td>I(Q)</td>
</tr>
<tr>
<td>Interferon</td>
<td>IFN</td>
</tr>
<tr>
<td>Interleukin</td>
<td>IL</td>
</tr>
<tr>
<td>Joining chain</td>
<td>J-chain</td>
</tr>
<tr>
<td>L-fucose</td>
<td>Fuc</td>
</tr>
<tr>
<td>Length of Molecule</td>
<td>L</td>
</tr>
<tr>
<td>Light chain</td>
<td>L chain</td>
</tr>
<tr>
<td>Linear accelerator</td>
<td>Linac</td>
</tr>
<tr>
<td>Lipopolysaccharide</td>
<td>LPS</td>
</tr>
<tr>
<td>Major histocompatibility complex</td>
<td>MHC</td>
</tr>
<tr>
<td>Mannose binding lectin</td>
<td>MBL</td>
</tr>
<tr>
<td>Mass spectrometry</td>
<td>MS</td>
</tr>
<tr>
<td>MBL-Associated Serine Proteases</td>
<td>MASP</td>
</tr>
<tr>
<td>Melting temperature</td>
<td>Tm</td>
</tr>
<tr>
<td>Membrane attack complex</td>
<td>MAC</td>
</tr>
<tr>
<td>Micro gram per millilitre</td>
<td>μg/ml</td>
</tr>
<tr>
<td>Molecular Weight</td>
<td>MW</td>
</tr>
<tr>
<td>Monoclonal antibodies</td>
<td>mAbs</td>
</tr>
<tr>
<td>N-Acetyl-Galactosamine</td>
<td>GalNAc</td>
</tr>
<tr>
<td>N-Acetyl-glucosamine</td>
<td>GlcNAc</td>
</tr>
<tr>
<td>N-Acetyl-Neuraminic acid</td>
<td>NeuAc</td>
</tr>
<tr>
<td>Nano gram per millilitre</td>
<td>ng/ml</td>
</tr>
<tr>
<td>Natural Killer Cell</td>
<td>NK Cell</td>
</tr>
<tr>
<td>Neonatal Fc receptor</td>
<td>FcRn</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Nuclear Factor of Activated T lymphocytes</td>
<td>NFAT</td>
</tr>
<tr>
<td>Nuclear magnetic resonance</td>
<td>NMR</td>
</tr>
<tr>
<td>Number, pressure and temperature</td>
<td>NPT</td>
</tr>
<tr>
<td>Number, volume, and temperature</td>
<td>NVT</td>
</tr>
<tr>
<td>One dimensional</td>
<td>1D</td>
</tr>
<tr>
<td>Optical Density</td>
<td>OD</td>
</tr>
<tr>
<td>Paired-distance distribution</td>
<td>$P(r)$</td>
</tr>
<tr>
<td>Partial specific volume</td>
<td>$\bar{v}$</td>
</tr>
<tr>
<td>Pattern recognition receptors</td>
<td>PRR</td>
</tr>
<tr>
<td>Peptide:N-glycosidase F</td>
<td>PNGase F</td>
</tr>
<tr>
<td>Phosphate buffered saline</td>
<td>PBS</td>
</tr>
<tr>
<td>Phosphatidylinositol (3,4,5)-trisphosphate</td>
<td>PIP$_3$</td>
</tr>
<tr>
<td>Phosphatidylinositol (4,5)-bisphosphate</td>
<td>PIP$_2$</td>
</tr>
<tr>
<td>Phosphatidylinositol-3-kinase</td>
<td>PI3K</td>
</tr>
<tr>
<td>Phospholipase Cg</td>
<td>PLC$_g$</td>
</tr>
<tr>
<td>Principal Component</td>
<td>PC</td>
</tr>
<tr>
<td>Principal Component Analysis</td>
<td>PCA</td>
</tr>
<tr>
<td>Protein Data base</td>
<td>PDB</td>
</tr>
<tr>
<td>Protein Structure file</td>
<td>PSF</td>
</tr>
<tr>
<td>Protium</td>
<td>$^1$H</td>
</tr>
<tr>
<td>Radius of Gyration</td>
<td>$R_g$</td>
</tr>
<tr>
<td>Ribonucleic acid</td>
<td>RNA</td>
</tr>
<tr>
<td>Root mean square deviation</td>
<td>RMSD</td>
</tr>
<tr>
<td>Scattered Radiation</td>
<td>$k_s$</td>
</tr>
<tr>
<td>Scattering Vector</td>
<td>$Q$</td>
</tr>
<tr>
<td>Science and Technology Facilities Council</td>
<td>STFC</td>
</tr>
<tr>
<td>Sedimentation coefficient</td>
<td>$S_{20,w}$</td>
</tr>
<tr>
<td>Size exclusion chromatography</td>
<td>SEC</td>
</tr>
<tr>
<td>Small angle Neutron scattering</td>
<td>SANS</td>
</tr>
<tr>
<td>Small angle scattering</td>
<td>SAS</td>
</tr>
<tr>
<td>Small angle X-ray scattering</td>
<td>SAXS</td>
</tr>
<tr>
<td>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
<td>SDS-PAGE</td>
</tr>
<tr>
<td>Somatic hypermutation</td>
<td>SHM</td>
</tr>
<tr>
<td>Term</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Son of sevenless</td>
<td>SOS</td>
</tr>
<tr>
<td>Static light scattering</td>
<td>SLS</td>
</tr>
<tr>
<td>Surface plasmon resonance</td>
<td>SPR</td>
</tr>
<tr>
<td>T Cell receptor</td>
<td>TCR</td>
</tr>
<tr>
<td>Target Station (1 and 2)</td>
<td>TS (1 and 2)</td>
</tr>
<tr>
<td>Three dimensional</td>
<td>3D</td>
</tr>
<tr>
<td>Time point (1, 6 and 10)</td>
<td>TP (1, 6 and 10)</td>
</tr>
<tr>
<td>Torsion Angle Monte Carlo</td>
<td>TAMC</td>
</tr>
<tr>
<td>Transforming growth factor β</td>
<td>TGFβ</td>
</tr>
<tr>
<td>Tumour necrosis factor</td>
<td>TNF</td>
</tr>
<tr>
<td>Two dimensional</td>
<td>2D</td>
</tr>
<tr>
<td>University College London</td>
<td>UCL</td>
</tr>
<tr>
<td>Uridine diphosphate</td>
<td>UDP</td>
</tr>
<tr>
<td>Uridine triphosphate</td>
<td>UTP</td>
</tr>
<tr>
<td>Variable domain</td>
<td>V domain</td>
</tr>
<tr>
<td>β2-microglobulin</td>
<td>β2m</td>
</tr>
</tbody>
</table>
### Units

<table>
<thead>
<tr>
<th>Unit</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ångstrom</td>
<td>Å</td>
</tr>
<tr>
<td>Absorbance units</td>
<td>AU</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>Degrees Celsius</td>
<td>°C</td>
</tr>
<tr>
<td>Femtosecond (10^{-15} seconds)</td>
<td>fs</td>
</tr>
<tr>
<td>Giga electron volts</td>
<td>GeV</td>
</tr>
<tr>
<td>Hour</td>
<td>h</td>
</tr>
<tr>
<td>Kilo Dalton</td>
<td>kDa</td>
</tr>
<tr>
<td>Kilo electron volts</td>
<td>KeV</td>
</tr>
<tr>
<td>Litre</td>
<td>l</td>
</tr>
<tr>
<td>Molar</td>
<td>M</td>
</tr>
<tr>
<td>Milli absorbance units</td>
<td>mAU</td>
</tr>
<tr>
<td>Milligram per millilitre</td>
<td>mg/ml</td>
</tr>
<tr>
<td>Minute</td>
<td>min</td>
</tr>
<tr>
<td>Millilitre</td>
<td>ml</td>
</tr>
<tr>
<td>Microlitre</td>
<td>l</td>
</tr>
<tr>
<td>Micromolar</td>
<td>M</td>
</tr>
<tr>
<td>Mass to charge ratio</td>
<td>m/z</td>
</tr>
<tr>
<td>Nanometre</td>
<td>nm</td>
</tr>
<tr>
<td>Nanomolar</td>
<td>nM</td>
</tr>
<tr>
<td>Revolutions per minute</td>
<td>rpm</td>
</tr>
<tr>
<td>Gas constant</td>
<td>R</td>
</tr>
<tr>
<td>Svedberg unit (10^{-13} seconds)</td>
<td>S</td>
</tr>
<tr>
<td>Second</td>
<td>sec</td>
</tr>
<tr>
<td>Volts</td>
<td>V</td>
</tr>
</tbody>
</table>
## Amino Acid Abbreviations

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Three Letter Code</th>
<th>One Letter Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Asparagine Acid</td>
<td>Asp</td>
<td>D</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
</tbody>
</table>
Chapter 1 Immunoglobulins
1.1 The Immune system

The immune system is vital in protecting multicellular organisms against a constant barrage of pathogens that can interrupt the normal function of the host. These pathogens are forever evolving, at rates faster than most complex multicellular organisms, therefore it is vital for the immune system to be able to adapt its response. Considering the range of pathogens, from tiny viruses to massive parasitic worms, that it has evolved to ward off is a testament of its impressive adaptability. However, the role of the immune system is not limited to defence; it also plays a role in maintaining proper functioning of the organism.

The mammalian immune system is divided into two main categories, the innate and adaptive response (Figure 1.1). The system’s role in defence starts once the physical barriers, i.e. the epithelial and mucosal surfaces, and chemical boundaries, such as low pH and anti-microbial proteins and peptides at these surfaces, are not sufficient to prevent entry of a pathogen into the host. The initial response is mediated by cells from the innate immune system. These include neutrophils and macrophages, which can eliminate an invading pathogen rapidly, by phagocytosing and destroying it, chemically and/or enzymatically, through releasing protein degrading enzymes, such as lysozyme. Granulocytes, including, basophils, eosinophils, mast cells and natural killer (NK) cells, degranulate when activated, in a process where secretory vesicles, referred to as granules, release cytotoxic proteins to eliminate the pathogen.

Dendritic cells, classed as antigen presenting cells (APCs), take up pathogens, process them into peptides and present them to the lymphocytes, cells of the adaptative response. These APCs display the digested peptides via the major histocompatibility complex (MHC). MHCs are divided into class I and class II. The distinction lies in slight variations in structures, the cell types that express them and which T lymphocytes subtype each class is able to recruit. Class I MHCs are located on the surface of all nucleated cells, and they present antigens located in the cytosol of the cell, such as viruses, and engage with CD8+ T cytotoxic lymphocytes. Class II MHCs are found primarily on APCs, which display antigens that they have engulfed and digested, such as bacteria. These are recognised by CD4+ T helper lymphocytes.
Figure 1.1 The cells of the immune system (legend overleaf).
Figure 1.1 The cells of the immune system (continued).

A. Illustrations of the cells of the innate response. Macrophages and neutrophils phagocytose pathogens. Vesicles containing these pathogens are referred to as phagosomes. Lysosomes, containing lysozyme, break down these pathogens once they are within the cells. Mast cells, basophils, eosinophils and NK cells release the cytotoxic contents of the granules, in process referred to as degranulation, upon activation. Dendritic cells take up antigens from the local environment and present them on MHC to T lymphocytes.

B. Illustrations of the cells of the adaptive immune response. Two main subsets of T lymphocytes are CD4+ T helper cells and CD8+ T cytotoxic cells. CD4+ T helper cells use the T cell receptor (TCR) to interact with dendritic cells via the MHC class I. These cells regulate the immune response by presenting antigens to B lymphocytes. Dendritic cells present antigens to T cytotoxic cells via the MHC class II and will eliminate pathogens found to have these same antigens. B lymphocytes bind antigens in either a T cell dependent manner, which requires antigen presentation by the T cell, or, B lymphocytes can be activated independent of T cells, whereby the B cell receptor (BCR) binds to antigens. Once the B lymphocyte is activated it will undergo somatic hypermutation (SHM), to produce antibodies with higher specificity for the antigen and differentiates into an antibody secreting plasma cell. These antibodies function to opsonise antigens, activate complement and effector cells, which will be discussed in more detail in section 1.1 of this chapter.
B lymphocytes will become activated in either a T cell dependent manner or T cell independent manner; this will be discussed further in section 1.2.5. Once activated, B lymphocytes will become differentiated into plasma cells and will secrete highly specific antibodies. The high antigen-specificity of antibodies is achieved through a process of somatic hypermutation (SHM), where the variable domains of the antibody genetically readjust to become more specific for the pathogen. Antibodies can eliminate a pathogen via several mechanisms, such as opsonisation, activating complement and effector cells, which will all be discussed in greater detail in section 1.1 of the chapter. These cells need to be activated, differentiated and proliferated, which can take several days. The adaptive immune system has a “memory” of past infections, conferred through the presence of long-lived cells that remain dormant. These cells are reactivated in the event of reinvasion whence they will now be able to rapidly mount a highly specific attack (Chaplin, 2010) (Owen et al., 2013).

1.1.1 Immunoglobulins

Immunoglobulins (Ig), also commonly referred to as antibodies, are antigen receptors of the adaptive immune response produced by B lymphocytes. An antigen is any material not produced by the host, i.e. foreign, such as a virus. Antibodies were first recognized in 1890 by immunologists Emil von Behring and Shibasaburō Kitasato. In their work, they immunized animals with attenuated forms of infectious pathogens, Clostridium tetani and Corynebacterium diphtheriae, that caused tetanus and diphtheria. Sera from these infected animals were collected and injected into unimmunized animals. They then infected these animals. They found that animals that had been injected with the serum showed no symptoms of an infection, and the control group, that had not been injected, had died. This work was the first to illustrate that there are molecules in serum that can offer protection, that arises after an infection and can be transferred between species to offer passive immunity. For this, Emil von Behring was awarded the first Nobel Prize in Physiology or Medicine, in 1901 (Owen et al., 2013).

Following this Paul Ehrlich described the first model of an antibody, in his side-chain theory, where he postulated antibodies or “receptors”, are branched
and contain multiple sites for antigen binding and complement activation (Davies et al., 1993). In 1959 Gerald Edelman and Rodney Porter published the molecular structure of antibodies and they were awarded the Nobel Prize in 1972 (Edelman, 1959; Porter, 1959). The first atomic structure of an antibody fragment was determined in 1972 (Inbar et al., 1972). In 1975 Georges Köhler and César Milstein developed a method to produce monoclonal antibodies (Köhler et al., 1975), that paved to the way for therapeutic antibodies to become a multi-billion dollar industry.

Immunoglobulins take on a characteristic Y-shaped structure that consists of two fragment of antigen binding (Fab) regions, which binding to antigens in a very specific manner, and the fragment crystallisable (Fc) region, which binds effector ligands to activate effector functions that enable the clearance of the antigen. The antibody is composed of four polypeptide chains, two heavy chains (H) and two light chains (L) (Figure 1.2A). As their names suggest, the H chains have a greater molecular weight and are longer than the L chains. Each H chain is paired with a L chain and connected via interchain disulphide bonds. The H and L chains each have one variable (V) domain, in addition, the L chain has one constant (C) domain and the H chain has up to four C domain. The V domains are divided into three regions of sequence hypervariability, called the complementarity determining regions (CDRs). The CDRs are followed by, four regions of conserved sequence, referred to as the framework regions. The antigen binding site is formed when three CDRs of the H chain are paired with three CDRs of the L chain. All domains take on a classic immunoglobulin fold, which consists of two twisted anti-parallel β-sheets, which are held together by hydrophobic interactions and intrachain disulphide bonds. (Figure 1.2B) (Schroeder et al., 2010). There are two types of immunoglobulin folds present in immunoglobulins, the V type, for the V domains and the C1 type for the constant domains (Sun et al., 2001).

The L chain C domain has two main sequence types, κ (kappa) and λ (lambda), which can be subdivided into λ1, λ2, λ3, and λ4. In humans approximately 60% of L chains are κ and 40% are λ (Owen et al., 2013). There are also five main types of H chain C domains, also varying in sequences, named α (alpha), δ (delta), ε (epsilon), γ (gamma) and μ (mu),
Figure 1.2 Immunoglobulin structure (legend overleaf).
Figure 1.2 Immunoglobulin structure (continued).

A. Schematic of the characteristic antibody structure, composed of two H chains, in dark pink, and L chains, in light pink. The Fabs are composed of two V domains and two C domains. The Fc comprises of two to three C domains, depending on the Ig class see Figure 1.3.

B. Schematic of the V type domain immunoglobulin fold organisation and VL domain for IgG1 crystal structure (PDB ID: 1HZH) (Saphire et al., 2001), which is labelled in a manner that corresponds to the schematic, showing two views (180°).

C. Schematic of C1 type domain immunoglobulin fold organisation. Corresponding to the schematic are two views (180°) of the CL domain from the same IgG1 crystal structure as in B (PBD ID:1HZH).

D. Crystal structure of IgG1 (PDB ID:1HZH). The Fc glycans are shown in blue sticks.

The B and C schematics are adapted from (Sun et al., 2001).
referred to as isotypes, giving rise to the different antibody classes, IgA, IgD, IgE, IgG, and IgM, respectively. The first to separate immunoglobulin classes were Arne Tiselius and Elvin Kabat in 1938 (Tiselius et al., 1939). Each of these classes take on different global structures and play different roles in the mammalian immune system. One notable difference is that in IgA, IgD and IgG, the H chains are joined together at a ‘hinge’ region through disulphide bonds, whereas in IgM and IgE, in lieu of a hinge, an additional C domain is present (Figure 1.3 and Table 1.1).

Another key structural feature of immunoglobulins, and the focus of this thesis, is the presence of conserved N-linked glycans in the Fc region of antibodies. In IgA, IgD and IgG, these glycans are located at the C\(H_2\) domain. In IgM and IgE they are located at the equivalent C\(H_3\) domain. These glycans play an important structural and functional role, however continued research to elucidate the role of antibody glycosylation in function is needed, to continue to improve the engineering of antibodies as therapeutics. Glycosylation and antibody glycosylation will be explored in more detail in Chapter 2 of this thesis.

Antibodies can mediate several effector functions that protect the host against a pathogen. These include neutralisation, when the antibody binds to the host-interaction site on the pathogen, blocking its ability to bind to other host proteins and thus inhibits its entry into cells. A second is opsonisation, derived from the Greek word opsōnein, which translates to “prepare for eating” was coined by Almroth Wright (Wright et al., 1904). This describes the process where the antibody binds to the pathogen and is then bound by Fc receptors on phagocytic cells, which will then phagocytose the pathogen. The phagocytic process is referred to as antibody-dependent cellular phagocytosis (ADCP). Complement fixation, is a process where an antibody-antigen complex is bound to complement components and subsequently the pathogen is either phagocytosed (via cells expressing C3 receptors), or lysed (as a result of pore formation by the complement components C7, C8, and C9). However, not all antibodies can carry out this function. The third and final effector function is, antibody-dependent cell-mediated cytotoxicity (ADCC). Antibody-antigen complexes bind to Fc receptor on cytotoxic cells including, natural killer (NK) cells and granulocytes, which will destroy the antigen, through the release of cytotoxic granules that

36
Figure 1.3 Structures of the five immunoglobulin classes (legend overleaf).
**Figure 1.3 Structures of the five immunoglobulin classes (continued).**
For all schematics, the V and C domains, H and L chains are labelled. O-linked glycans are shown in red lines with a circle, and N-linked glycans are shown in blue lines with a circle.

A. Schematic of IgG, which has three C domains, and two V domains. A hinge connects the CH2 and CH3 domains, with disulphide bonds connecting the H chains. The number of disulphide bonds varies between IgG isotypes (Figure 1.5).

B. Schematic of IgD, which has a long hinge with several O-linked glycans, as well as tail pieces connected to the CH3 domain.

C. Schematic of IgE, where each H chain has an additional CH3 domain because IgE lacks a traditional hinge.

D. Schematic of IgA, which typically takes on a dimeric form, connecting two monomers via the tail pieces, joined by a joining (J)-chain. The IgA hinge has several O-linked glycans.

E. Schematic of IgM, which typically takes on a pentameric form, connected at the tail pieces connect and J-chain. IgM does not have a traditional hinge but has a CH3 domain in its place.
The schematic outlines the main antibody functions, which are neutralisation of pathogens, ADCC, opsonisation and complement activation. Adapted from (Owen et al., 2013).
<table>
<thead>
<tr>
<th>Ig Subclass</th>
<th>Heavy Chain</th>
<th>Molecular Forms</th>
<th>Serum Concentration (%)</th>
<th>Function</th>
<th>Complement Activation</th>
<th>FcR Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgA1</td>
<td>α1</td>
<td>Monomeric</td>
<td></td>
<td>Mucosal</td>
<td>-</td>
<td>FcαR</td>
</tr>
<tr>
<td>IgA2</td>
<td>α2</td>
<td>Dimeric, Secretory</td>
<td>15%</td>
<td>Immunity</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>IgD</td>
<td>δ</td>
<td>Monomeric</td>
<td>&lt;0.5%</td>
<td>Homeostasis</td>
<td>-</td>
<td>FcδR</td>
</tr>
<tr>
<td>IgE</td>
<td>ε</td>
<td>Monomeric</td>
<td>&lt;0.01%</td>
<td>Allergy</td>
<td>-</td>
<td>FcεR</td>
</tr>
<tr>
<td>IgG1</td>
<td>γ1</td>
<td>Monomeric</td>
<td>67%</td>
<td>Secondary immune response, toxins and virus neutralisation</td>
<td>+</td>
<td>FcγR I, II, III</td>
</tr>
<tr>
<td>IgG2</td>
<td>γ2</td>
<td>Monomeric</td>
<td>22%</td>
<td></td>
<td>+</td>
<td>FcγR II</td>
</tr>
<tr>
<td>IgG3</td>
<td>γ3</td>
<td>Monomeric</td>
<td>7%</td>
<td></td>
<td>+</td>
<td>FcγR I, II, III</td>
</tr>
<tr>
<td>IgG4</td>
<td>γ4</td>
<td>Monomeric</td>
<td>4%</td>
<td></td>
<td>-</td>
<td>FcγR I, II</td>
</tr>
<tr>
<td>IgM</td>
<td>μ</td>
<td>Pentameric</td>
<td>10%</td>
<td>Primary immune response</td>
<td>+++</td>
<td>FcμR</td>
</tr>
</tbody>
</table>

Table 1.1 Characteristics of the human immunoglobulin classes. Adapted from (Schroeder et al., 2010).
contain perforin and granzymes, or trigger apoptosis of virally infected cells (Figure 1.4) (Owen et al., 2013). Immunoglobulins are of interest to the scientific community because of their ability to very specifically bind antigens. This unique feature has been studied and exploited for the generation of therapeutics over the last several decades.

1.1.2 Immunoglobulin A

IgA makes up two-thirds of the body’s antibody production, and is predominantly found at mucosal surfaces, for example the intestines and lungs. It is also the main Ig class found in secretions, such as breast milk. It is found in serum at a concentration of about 2-3 mg/ml. There are two subclasses of IgA: IgA1, which is found in serum, and IgA2 found in secretions. The main structural differences between the subclasses is that IgA1 has a more extended hinge. IgA exists in three molecular forms; monomeric (IgAm), found in serum, dimeric (dIgA), where the two monomers are joined by a J-chain and secretory IgA (S-IgA), which is a combination of dIgA and a secretory component (Woof et al., 2006). A J-chain is a small polypeptide, with a molecular weight of about 15 kDa, that regulates the polymer formation of IgA and IgM (Johansen et al., 2000).

IgA1 has two conserved N-linked glycans in the H chain at Asn263 and Asn459. IgA2 has two-three additional N-linked glycans (Mattu et al., 1998). These N-linked glycans are predominantly a biantennary complex-type (Field et al., 1994). IgA1 has eighteen potential, O-linked glycosylation sites, nine on each H chain, however the occupancy of these sites and structures of these glycans differs depending on which tissue/cell type is expressing the IgA (Mattu et al., 1998). Furthermore, the IgA1 hinge region alone contains three to five O-linked glycans, whereas IgA2 does not have any carbohydrates attached to its hinge (Torano et al., 1977). The variations in the structure of these glycans are thought to contribute to autoimmune disease pathology, such as in IgA nephropathy (Tomana et al., 1997). IgA serves a vital function in the mucosal immunity, by controlling the mucosal microbiota and dampening the inflammatory immune response. IgA lacks the necessary binding site to interact with C1q and activate the classical complement pathway. However, there are reports of complement activation via the mannose-binding lectin (MBL) pathway (Roos et al., 2001). There is also some evidence for polymeric IgA activating the alternative
complement pathway, however, the methods used to demonstrate this seem to perturb the system significantly and therefore further studies are required to validate these findings (Hiemstra et al., 1987) (Woolf et al., 2011). Some studies have reported that IgA-antigen complexes can inhibit complement activation induced by IgG and IgM (Griffiss et al., 1983; Russell et al., 1989). Thus far there are no high-resolution full-length structures of IgA available, however, there have been several low-resolution solution structures reported (Furtado et al., 2004; Bonner et al., 2009; Hui et al., 2015). Electron microscopy structures of IgA have been also published (Parkhouse et al., 1971), including electron microscopy studies of IgA-antigen complexes (Munn et al., 1971).

1.1.3 Immunoglobulin D

IgD has been somewhat of an enigma. It is evolutionary conserved from organisms as biologically divergent as bony fish to humans and has likely been a key player in immunity since the inception of the adaptive immune system (Ohta et al., 2006). However, the roles of IgD are still being elucidated. IgD occurs in two monomeric forms: membrane bound IgD found on the surface of mature B lymphocytes, and acts as an antigenic receptor; and a secreted form (secIgD), that makes up about 3 μg/ml of all immunoglobulin content in the serum (Rogentine et al., 1966). B lymphocytes co-express IgD and IgM, with the same VH and VL domains, thus with identical antigenic specificity.

Structurally, IgD is the only human immunoglobulin in which the H chains are held together by one disulphide bond (Takahashi et al., 1982). This is especially unusual as it also has the longest hinge composed of 64 amino acids. Additionally, like IgA1, the hinge is glycosylated, with up to seven potential O-linked glycosylation sites. Additionally, IgD has three N-linked glycosylation sites on each H chain. At the C-terminus of the H chain there is a tail piece composed of seven to eight amino acids. The structure of IgD has been determined using a solution scattering approach, however, as yet, a full length high resolution structure has not been published (Sun et al., 2005).

Functionally IgD, has been poorly understood, although recent advances have started to shed light on this. IgD plays a role in enhancing mucosal homeostasis by monitoring the commensal microbiota at mucosal surfaces
(Gutzeit et al., 2018). Furthermore, it mediates immune surveillance by binding pathogenic mucosal antigens, and subsequently interacting with basophils and mast cells via the Fc region. This interaction serves to stimulate the production of IgE and IgG as well as amplifying T Helper cell responses, which will have a high affinity for the antigen targeted by IgD. Interestingly, basophil-bound IgD was found to reduce IgE degranulation and thus dampening down the immune response (Shan et al., 2018).

### 1.1.4 Immunoglobulin E

IgE is well characterised as mediating allergic reactions, through the development of IgE with specificity for generally innocuous antigens, resulting in pathologies such as food allergies. To this end, anti-IgE therapies to regulate conditions such as severe allergic asthma have been approved (Busse et al., 2001). The role IgE is thought to have evolved to protect against parasites and venoms (Mukai et al., 2016). IgE activates basophils and mast cells via the FcεRs, which phagocytose IgE-bound antigens. It is the least abundant Ig in serum with a concentration of about 100 ng/ml. Additionally, IgE does not activate the complement pathways (Oettgen, 2016).

There is no full-length high-resolution structure of IgE available however, there are several IgE-Fc structures, largely in complex with receptors or antigens; (Wan et al., 2002; Gould et al., 2003; Dhaliwal et al., 2012; Pennington et al., 2016; Doré et al., 2017; Chen et al., 2018; Jabs et al., 2018). A remarkable feature of the IgE-Fc is its high degree of flexibility, adopting both bent and elongated conformations. IgE has seven N-linked glycosylation sites on each H chain (Arnold et al., 2004).

### 1.1.5 Immunoglobulin M

IgM is unique in its ability to form hexameric or pentameric structures, held together by a J-chain (Munn et al., 1971). IgM plays a critical role in the initial defence against pathogens (Ochsenbein et al., 1999) and cancerous cells (Vollmers et al., 2006). Due to its frontline role, IgM has a low antigenic affinity. This is owing to the B lymphocytes that express IgM having not yet undergone affinity maturation in response to the antigen. Despite this, IgM is still highly
effective at clearing pathogens. This is due to the prevalence of IgM, where its serum concentration is between 0.5-2.00 mg/ml, and the enhanced ability of oligomeric IgM to bind C1q, and activate the classical complement pathway (Feinstein et al., 1986). As oligomeric IgM has a superior ability to activate complement, when compared to IgG, there is some interest in developing IgM-based immunotherapy (Samsudin et al., 2020).

Monomeric IgM is comprised of 14 domains across the four polypeptide chains. IgM, like IgA has a short tail piece at the C-terminus of the H chain, which is thought to be critical in establishing the correct geometry for the oligomeric forms of IgM (Pasalic et al., 2017). Early electron microscopy-derived structures postulated that the pentamer takes on a structure with a central Fc5 disc and five Fab’2 external arms and once the monomers are bound the pentamer is stabilised in a five-legged “staple-like” or “table-like” conformation (Feinstein et al., 1969). More recent electron microscopy-derived structures have been able to solve the J-chain in a IgM-Fc pentamer (Li et al., 2020). Solution scattering has been used in order to examine the structure of oligomeric IgM, from which a model for pentameric IgM was created (PDB ID: 2RCJ) (Perkins et al., 1991). IgM has five N-linked glycosylation sites per H chain, including one at the tail piece peptide (Arnold et al., 2005).

1.2 Immunoglobulin G

IgG is the best characterized Ig class; this is due to a wide and ever-growing interest in exploiting IgG for immunotherapies. As of March 2021, there were 89 IgG based therapies approved for use in the European Union and the United States (up-to-date records are maintained by Antibody Society and can be found at this weblink: https://www.antibodysociety.org/resources/approved-antibodies/) (Reichert, 2021). IgG is the most abundant Ig class in human serum making up 10-20% of total plasma protein.

The IgG structure is composed of 12 domains (Figure 1.5). The L chain is approximately 25 kDa and the H chain is approximately 50 kDa. The H chain has a variable domain (VH) and three constant chains (CH1, CH2 and CH3). The L chain has a variable domain (VL) and a constant domain (CL). The Fab regions are composed of VL, CL, VH and CH1 domains, and the Fc region is composed of
two C\(\text{H}2\) and two C\(\text{H}3\) domains. The Fab and Fc regions are linked together by a hinge region, composed of the upper hinge, middle hinge and the lower hinge. The residues closest to the lower hinge region at the C\(\text{H}2\) domain are responsible for the IgG effector functions, as it contains the overlapping binding sites for C1q and F\(\gamma\)R. A conserved structural feature of the C\(\text{H}2\) domain is an N-linked glycan at Asn\(^{297}\), whose functional role is not clear.

There are four subclasses of IgG namely, IgG1, IgG2, IgG3 and IgG4, whose order represents their abundance in serum (Figure 1.5). These subclasses are highly conserved with approximately 90% sequence similarity. The primary differences are found at the hinge and the upper C\(\text{H}2\) domain (Figure 1.6). Due to these differences each of the subclasses has a varying capacity to initiate effector functions, including complement activation, phagocytosis and ADCC. This is based on their binding affinity to different F\(\gamma\)Rs and their propensity to bind C1q. One of the primary factors that contributes to IgG affinity to bind ligands is hinge flexibility. IgG3 is the most flexible followed by IgG1, IgG4 and IgG2 in that order (Roux et al., 1997). Moreover, the IgG-Fc region can bind the neonatal Fc receptor (FcRn), conferring a longer half-life to IgG in addition to allowing for transportation to the placenta and mucosal surfaces (Vidarsson et al., 2014).

Additional variability to the IgG subclasses comes from allelic variation in the H chain genes, and this gives rise to allotropes. This was first observed when human sera would agglutinate erythrocytes coated with “incomplete” anti-Rh antibodies. These allotropes differ between individuals and population groups. Different allotropes can be identified as non-self and induce an immunogenic response, which can be the case in blood transfusions or during pregnancy. There is also the concern of whether or not monoclonal antibody therapies can be immunogenic to people if the allotype is not considered (Jefferis et al., 2009). The nomenclature of these allotypes is composed of the genetic marker (Gm) and subclass. For example, IgG1 would be G1m. Additionally, this would then be followed by the allotype number or letter, for instance, G1m1. Variations
The isotypes in the IgG subclass are all composed of H chains with one V domain and three C domains. The L chains possess one V domain and one C domain. The isotypes differ in the number of amino acid residues in the hinge and disulphide bonds between H chains. IgG1 has two disulphide bonds, IgG2 has three, IgG3 has eleven and IgG4 has two.

Figure 1.5 Structures of the four isotypes of the immunoglobulin G.

and three C domains. The L chains possess one V domain and one C domain. The isotypes differ in the number of amino acid residues in the hinge and disulphide bonds between H chains. IgG1 has two disulphide bonds, IgG2 has three, IgG3 has eleven and IgG4 has two.
in the κ L chain are denoted as Km (Jefferis et al., 2009). For each subclass a number of allotypes have been identified. For both IgG1 and IgG2 four allotypes have been identified in each. For IgG4 no serologically relevant allotypes have been identified thus far. IgG3 is particularly polymorphic, with several variants coding for significant amino acid changes that can result in functionally different molecules (Vidarsson et al., 2014).

1.2.1 Immunoglobulin G1

IgG1 is the most abundant IgG in serum, with a concentration of 5–11 mg/ml (Nirula et al., 2011). The IgG1 hinge is 15 amino acids long and is very flexible (Vidarsson et al., 2014). IgG1 is able to activate complement through C1q and FcγR binding which results in ADCC and ADCP. IgG1 differs to the other IgG subclasses in the position of the C_H1 cysteine which forms a disulphide bonds with C_L. For IgG1 this C_H1 cysteine is at Cys^{220}, however in other IgG subclasses it is Cys^{131}. These two cysteine positions are spatially juxtaposed and result in differences in function (Vidarsson et al., 2014).

Thus far, there has been one full-length high-resolution IgG1 X-ray crystal structure published (PDB ID: 1HZH) (Saphire et al., 2001). However, due to the inherent flexibility of the IgG1 structure, this crystal structure only offers information on one possible conformation. Previous solution scattering and molecular modelling studies on the IgG1 structure, illustrated that IgG1 can adopted stable conformations that accommodate C1q and FcγR binding without steric clashes (Rayner et al., 2015). Solution structures of IgG1 with and without glycans will be explored in Chapter 4 and the role of glycosylation on thermostability will be explored in Chapter 6 of this thesis.

1.2.2 Immunoglobulin G2

IgG2 responds to bacterial capsular polysaccharide antigens (Siber et al., 1980). IgG2 has the shortest hinge in the IgG class, composed of 12 amino acids and lacking one of the double glycines found at positions 235-6. Additionally, the IgG2 hinge is also incredibly rigid due to the poly-proline helix, which is stabilised by up to four disulphide bonds (Vidarsson et al., 2014). So far, nine antibody therapies based on IgG2 have been developed (Reichert, 2021).
There are three isoforms of IgG2; IgG2A, IgG2A/B and IgG2B, that vary in the disulphide connectivity at the lower Fab region. This disulphide connectivity is dynamic in vivo, with IgG2A being the initial form produced by the B lymphocytes. This is rapidly converted to IgG2A/B while circulating in blood, and then is eventually converted fully to IgG2B. This disulphide reshuffling indicates the protein’s age and alters its function (Liu et al., 2008). Additionally, IgG2 monomers can form covalent dimers (Yoo et al., 2003).

A solution scattering study combined with atomistic modelling has served to develop a structural understanding of why IgG2 is able to bind FcγRII/FcγRIII but unable to bind FcγRI (Hui et al., 2019). Additionally, a three-dimensional structure of human myeloma IgG2 has been obtained by electron microscopy, differential scanning microcalorimetry (DSC) and fluorescence. When combined, these latter studies demonstrated that one of the Fab arms obstructs the Cn2 domain of the Fc region (Ryazantsev et al., 2013).

1.2.3 Immunoglobulin G3

IgG3 is a potent pro-inflammatory antibody, however, it has a shorter half-life than other subclasses. This is due to its susceptibility to proteolysis, possibly due to its elongated hinge. Owing to this, IgG3 has not been considered an ideal therapeutic candidate (Saito et al., 2019). IgG3 has the longest hinge in the IgG class, composed of up to 62 amino acids, including 21 prolines and 11 cysteines, which form disulphide bonds (Lakbub et al., 2016). The hinge takes on a poly-proline helix structure, with limited flexibility. However, the length of the hinge varies between allotypes of IgG3. Due to its length, the Fab fragments are relatively far away from the Fc region, allowing them to be more flexible. This extended hinge is the result of duplications of a hinge exon, which is generally only encoded by one exon in other IgG subclasses, but in IgG3 there are up to four exons encoding for the hinge (Vidarsson et al., 2014). Serum-derived and monoclonal human IgG3 was found to have O-linked glycosylation sites, with approximately 10% and 13% occupancy rates, respectively, in the hinge region (Plomp et al., 2015).

Electron microscopy images of myeloma IgG3 molecules have demonstrated that the hinge is in a compact state (Ryazantsev et al., 1990).
There is no full length crystal structure of IgG3 as yet, however some solution studies have suggested that the Fab regions in wild type IgG3 are directed away from the Fc plane (Lu et al., 2007). A structure for the IgG3 Fc is available (PDB ID: 5W38) (Shah et al., 2017) but there is no structure for the hinge, and no structures for human IgG3 Fabs. Chapter 7 of this thesis will focus on solution scattering and atomistic modelling of IgG3 to address this lack of knowledge.

1.2.4 Immunoglobulin G4

IgG4 is generally associated with allergy pathology (Devey et al., 1976). Its normal function is considered to be anti-inflammatory by blocking the function of other antibodies, namely IgE, in order to mitigate the proinflammatory effects of repeated exposure to an antigen (Aalberse et al., 2009). IgG4 does not bind to C1q, and therefore is not able to activate the classical complement pathway. Moreover, relative to the other IgG subclasses, IgG4 has a low affinity for most FcγRs. Due to its reduced effector capacity, IgG4 can be a desirable therapeutic, in cases where proinflammatory effector functions are problematic. There are currently 15 approved IgG4 therapies, as of March 2021 (Reichert, 2021).

The IgG4 hinge is 12 amino acids in length, therefore is shorter than IgG1. The flexibility of the IgG4 hinge lies between that of IgG1 and IgG2 (Roux et al., 1997). There are two different isomers of IgG4 that differ in the disulphide bonding of hinge cystines. One isomer has the normal inter-H chain disulphide bonding, forming covalently linked half-molecules, and the other has intra-H chain disulphide bonding, i.e. non-covalently linked half-molecules. IgG4 is a highly dynamic molecule and can exchange Fab arms by swapping a H chain and attached L chain (half-molecule) with a half-molecule from another antibody molecule, resulting in bispecific molecules. This process is known as Fab arm exchange (FAE) (van der Neut Kolfschoten et al., 2007) and is thought to be mediated by local redox conditions in vivo (Rispens et al., 2011). The resulting bispecific IgG4 molecules are unable to effectively cross-link antigens and has lower avidity because multivalent target binding is not possible. Mutating Ser228 to Pro, abolishes the formation of intra-chain disulphide bonding, making the IgG4 molecule IgG1-like (Silva et al., 2015).
Table 1.2 Characteristics of the isotypes of the IgG subclass.
Adapted from (Vidarsson et al., 2014), which complied receptor K\textalpha values from (Bruhns et al., 2009).

<table>
<thead>
<tr>
<th></th>
<th>IgG1</th>
<th>IgG2</th>
<th>IgG3</th>
<th>IgG4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Molecular mass</strong></td>
<td>146</td>
<td>146</td>
<td>170</td>
<td>146</td>
</tr>
<tr>
<td>(kDa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amino acids in</strong></td>
<td>15</td>
<td>12</td>
<td>62</td>
<td>12</td>
</tr>
<tr>
<td>hinge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hinge disulphide</strong></td>
<td>2</td>
<td>4</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>bonds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average adult</strong></td>
<td>7.00</td>
<td>3.80</td>
<td>0.51</td>
<td>0.56</td>
</tr>
<tr>
<td>serum concentration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mg/ml)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Half-life (days)</strong></td>
<td>21</td>
<td>21</td>
<td>7-21</td>
<td>21</td>
</tr>
<tr>
<td><strong>Response to:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteins</td>
<td>++</td>
<td>+/-</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>+</td>
<td>+++</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Allergens</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td><strong>Fc\gamma R K\alpha values</strong></td>
<td>x10^6 M^{-1}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fc\gamma R I</td>
<td>65</td>
<td>-</td>
<td>61</td>
<td>34</td>
</tr>
<tr>
<td>Fc\gamma R IIa</td>
<td>3.50-5.20</td>
<td>0.10-0.45</td>
<td>0.89-0.91</td>
<td>0.17-0.21</td>
</tr>
<tr>
<td>Fc\gamma R IIb</td>
<td>0.12</td>
<td>0.02</td>
<td>0.17</td>
<td>0.20</td>
</tr>
<tr>
<td>Fc\gamma R IIc</td>
<td>0.12</td>
<td>0.02</td>
<td>0.17</td>
<td>0.20</td>
</tr>
<tr>
<td>Fc\gamma R IIIa</td>
<td>1.20-2.00</td>
<td>0.03-0.07</td>
<td>7.70-9.80</td>
<td>0.25</td>
</tr>
<tr>
<td>Fc\gamma R IIIb</td>
<td>0.20</td>
<td>-</td>
<td>1.1</td>
<td>-</td>
</tr>
<tr>
<td><strong>FcRn (at pH &lt;6.5)</strong></td>
<td>+++</td>
<td>+++</td>
<td>+/+/+</td>
<td>+++</td>
</tr>
<tr>
<td><strong>C1q binding</strong></td>
<td>++</td>
<td>+</td>
<td>+++</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 1.6 Sequences of the immunoglobulin G heavy chain isotypes (legend overleaf).
Figure 1.6 Sequences of the immunoglobulin G isotypes (continued).
Sequence alignment of IgG1 (Uniprot ID: P01857), IgG2 (Uniprot ID: P01859), IgG3 (Uniprot ID: P01860) and IgG4 (Uniprot ID: P01860). Sequences were aligned using Clustal Omega (Sievers et al., 2011). The sequences have been numbered according to EU numbering. Highlighted in yellow is the conserved IgG glycosylation site at Asn\textsuperscript{297}. Beneath the alignments, consensus symbols indicated the degree of conservation, where (*) indicates full conservation, (:) indicates conservation between groups of strongly similar properties based on the Gonnet PAM 250 matrix, (.) indicates conservation between groups of weakly similar properties, and a space indicates no conservation.

A. Sequences corresponding to the C\textsubscript{H}1 domain.
B. Sequences corresponding to the hinge.
C. Sequences corresponding to the C\textsubscript{H}2 domain.
D. Sequences corresponding to the C\textsubscript{H}3 domain.
There have been two high-resolution full-length structures published to date (PDB ID: 5DK3) (Scapin et al., 2015) and (PDB ID: 6GFE) (Blech et al., 2019). Previous solution scattering structures and atomistic modelling have also been produced (Rayner et al., 2014). Further solution scattering work highlighting the implications of glycosylation on the structure of IgG4 will be examined in Chapter 5 and Chapter 6.

1.2.5 B lymphocyte immunoglobulin G production

The primary function of B lymphocytes is to secrete Ig. This is achieved by a process of clonal selection, activation, proliferation and deletion. Work in understanding antibody production was contributed significantly by Neils Jerne, David Talmage, Sir Peter Medawar, and Sir Frank MacFarlane Burnet. Medawar and Burnet were jointly awarded the Nobel Prize for Physiology or Medicine in 1960 for their discovery of “acquired immunological tolerance”. Jerne was awarded a Nobel prize with Georges J.F. Köhler and César Milstein in 1984 for “theories concerning the specificity in development and control of the immune system and the discovery of the principle for production of monoclonal antibodies”.

Burnet’s clonal selection hypothesis proposed that the receptors found on leukocytes and the antibodies produced by these same cells had the same antigen-binding specificities. Additionally, he suggested that a single B lymphocyte can develop into a clone, that has the same antigen-binding specificity as the first B lymphocyte and that these clones would migrate to, replicate and function within the secondary lymphoid organs. That these lymphoid organs harbour progeny cells from these B lymphocytes, which remain here to fend off future infections by the same pathogen, serving a memory function. He also imagines some ‘randomisation’ of the globulin molecule, which was later to become known as somatic hypermutation. He also mentions the need for clonal deletion, to eliminate any cells that may arise with specificity for self-antigens. All this was captured in one fell swoop in Burnet’s three-page seminal paper,
Figure 1.7 B lymphocyte activation and antibody production (legend overleaf).
Figure 1.7 B lymphocyte activation and antibody production (continued).

A. Schematic of type 1 T-independent B lymphocyte activation. With low antigen concentrations, antigens bind to BCRs. With high antigen concentrations, antigen binds PRRs of B-1 B cells, a subset of B lymphocytes. B-1 B cells do not undergo SHM and secrete low-affinity IgM.

B. Schematic of type 2 T-independent B lymphocyte activation. Antigen cross-links CD21 and BCRs and are able to recruit complement components C3d and C3dg. Cytokines released by macrophages and dendritic cells drive antibody production.

C. Schematic of T dependent B lymphocyte activation. Antigen binds BCRs on B-2 B cells and cross-links them. The T cell CD40L binds to B cell CD40. Antigen is also presented to CD4+ T helper cell via an interaction between the MHC and TCR. Activated B-2 B cells can migrate to the lymph nodes or the spleen, where they form clusters of activated B cells, called primary foci. These cells differentiate into a plasma cells, that secrete IgM. Some of the plasma cells will later die, while a few will remain in the bone marrow for a prolong period of time. Activated B-2 B cells can also migrate to follicles within lymphoid tissues. Here these cells will undergo a process of SHM and CSR and then differentiate into plasma cell that secretes antibodies with a high affinity for the antigen. Some of these plasma cells will remain in the lymphoid tissues as memory cells.

Adapted from (Owen et al., 2013).
“A modification of Jerne’s theory of antibody production using the concept of clonal selection” (Burnet, 1976), which would go on to underpin our understanding of immune tolerance (Owen et al., 2013).

There are two main B-cell responses, elicited depending on the type of antigen (Figure 1.7). One response is the T-independent response (TI), that is mainly mediated by the B-1 B cells and the marginal zone B-cell subtypes. This response is initiated by T-independent antigens, that are divided into two types. The first type, referred to as TI-1 antigens, are mitogenic antigens such as lipopolysaccharide (LPS). When bound these antigens induce B lymphocyte proliferation. At high antigenic concentrations this response occurs through the binding of the antigen to pattern recognition receptors (PRRs) on B lymphocytes. At low antigenic concentrations, B lymphocytes are activated through interactions with the B cell receptors (BCRs). B-1 B lymphocytes are predominantly activated in a TI manner. These cells are considered ‘innate-like’, as they predominantly secrete low-affinity IgM and they do not undergo SHM.

Another class of antigens are TI-2 antigens. These are generally highly repetitive antigens, such as bacterial capsular polysaccharides. These antigens are capable of cross-linking receptors found on the B lymphocyte cell surface, including CD21 and BCRs, allowing B lymphocyte activation. They are able to bind with complement components C3d and C3dg (Section 1.4). TI-2 antigens are assisted by monocytes, macrophages and dendritic cells, that release cytokines that drive antibody production in these B lymphocytes.

The second response requires the contribution of CD4+ T-helper cells. This response is referred to as a T-dependent (TD) response, where B-2 B cells bind to TD antigens. B-2 B cells are the predominant subset of B lymphocytes. This response starts when an antigen binds and cross-links BCRs and the T lymphocyte CD40L binds the B lymphocyte via its CD40 receptor. Once the B lymphocyte is activated, via the TD response, it will migrate to special lymphoid regions in the lymph nodes or the spleen. Here, they will form clusters of activated B lymphocytes, known as primary foci, where these activated B lymphocytes differentiate into plasma cells. These plasma cells will initially release an abundance of IgM. Some of these cells die after the initial response but a few will remain in tissues such as the bone narrow for prolonged residence. Some activated B lymphocytes will not enter the primary foci but will instead enter the
follicles in the lymphoid tissue. Here, they undergo additional differentiation, where they will change morphology and form germinal centres. These cells undergo a process of SHM, in which the variable domains undergo mutagenesis, which alters the sequence in the antigen-combining site. This increases the antibody’s affinity for the antigen. These higher affinity antibodies will be in circulation within 6-10 days after the initial infection. In addition, CD4+ T helper cells, influence B lymphocytes to undergo class switch recombination (CSR) in order to produce antibody classes, other than IgM. Both SHM and CSR are mediated through activation induced by cytidine deaminase. After the primary immune response, memory B lymphocytes will reside in the lymphoid tissue and retain these mutated BCRs. Upon a second infection these memory cells will be able to secrete these high affinity antibodies at a faster rate than in the first infection (Owen et al., 2013).

### 1.2.6 Production of therapeutic antibodies

Monoclonal antibodies (mAbs) are a major class of therapeutics. The production of mAbs on an industrial scale has been made possible by the hybridoma technique (Köhler et al., 1975), which produces pure antibodies, in large amounts. Currently, there are several methods of producing therapeutic antibodies, including the use of mouse hybridoma to achieve chimeric or humanised mAbs, phage display, transgenic mice, or singe B cell methods, that produce fully human mAbs. There are a few key biochemical and biophysical characteristics to be considered in producing mAbs, including their antigen specificity and affinity, antibody stability, and immunogenicity.

The first approved antibody therapy was that for muromonab-CD3, produced using a hybridoma method. Muromonab-CD3 is a mouse monoclonal IgG2A, the suffix ‘-monab’, indicates its murine origin. Chimeric antibodies were the first iteration on the hybridoma method. These were advanced through a series of developments. The first was to create clones of the murine V_H and V_L domains (Orlandi et al., 1989), the second was the ability to express the L and H chains in a stable human cell line (Neuberger, 1983; Williams et al., 1986; Beidler et al., 1988), and finally the ability to make antibodies that possessed murine V_H and V_L chains fused with human C domains (Boulianne et al., 1984; Morrison et
al., 1984). Chimeric antibodies retain 30-35% of the murine sequence, which can result in increased immunogenicity in human (An, 2009). The first chimeric IgG on the market was rituximab (Reff et al., 1994). Chimeric antibodies have the suffix ‘-ximab’.

Humanised antibodies are created by grafting non-human CDRs onto a human Ig framework in order to maintain antigen specificity (Jones et al., 1986). The development of humanised antibodies was vital in accelerating the use of antibodies as therapeutics, as they were better clinically tolerated for reason of their lower immunogenicity, when compared to chimeric antibodies. Humanised antibodies were advantageous to the hybridoma method because, despite the low cost and faster production associated with mouse mAbs, the use of mouse antibodies resulted in patients producing anti-mouse antibodies, rendering the therapy immunogenic. Moreover, the mouse Fc region is not able to effectively initiate the human ADCC response. Thus, humanisation of antibodies circumvented these issues (Lu et al., 2020). The first approved humanised mAb was daclizumab in 1997 (Vincenti et al., 1998). One of the most renowned humanised mAb is trastuzumab, used in the treatment of breast cancer, approved in 1998 (Baselga et al., 1998). Humanised mAbs, are given the suffix ‘-zumab’.

The next generation of therapeutic antibodies were fully human antibodies. This was achieved in a number of ways. One method was phage display, a tool that allows for the selection of mutants. Diverse exogenous genes are incorporated into filamentous bacteriophages creating a library, in which the bacteriophages display the corresponding protein, which can be screened against other proteins and the best binding ones can then be isolated. Enrichment of the library is an iterative process that mimics natural selection, by allowing for the generation and isolation of binders high affinity and specificity (McCafferty et al., 1990). The first phage display generated therapy to enter the market was adalimumab (Den Broeder et al., 2002). Human antibodies are given the suffix, ‘-umab’. Work on phage display was awarded a Nobel prize in 2018, that was shared between George P. Smith (Smith, 1985) and Gregory P. Winter.

Another method of generating human antibodies is to create transgenic ‘humanised’ mice, that have been engineered to replace murine antibody genes with human genes. The advantage to using mice is that they are easy to immunise, they do not tolerate most human antigens, and their B lymphocytes
form stable hybridomas (Green et al., 1994; Lonberg et al., 1994). Additionally, mice will undergo the natural affinity maturation process in vivo in response to an antigen. The first therapy produced in transgenic mice, was panitumumab (Yang et al., 2001), using a transgenic mouse called XenoMouse® (Jakobovits, 1998). However, this requires finding an antigen that can illicit an immunogenic response, which is not a limitation when using phage libraries. Moreover, using transgenic mice requires the creation of new libraries for each target, which is also not a limitation when using phage libraries.

Single B cell methods can also be used to produce human antibodies. A technique for immortalising human B cells using Epstein-Barr virus was created. This method has some limitations, predominantly that it is inefficient in some patients and it can be difficult to maintain the transformed clones (Traggiai et al., 2004). However, it offers some advantages to the transgenic mice, as the immune reactivity is triggered more robustly in humans than in mice, which is critical in infectious diseases. Additionally, single B cell technologies only require a few cells, making isolation of potential mAbs far easier. The advantages of B cell technologies over phage display is that H and L chain pairing are preserved, while this can be random in phage display. These randomly paired mAbs can have decreasing binding affinity or develop self-reactivity when transferred from single chain variable fragment (scFv) to intact IgG (Lu et al., 2020). Antigen specific B cells can be identified in a number of ways, in a process called antigen baiting, including antigen conjugated fluorescent beads (Correa et al., 2018), antigen-coated magnetic beads (Lundkvist et al., 1993). Antigen baiting is a preliminary selection tool, after which cloning of each H chain and L chain is performed and expressed in other mammalian cell lines such as Chinese hamster ovary (CHO) cells.

Humanised antibodies may lose antigen affinity in the humanisation step (Makabe et al., 2008). For mAb based therapies good affinity is considered about 1 nM or less for the target antigen. Moreover, antibody engineering, such as point mutations to increase stability, may result in lower affinities. Therefore, further affinity maturation may be required, by generating large libraries of randomly mutated CDR or V domain sequences, from which higher affinity mutants are selected. Another procedure could be to create a small library through targeted mutagenesis that attempts to mimic in vivo affinity maturation, for which the high
affinity mutant is selected and randomised at CDR hotspots. Combinations of different mutations may have an additive effect in increasing affinity (Lu et al., 2020).

1.2.7 Expression, purification and characterisation of monoclonal antibodies

Mammalian cell culture is one of the preferred platforms for the expression of mAbs. The advantages of this platform are speed, flexibility, mammalian post-translation modifications, and sustained high yields in stable cell lines. CHO and non-NS0 cells are used for the production of biopharmaceuticals. These cell lines have been significantly engineered in order to achieve maximum stability and yield. Mammalian cell lines, such as human embryonic kidney (HEK)-293, are often used for transient expression at the trial stages of drug development, when significant Ig quantities are required, but the intensive engineering of a dedicated cell line is not warranted at such an early stage. Transient expression involves the introduction of DNA into the cells via transfection reagents. Transfection reagents are necessary because mammalian cells do not readily integrate exogenous DNA, therefore cationic polymers, cationic lipids, or calcium phosphate allow for efficient uptake. Vectors for H and L chains can be plasmid DNA and viral vectors. Viral vectors can be harder to construct, so may not be used at the initial trial phases.

Expression can take place at a range of scales, from static cell culture to thousand litre stirred bioreactors with adapted suspension cultures in batch or fed-batch mode. Bioreactors come in all shapes and sizes, can be fixed or disposal, depending on the production and economic objectives of the producer. All variables in this bioreactor such as dissolved oxygen, pH, CO₂ and more can be closely controlled in order to limit batch-to-batch variability and to achieve maximum yield.

Once the antibodies have been expressed, modified (in the endoplasmic reticulum (ER) and Golgi apparatus) and subsequently secreted into the medium, they can be recovered and purified. Cells can be removed by batch centrifugation, continuous centrifugation or microfiltration, all of which results in a clarified conditioned medium. Following this, the medium is concentrated,
generally by ultracentrifugation, with a 30 kDa molecular weight cut-off membrane. This retains the 150 kDa mAbs while bulk medium filters through the membrane. Purification protocols can differ depending on the antibody. Predominantly this starts with affinity chromatography such as protein A or protein G, both of which have a high affinity for IgG-Fc regions. Another priority is to remove endotoxin, a pyrogen and component of the gram-negative cell wall. Endotoxin is highly negatively charged and can therefore be removed using anion exchange chromatography. Once the mAbs have reached maximum purity the mAbs undergo buffer exchange to remove harsh buffers used in purifications. mAbs are generally exchanged into phosphate buffered saline for analytical use and into histidine-based buffers for in vivo or cell assays. To achieve buffer exchange, the procedure used is similar to that of concentrating samples. This time fresh formulation buffer is added sequentially, topping up the buffer volume to the sample original volume at each buffer addition.

Characterisation considerations are dose, purity, identity such as antibody modifications (predominantly glycosylation), aggregation and thermal stability, and potency. Dose and concentration can be achieved using UV spectroscopy, colorimetric assay, and size exclusion chromatography (SEC). Purity can be determined using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), and western blot. The identity and mass of purified mAbs can be done using mass spectrometry, cation exchange chromatography and capillary isoelectric focusing, reverse-phase high pressure liquid chromatography (HPLC). Aggregation and thermal stability can be measured using SEC, dynamic light scattering, and capillary differential scanning calorimetry. Finally, potency can be measured using assays such as enzyme-linked immunosorbent assay (ELISA) and surface plasmon resonance (SPR) (An, 2009).
1.3 Receptors of Immunoglobulin G (FcγR)

The different antibody classes and subclasses achieve their diverse effector functions through their ability to bind to specific FcRs located on cell surfaces. FcRs, as the name suggests, binds specifically to the Fc region of the Ig and were first identified in 1960 (Boyden et al., 1960). The mechanism of antibody binding, i.e. through the Fc region, was described in 1965 (Berken et al., 1966). FcRs are type I transmembrane glycoproteins, except for FcγRIIIb, which is glycosylphosphatidylinositol (GPI)-anchored. FcRs require several antigen-bound antibodies that form immune complexes, in order to be cross-linked and activated. The diverse effector functions that are mediated by antibody binding of FcRs are determined by several factors namely, the antibody class and subclass, the FcR, the cell type on which the FcR is located (NK cell, T and B lymphocyte, macrophage, etc) and finally the signalling pathway triggered within the cells (Figure 1.8) (Nimmerjahn et al., 2008).

The structures of human FcRs are composed of two extracellular domains (D1 and D2), the exception being FcγRI, which possess an extra domain (D3). These domains take on a classic immunoglobulin fold and have accessory signal-transducing polypeptides, embedded in the cell membrane. Additionally, they are associated with intracellular signalling proteins that contain immunoreceptor tyrosine activation motif (ITAM) or immunoreceptor tyrosine inhibition motif (ITIM). As their names imply these signal proteins initiate an activating or inhibiting signalling cascade (Owen et al., 2013).

ITAM signalling starts by recruiting the tyrosine kinase LYN, which phosphorylates tyrosine residues on the ITAM, which recruits another tyrosine kinase SYK and thus activates son of sevenless homologue (SOS) or phosphatidylinositol-3-kinase (PI3K). SOS goes on to activate RAS which goes on to activate downstream signalling pathways. PI3K phosphorylates phosphatidylinositol (4,5)-bisphosphate (PIP2), generating phosphatidylinositol (3,4,5)-trisphosphate (PIP3), this recruits Bruton’s tyrosine kinase (BTK) and phospholipase Cγ (PLCγ). BTK goes on to activate RAC, which activates RAS homologue, RHO, to activate downstream signalling pathways. PLCγ will activate diacylglycerol (DAG), which goes on to activate PKC that activates other downstream signalling pathways.
Figure 1.8 Activating and inhibitory signalling pathways mediated by ITAMs and ITIMs (legend overleaf).
**Figure 1.8 Activating and inhibitory signalling pathways mediated by ITAMs and ITIMS (continued).**

**A.** Schematic of activating signalling mediated by ITAMs. Signalling pathway starts by recruiting LYN, which phosphorylates tyrosine residues on the ITAM, which recruits SYK, which activates SOS or PI3K. SOS goes on to activate RAS which goes on to activate downstream signalling pathways. PI3K phosphorylates PIp<sub>2</sub> generating PIp<sub>3</sub>, this recruits BTK and PLC<sub>γ</sub>. BTK goes on to activate RAC, which activates RAS homologue, RHO, this activates downstream signalling pathways. PLC<sub>γ</sub> will activate DAG, which goes on to activate PKC that activates downstream signalling pathways. PLC<sub>γ</sub> also activates InsP<sub>3</sub> that stimulates the release of Ca<sup>2+</sup> from the ER. Ca<sup>2+</sup> modulates the effect of calmodulin and calcineurin, which dephosphorylate the transcription factor NFAT. NFAT can now enter the nucleus and bring about gene activation and subsequent protein expression. The activation of these downstream signalling pathways cumulates in cellular responses such as ADCC, phagocytosis, cytokine release or oxidative burst.

**B.** Schematic of inhibitory signalling mediated by ITIMs. ITIMs recruit phosphatases, that dephosphorylate the structures that where phosphorylated in activating signalling. The signalling cascade starts with the phosphorylation of the ITIM by LYN, that acts as a docking site for SHIP, which bind to the ITIM at the SH2 domain. SHIP hydrolyses PIp<sub>3</sub> to PIp<sub>2</sub>, which prevents the recruitment of BTK and PLC<sub>γ</sub>. In some instances, ITIM activation results in B lymphocyte apoptosis, this process is SHIP independent, and instead involves cABL apoptosis, this process is SHIP independent, and instead involves cABL kinase, which phosphorylates the ITIM, recruiting BTK, that activates c-Jun N-terminal kinase (JNK).

Adapted from *Nimmerjahn et al., 2008*. 
PLC\(\gamma\) also activates inositol-14,5-triphosphate (InsP3) that stimulates the release of Ca\(^{2+}\) from the ER. Ca\(^{2+}\) modulates the effect of calmodulin and calcineurin, which dephosphorylate the transcription factor Nuclear Factor of Activated T lymphocytes (NFAT). NFAT can now enter the nucleus and bring about gene activation and subsequent protein expression. The activation of these downstream signalling pathways cumulates in cellular responses such as ADCC, phagocytosis, cytokine release or oxidative burst (Nimmerjahn et al., 2008).

ITIMs recruit phosphatases, that function by dephosphorylating the structures that where phosphorylated in activating signalling. The signalling cascade starts with the phosphorylation of the ITIM by LYN, that in turn serves as a docking site for inositol polyphosphate 5'-phosphatase (SHIP), which bind to the ITIM at the SH2 domain. SHIP hydrolyses PIP\(_3\) to PIP\(_2\), which prevents the recruitment of BTK and PLC\(\gamma\). In some instances, ITIM activation results in B lymphocyte apoptosis, this process is SHIP independent, and instead involves cABL kinase (Nimmerjahn et al., 2008). ITIMs are vital as they set the threshold for activation, preventing cells from being activated by low-level signals. Inhibitory and activating receptors can be co-expressed on the same cells. For example, anti-inflammatory cytokines such as transforming growth factor (TGF)\(\beta\) and interleukin (IL)-4, stimulate the up-regulation of inhibitory FcRs and proinflammatory cytokines. Tumour necrosis factor (TNF) and interferon (IFN)\(\gamma\) up-regulate the expression of some activating FcRs (Owen et al., 2013).

There are three main Fc\(\gamma\)Rs families in humans, Fc\(\gamma\)RI (CD64), Fc\(\gamma\)RII (CD32), and Fc\(\gamma\)RIII (CD16), that bind IgG antibodies (Figure 1.9). These three receptors vary in their affinity for the different IgG subclasses, with Fc\(\gamma\)RI being the highest affinity receptor. The other Fc\(\gamma\)Rs have a distinctly lower affinity, and this may be to prevent endemic antibodies from non-specifically activating proinflammatory responses. Structural evidence suggests that mechanistically the interaction between IgG and Fc\(\gamma\)Rs has 1:1 stoichiometry (Figure 1.10) (Sondermann et al., 2000; Radaev, Motyka, et al., 2001; Lu, Chu, et al., 2015). The interface of Fc binding to FcRs occurs via the lower hinge and C\(_\text{H}2\) domain, with binding being asymmetric. The major contact points in all solved Fc\(\gamma\)R-Fc complexes appear to
Figure 1.9 Extracellular domain structures of the FcγRs and FcRn.

A. Extracellular domain structure of FcγRI, which has three extracellular domains, D1, D2 and D3. Glycans are shown in green sticks. (PDB ID: 3RJD) (Lu et al., 2011).

B. Extracellular domain structure of FcγRII, which has two extracellular domains, D1, D2. (PDB ID: 3RY5) (Ramsland et al., 2011).

C. Extracellular domain structure of FcγRIII, which has two extracellular domains, D1, D2. (PDB ID: 3SGJ) (Ferrara et al., 2011).

D. Extracellular domain structure of rat FcRn, which has three extracellular domains, α1, α2 and α3, and an associated β2 microglobulin domain. (PDB ID: 1EXU) (Burmeister et al., 1994).
Figure 1.10 Structures of the FcγRs and FcRn in complex with an IgG-Fc (legend overleaf).
Figure 1.10 Structures of the FcγRs and FcRn in complex with an IgG-Fc (continued).

A. Extracellular domain structure of FcγRI in complex with IgG1 Fc. (PDB ID: 4X4M) (Lu et al., 2015).
B. Extracellular domain structure of FcγRII in complex with IgG1-Fc. (PDB ID: 3WJJ) (Mimoto et al., 2013).
C. Extracellular domain structure of FcγRIII in complex with IgG1-Fc. (PDB ID: 3AY4) (Mizushima et al., 2011).
D. Extracellular domain structure of rat FcRn in complex IgG-Fc. (PDB ID: 1FRT) (Burmeister et al., 1994).

For all, glycans are shown in green sticks. All are shown in two views (90° or 180° rotations).
be within the same regions on both the Fc and the FcγR. In the Fc the FG loop in the C\text{H}2 domain makes contacts with the receptor D1-D2 domain linker and D2 BC loop through a hydrophobic proline sandwich interaction, in which Pro\textsuperscript{329} in the Fc is positioned between two Trp residues from the receptors. These Trp residues are conserved in the FcγRs, however the residue adjacent to the proline sandwich varies between receptors, Arg\textsuperscript{102} in FcγRI, Ser\textsuperscript{88} in FcγRII and Ile\textsuperscript{88} in FcγRIII. This may account for variations in affinity. A second interaction site is found between the receptor D2 domain C and C’ strands and the Fc region BC and DE loops. This interaction is mediated through salt bridges, hydrogen bonds and van der Waals interactions (Davies et al., 2015).

Another proposed mechanism for binding is via an angular change between the C\text{H}2 and C\text{H}3 domains. This provides an explanation for allosteric interactions between different receptors, for example, if an antibody is already in complex with an FcRn molecule it could not bind another FcγR. However, this proposed angular change has not been observed experimentally (Sondermann et al., 2000). A limiting factor in understanding the mechanisms for subclass receptor affinity and binding is that structural studies are mostly focussed on Fc regions from IgG1 molecules. Despite being able to make comparisons between the subclasses using sequence data, further structural studies of these receptors in complex with Fc domains from the other subclasses will undoubtedly be of benefit to the field.

Different cell types express different FcγRs, and effector cells can co-express activating and inhibitory FcγRs. However, some cells do not co-express these receptors. For example, NK cells only express activating FcγRIII, where B lymphocytes only express inhibitory FcγRIIb. FcγR expression is not limited to cells of the immune system but are also expressed by endothelial cells, microglial cells, osteoclasts and mesangial cells. The functional role that these receptors play when expressed by these cells is still poorly understood.

1.3.1 FcγRI (CD64)

FcγRI is the high affinity receptor, which binds IgG1 and IgG3 with higher affinity that the other subclasses. This receptor has an additional immunoglobulin-like domain (D3), which is not present in the other FcγRs. The
D3 domain has been postulated to be the reason behind the higher affinity of the FcγRI (Allen et al., 1989). FcγRI is a 72 kDa transmembrane glycoprotein, that preferentially recruits IgG1, IgG3 and IgG4, but does not recruit IgG2. There are seven potential N-linked glycosylation sites in the FcγRI extracellular domain (Patel et al., 2019). These interactions have a high affinity ($K_D 10^{-9} - 10^{-10}$ M) (Kiyoshi et al., 2015). A crystal structure of the extracellular domains of FcγRI is available (PDB ID: 3RJD) (Lu et al., 2011). As well as crystal structures of FcγRI in complex with IgG1-Fc (PDB ID: 4X4M) (Lu, Chu, et al., 2015), (PDB ID: 4W4O) (Kiyoshi et al., 2015), and (PDB ID: 4ZNE) (Oganesyan et al., 2015).

These structures help to develop an understanding of mode of binding, however there are some discrepancies between the structures. The primary difference is that Lu et al. (2015) found that there is a greater emphasis on the role of Fc-glycan recognition for high affinity binding, whereas Kiyoshi et al. (2015) reported little interaction between the Fc-glycan and the receptor, with only three residues making contact with the Fc-glycan. This paucity of contacts with the Fc-glycan was confirmed by Oganesyan et al. (2015). It is vital to understand that the differences between these complexes may be due to slight differences in sequences of FcγRI. In particular Kiyoshi et al. (2015) and Lu et al. (2015) used an FcγRI mutant with 19 mutations in the sequence that improved expression and thermostability. Moreover, FcγRI is a glycosylated protein, however, the Lu et al. (2015) FcγRI structure does not have any glycans present. This may have contributed to the differences in reported Fc-glycan contribution to the interaction between the Fc and the receptor.

When the FcγRI-Fc complexes are compared with the Fc in complex with other FcγRs, the FcγRI-Fc complexes had the largest buried surface area between antibody and receptor, this being over 2100 Å². Additionally, Arg102, which is adjacent to the proline sandwich, is able to hydrogen bond with Pro329 in the IgG1-Fc Cε2 FG loop (Davies et al., 2015). The D3 domain is thought to act as a spacer, that helps to position the Fab regions on the cell surface to enable effective FcγRI-Fc binding (Kiyoshi et al., 2015). In all solved complexes the D3 domain does not appear to directly interact with the Fc region.
1.3.2 FcγRII (CD32)

FcγRII can mediate both activating and inhibitory effector functions. Several isoforms have been identified, FcγRIIa, FcγRIIb and FcγRIIc. These receptors are predominantly membrane bound. FcγRII has two to three N-glycosylation sites in the extracellular domains (Patel et al., 2019). The extracellular domains are conserved between isoforms, consequently the main source of diversity between them is due to differences in the transmembrane and cytoplasmic domains, which give rise to variations in the response. In addition to the main isoforms, FcγRIIa and FcγRIIb have additional variants, namely there are two forms of FcγRIIa, FcγRIIa1 and FcγRIIa2, and FcγRIIb is likewise expressed as FcγRIIb1 and FcγRIIb2 (Nimmerjahn et al., 2008).

FcγRIIa triggers proinflammatory effects. It is unique to primates and is the most abundant FcγR, being expressed by platelets, placental endothelial cells, dendritic cells, eosinophils, basophils, neutrophils and mast cells. FcγRIIa is the only FcγR that is expressed in platelets. It is a 40-60 kDa glycoprotein and is unique amongst FcRs in its ability to form dimers. FcγRIIa has a low affinity for IgG with a Kₐ of < 10⁶ M⁻¹. IgG1 and IgG3 are bound preferentially, and while IgG2 and IgG4 can still bind FcγRIIa, this interaction is less favourable. Signalling is mediated through an ITAM. FcγRIIa is thought to play a role in destructive immunity. Studies with mice transgenic for human FcγRIIa, showed a sensitivity to pathogenic antibodies and developed autoimmunity (Sardjono et al., 2003). A structure of IgG1-Fc in complex with FcγRIIa is available (PDB ID: 3RY6) (Ramsland et al., 2011). There are several unbound structures available, (PDB IDs: 1FCG, 1H9V, 3RY4, 3RY5) (Maxwell et al., 1999; Sondermann et al., 2001; Ramsland et al., 2011).

FcγRIIb is the main inhibitory FcγR, which when engaged inhibits the activity of the cells it is expressed on. This inhibitory activity is mediated by an ITIM domain. FcγRIIb is the only receptor present on B lymphocytes and when it cross-links with a neighbouring BCR, the activation threshold is increased, and antibody production is suppressed. This receptor is also expressed on other immune cells, including dendritic cells, macrophages, neutrophils, mast cells and basophils. FcγRIIb1 is expressed predominantly on B lymphocytes and cannot effectively endocytose cross-linked immune complexes. FcγRIIb2 is expressed
by monocyte, dendritic cells and macrophages and can effectively endocytose immune complexes upon cross-linking. FcγRIlb3 is the soluble form of this receptor, as it lacks the transmembrane domain and inhibits the presentation of IgG-antigen complexes (Smith et al., 2010). Studies have found that impaired FcγRIlb is implicated with a higher propensity of developing autoimmune pathologies, suggesting that it plays a role in immune tolerance. Two crystal structure of the IgG1-Fc in complex with FcγRIlb are available (PDB ID: 3WJJ and 3WJL) (Mimoto et al., 2013), as well as a crystal structure of the unbound receptor (PDB ID: 2FCB) (Sondermann et al., 1999).

FcγRIIc is expressed on 7-15% of NK cells and B lymphocytes. It has an equal affinity for IgG1 and IgG3, followed by IgG4 and has the lowest affinity for IgG3. It has two glycosylation sites (Patel et al., 2019). FcγRIIc signals via an ITAM. The extracellular domains of FcγRIIa and FcγRIIc are 95% identical (Anania et al., 2019). NK cells expressing FcγRIIc has increased ADCC upon receptor cross-linking (Metes et al., 1998). There is a paucity of structural information on FcγRIIc, with no structures available.

1.3.3 FcγRIII (CD16)

FcγRIII is the lowest affinity receptor, with an affinity range of $K_a$ 0.03-9.8-10$^6$ M$^{-1}$ (Table 1.2). It exists in two isoforms FcγRIIIa and FcγRIIIb, which share 96% sequence similarity in their extracellular binding regions. Crystal structures of FcγRIII in complex with IgG1-Fc is available (PDB ID: 1T89) (Radaev, Motyka, et al., 2001), (PDB ID: 1E4K) (Sondermann et al., 2000). The complexes illustrated that receptor binding eliminated the dyad symmetry for the Fc region, forcing the structure into an asymmetric interface, whereby identical residues on either chain of the hinge interacts with different and unrelated parts of the bound receptor. This asymmetric binding excludes the possibility for a second receptor to bind. Moreover, receptor binding stabilises the lower hinge in both complexes (Radaev et al., 2002). An unbound structure is also available (PDB ID: 1FNL) (Zhang et al., 2000).

FcγRIIIa is predominantly expressed on the cell membrane of NK cell, as well as macrophages, mast cells and basophils. FcγRIIIa can be shed from NK
Table 1.3 Summary table of the key characteristics of FcγRs and FcRn.

Adapted from (Bruhns et al., 2015; Pyzik et al., 2015; Patel et al., 2019).

<table>
<thead>
<tr>
<th>Cellular expression</th>
<th>FcγRI (CD64)</th>
<th>FcγRIIa (CD32a)</th>
<th>FcγRIIb (CD32b)</th>
<th>FcγRIIc (CD32c)</th>
<th>FcγRIIa (CD16a)</th>
<th>FcγRIIb (CD16b)</th>
<th>FcRn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monocyte</td>
<td>Activating</td>
<td>Activating</td>
<td>Inhibitory</td>
<td>Activating</td>
<td>Activating</td>
<td>Inhibitory</td>
<td>IgG recycling and antigen presentation</td>
</tr>
<tr>
<td>Macrophage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dendritic cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mast cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutrophil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eosinophil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macrophage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dendritic cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mast cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutrophil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monocyte</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NK cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NK cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monocyte</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basophil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eosinophil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutrophil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monocyte, macrophage, dendritic cells, endothelium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IgG affinity</th>
<th>IgG1=IgG3</th>
<th>IgG1=IgG3&gt;lgG3</th>
<th>IgG1=lgG3</th>
<th>IgG1=IgG3</th>
<th>IgG3&gt;lgG1&gt;lgG2</th>
<th>IgG3&gt;lgG1=lgG4</th>
<th>IgG1=lgG2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;lgG4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;lgG2 =IgG4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>=lgG4&gt;lgG2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>=lgG4&gt;lgG2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;lgG4=lgG2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>=lgG4&gt;lgG3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
cells, forcing the NK cell to disengage from the immune synapse and this reinitiates the cell’s ability to perform ADCC. FcγRIIIa has the highest affinity to IgG3, followed by IgG1, and about equal affinity to IgG4 and IgG2. FcγRIIIa is glycosylated at five sites. Glycoforms play a significant role in its affinity for antibody binding, indeed certain glycoforms, in particular the afucosylated glycoform, can make its IgG affinity akin to that of CD64 (Subedi et al., 2019). A study found that repeated binding of FcγRIIIa decreased the amount of perforin secreted but shedding of FcγRIIIa allowed for replenishment of perforin (Srpan et al., 2018). FcγRIIIa has been studied in complex with IgG1 using high-speed atomic force microscopy (HS-AFM) and hydrogen/deuterium exchange mass spectrometry (HDX-MS) and illustrated that the IgG1 binding was not solely limited to the Fc but that the Fab regions also interacts with the receptor (Yogo et al., 2019).

FcγRIIIb is located predominantly on the surface of neutrophils and a small population of basophils. FcγRIIIb is attached to a GPI-anchor and lacks a signalling component (Zhang et al., 2000). There are between four to six N-glycosylation sites in the extracellular domain. FcγRIIIb has the highest binding affinity for IgG3, followed by IgG1 (Patel et al., 2019). FcγRIIIb is also present as a soluble receptor, which can bind myeloid cells, NK cells, B lymphocytes, a subset of T lymphocytes and monocytes. This binding is mediated through complement receptors. The soluble form of FcγRIIIb was found to activate complement receptor type 3 (CR3)-dependent inflammatory process (Galon et al., 1996) and it can also bind complement receptor type 4 (CR4). FcγRIIIb has also been studied using solution scattering (Yogo et al., 2017).

1.3.4 FcRn
The neonatal Fc receptor (FcRn) is found on the surface of molecules that line the bloodstream. These receptors help maintain antibody levels in serum. It is also vital for the transfer of immunity from mother to foetus. The FcRn was identified after the other FcγRs were identified, despite early evidence of the transfer of IgG between mother and foetus (Simister et al., 1985). Apart from recycling IgG, FcRn is also involved in human serum albumin recycling (Oganesyan et al., 2014). The FcRn is composed of β2-microglobulin (β2m) and
three α-chains, which are anchored to the membrane. The β2m is related to the major histocompatibility complex (MHC) but cannot present peptides. There are several unbound structures of the FcRn available (PDB ID: 1EXU) (West et al., 2000), (PDB ID: 6C97) (Stöppler et al., 2018). It is also available in complex with Fab regions (PDB: 5WHK, 6NHA and 6FGB) (Kenniston et al., 2017; Smith et al., 2018; Blumberg et al., 2019), as well as, in complex with the Fc region (PDB ID: 1FRT and 1I1A) (West et al., 2000; Martin et al., 2001). Also available is a structure of FcRn in complex with albumin and Fc region simultaneously (PDB ID: 4N0U) (Oganesyan et al., 2014).

A seemingly conserved FcRn binding site in the Fc region exists at the interface between C\textsubscript{H}2 and C\textsubscript{H}3, with residues Met\textsuperscript{252}, Ile\textsuperscript{253}, Ser\textsuperscript{254}, Asn\textsuperscript{434}, His\textsuperscript{435} and Tyr\textsuperscript{436} being conserved in IgG1, IgG2 and IgG4, that form a 525 Å\textsuperscript{2} section of solvent-accessible surface area (DeLano et al., 2000). However, in IgG3 His\textsuperscript{435} and Tyr\textsuperscript{436} are substituted for Arg\textsuperscript{434} and Phe\textsuperscript{436} (Davies et al., 2015). This binding is pH-dependent, occurring at pH 6.0-6.5 and dissociates at a pH of approximately 7.5 (Rodewald, 1970, 1976). The FcRn-Fc complex reveals a structural change in the Fc when the receptor is bound and the formation of three titratable salt bridges that are thought to confer this pH-dependent binding (Martin et al., 2001). FcRn can contribute to the development of autoimmune disorders by allowing for high levels of self-IgG to continue to circulate in plasma. Therefore, there is some interest in developing an antagonist to FcRn that may help clear pathogenic IgG (Kenniston et al., 2017).

1.4 The Complement System

The complement system was discovered in the 1890s by Jules Bordet, when he recognised that bacteria can be haemolysed by antibodies operating with other heat-labile serum proteins that functioned to “complement” the clearance of pathogens. For this Bordet was awarded a Nobel prize in 1919 (Silverstein, 2001). The complement system is composed of more than thirty known glycoproteins, both soluble and membrane bound, that cooperate with the innate and adaptive response. The complement system is activated when complement proteins bind to pathogens, directly or via antibodies, and results in a cascade of enzymatic reactions. This cascade involves the cleavage and
activation of zymogens, that go on to incite a range of functions, including opsonin induced phagocytosis, incite inflammation through anaphylatoxins and ultimately assemble the membrane attack complex (MAC) that creates pores in pathogenic membranes, by inducing cell lysis. These functions are often mediated through complement receptors on phagocytic cells, granulocytes, or erythrocytes. To be noted, many complement proteins follow the uniform nomenclature, denoted by ‘C’, followed by a number, which indicates the order in which the proteins were discovered. This order is not strictly related to their position in the pathway. There are three complement pathways; classical, alternative and lectin, which all converge at the formation of C5 convertase (Figure 1.11).

The pathways are distinct in the complement components that initiate them; the classical pathway is initiated by C1q, the lectin pathway by mannose binding lectin (MBL) and the alternative pathway by ficolins (Sarma et al., 2011; Owen et al., 2013).

1.4.1 The Classical Pathway

The classical pathway is technically considered part of the adaptive immune response because it requires the formation of an antigen-antibody complexes. These complexes can be soluble, referred to as immune complexes, or epitopes, situated on pathogenic membranes. Only IgM and IgG antibodies are able to activate the classical complement pathway. The formation of the antigen-antibody complexes induces a structural change in the Fc region, which allows for C1 binding. C1 is composed of C1q and two molecules each of C1r and C1s, which are serine proteases. This C1 complex is held together by Ca2+, resulting in C1qr2s2. The antibody must be antigen bound in order to bind to C1q, because at least two proximate Fc binding sites are required for C1q to bind. In the case of IgM, when no antigen is bound, IgM adopts a planar conformation, where none of the Fc binding sites are exposed, however when antigen is bound IgM adopts a staple conformation, which allows for at these binding sites to be exposed. IgG only has one Fc binding site per molecule, even upon antigen binding, therefore, it requires up to 1000 membrane-bound IgG molecules to activate the classical pathway and induce lysis, whereas it requires fewer than 10 IgM molecules to activate complement.
Figure 1.11 Schematic illustrating the classical, alternative and lectin complement pathways (legend overleaf).
Figure 1.11 Schematic illustrating the classical, alternative and lectin complement pathways (continued).

The classical pathway is initiated by antibody-antigen complexes that bind C1q. This resulting in a conformational change in one of the C1r molecules, converting it into an active serine protease, which then cleaves its partner C1r molecule. The cleaved C1r pair go on to cleave and activate the two C1s molecules. The C1s hydrolyses C4 into C4a and C4b. C2 is cleaved by C1s, into C2a and C2b. C2b diffuses away, leaving a complex of C4b2a, a C3 convertase. This converts C3 into anaphylatoxin C3a and C3b. C3b binds C4b2a to form C4b2a3b, a C5 convertase. This complex cleaves C5 into C5b and C5a. C5b is the first complement protein that goes on to form the MAC, with the addition of C6, C7, C8 and several C9 molecules.

MBL binds several carbohydrate moieties on pathogenic surfaces. MBL is associated with MASP1, 2 and 3. MASP2 acts as C1s, cleaves both C2 and C4, generating C4b2a a C3 convertase. Following this point this pathway follows the same trajectory as the classical pathway.

High serum C3 concentrations allows for spontaneous formation of C3(H₂O). In the presence of Mg²⁺ C3(H₂O) binds factor B. Factor B is cleaved factor D, releasing Ba that diffuses away and Bb remains bound to C3(H₂O) forming C3(H₂O)Bb, a C3 convertase active in plasma, that cleave C3 into C3a and C3b, which can go on to interact with factor B to form C3bB and this interacts with factor D to form C3bBb. This is the second C3 convertase of the alternative pathway, distinct from C3(H₂O)Bb, as C3bBb is membrane bound rather than being in solution. This complex is stabilised by properdin, a serum protein and generates C3b at the microbial surface. These C3b molecules bind to factor B, enabling more Bb to be generated amplifying the pathway. C5 convertase is composed of C3bBbC3b and is stabilised by factor P. As in the other pathways C5 convertase cleaves C5 and leads to the formation of the MAC.

Outcomes of these complement pathways is inflammation, opsonisation and cell lysis.

Adapted from (Owen et al., 2013)
When C1q binds to C_H2, it elicits a conformational change in one of the C1r molecules, converting it into an active serine protease, which then cleaves its partner C1r molecule. The cleaved C1r pair go on to cleave and activate the two C1s molecules. The C1s hydrolyses C4 into C4a and C4b. This C4 cleavage creates an unstable internal thioester in C4b. This allows for the covalent attachment of C4b to the target membrane in proximity to C1, then C4b binds to C2. This makes C2 susceptible to cleavage by C1s, with C2b diffusing away to leave a complex of C4b2a, referred to as C3 convertase. Membrane-bound C3 convertase, as the name implies, converts C3 into an active form C3b as well as creating a small anaphylatoxin C3a. One C3 convertase molecule can generate over 200 molecules of C3b, resulting in an amplification of the classical pathway. C3b binds to pathogenic membranes, acting as an opsonin, signalling to phagocytic cells with C3b receptors to engulf the pathogen. C3b can also bind to the Fc region of antibodies in immune complexes, this being the signal for phagocytic cells to phagocytose the immune complex, or for erythrocytes to be transported to the liver to be destroyed. Additionally, C3b can bind C4b2a to form C4b2a3b, referred to as the C5 convertase. This complex cleaves C5 into C5b and C5a. C5b is the first complement protein that goes on to form the membrane attack complex (MAC), with the addition of C6, C7, C8 and several C9 molecules (Owen et al., 2013).

1.4.2 C1q

Activation of the classical complement pathway is mediated by C1q, which acts as the recognition unit. It is a glycoprotein assembled from 18 polypeptide chains with a C-terminal globular head region that mediates recognition of molecular structures and a N-terminal collagen-like tail, that modulates immune effector functions. It was discovered in late 19th century and purified from serum in the 1940s (Sim et al., 2016), but it was not until the 1960s that C1q and its proteases C1r and C1s were officially defined (Lepow et al., 1963). C1q appears early in evolution, a recent study showed that gC1q protein from amphioxus, an invertebrate, can bind human IgG and activate the classical complement pathway (Gao et al., 2014).
The first C1q structure was solved by electron microscopy, and it was described as having a “fragile” and “delicate” structure (Shelton et al., 1972; Svehag et al., 1972), that some have, rather romantically, compared to a bouquet of tulips. It has six peripheral globular regions joined together by fibrillar strands to a central bundle of fibres. Crystal structures of the globular heads have been solved and shed light on the mechanisms behind the ability of C1q to recognise so many varied ligands (Gaboriaud et al., 2003).

IgG and IgM can activate complement through binding with C1q. However, there is some variation between IgG subclasses in regard to the affinity to C1q. IgG1 and IgG3 have the highest affinity for C1q (Bindon et al., 1988), but IgG2 and IgG4 have a reduced affinity and IgG2 only binds C1q under certain conditions. C1q binds to IgG at the C\textsubscript{H}2 domain, the residues that are vital for this interaction are thought to be Leu\textsuperscript{235}, Asp\textsuperscript{270}, Lys\textsuperscript{322}, Pro\textsuperscript{329}, and Pro\textsuperscript{331}; variations in amino acid sequence at the C\textsubscript{H}2 domain in IgG2 and IgG4 may explain the reduced binding affinity for C1q. Studies have found that IgG hexamers form via interactions at the Fc region, assemble at the cell surface and can bind C1q (Diebolder et al., 2014; van den Bremer et al., 2015).

1.4.3 Lectin Pathway

The lectin pathway is distinct from the classical pathway because the initiator molecule is not activated by an antigen-antibody complex but rather a lectin molecule. Thus, this pathway is considered to form part of the innate response. Lectins are carbohydrate components found on microbial membranes. MBL binds several carbohydrate moieties including mannose, N-acetyl glucosamine, D-glucose, L-fucose, present on microbial surfaces. MBL can also bind ficolins. MBL is similar to C1q in structure. In serum MBL is associated with MBL-Associated Serine Proteases (MASPs). There are three MASP proteins, MASP1, MASP2 and MASP3. However, MASP2 plays a central role in the lectin pathway. MASP2 acts in the same way as C1s, and can cleave both C2 and C4, generating the C4b2a complex that cleaves C3 to C3b. The pathway hereafter follows the same route as the classical pathway (Owen et al., 2013).
1.4.4 Alternative Pathway

The alternative pathway is also initiated by an initiator independent of antigen-antibody interactions and is therefore also considered to be part of the innate response. Additionally, the alternative pathway is unique as it uses C3 and C5 convertases whose components are different to those in the classical and lectin pathway. The C3 convertase is composed of C3b and Bb, a molecule exclusive to the alternative pathway. The C5 convertase is composed of an additional C3b molecule to the C3 convertase. The alternative pathway can be activated in a number of ways. The first is referred to as the ‘tickover’ pathway, that uses C3, factor B, factor D and properdin. This pathway is so dubbed ‘tickover’ due to the fact that C3b is constantly being generated and then being spontaneously inactivated. Another mode of activation is via properdin, and a third mode is through thrombin and proteases.

The tickover pathway is initiated when C3 has a high serum concentration, and undergoes spontaneous thioester bond hydrolysis, forming C3(H₂O). In the presence of Mg²⁺ C3(H₂O) binds factor B. Factor B is cleaved by serine protease factor D, releasing Ba that diffuses away and Bb which remains bound to C3(H₂O) forming C3(H₂O)Bb, a C3 convertase active in plasma, that cleaves C3 into C3a and C3b. C3b can go on to interact with factor B to form C3bB and this interacts with factor D to form C3bBb. This is the second C3 convertase of the alternative pathway, distinct from C3(H₂O)Bb, as C3bBb is membrane bound rather than being in solution. This complex is stabilised by properdin, a serum protein and generates more C3b at the microbial surface. These C3b molecules bind to factor B, enabling more Bb to be generated and thus amplifying the pathway. C5 convertase is composed of C3bBbC3b and is stabilised by factor P. As in the other pathways C5 convertase cleaves C5 and leads to the formation of the MAC.

Another initiator of the alternative pathway is postulated to be properdin which acts as a stabilising factor in the tickover pathway. *In vitro* studies have shown that surface immobilised properdin can bind C3b and factor B in the presence of Mg²⁺. The bound factor B was then cleaved by factor D and resulted in C3bPBb, that acted as an effective C3 convertase. Possible physiological evidence for this comes from clinical evidence that properdin deficient patients are more susceptible *Neisseria gonorrhoeae* induced meningococcal disease.
However, properdin’s capacity to initiate the alternative pathway still needs to be fully determined. Finally, the alternative pathway is thought to be activated through interactions between the complement cascade and the coagulation cascade, through thrombin, which can cleave C3 and C5 (Owen et al., 2013).
Chapter 2 Glycosylation
2.1 Glycobiology

A core paradigm in biology is that DNA is transcribed into RNA and is then translated into protein. This elegant template-based view of proteins has dominated scientific thinking. However, this is only part of the picture. In reality many proteins are associated with other molecules, generally carbohydrates and lipids, referred to as co/post-translational modifications, as they are attached during or after the protein is translated. These modifications are often integral to the function of the protein. These molecules are diverse in nature and makeup, contributing significantly to the heterogeneity of the overall molecule. Sugars can be attached to proteins covalently as monosaccharides, or as chains of several sugars, (oligosaccharides), generally referred to as glycans.

2.1.2 Chemistry of Carbohydrates

A monosaccharide is a basic component, that cannot be hydrolysed into smaller subunits. There are two main types, aldoses and ketoses. Aldoses have a carbonyl group at the end of the carbon chain, i.e. an aldehyde group, and ketoses have a carbonyl group on the inner carbon, i.e. a ketone group. Free monosaccharides can exist in either a chair or open linear conformation, however, oligosaccharides always take on a chair conformation.

In an oligosaccharide, monosaccharide subunits are attached to one another by glycosidic bonds. In the ring form, aldose sugars have a chiral centre at carbon 1 (C1) and ketose sugars have a chiral centre at C2. A glycosidic bond is formed between hydroxyl group of one monosaccharide and the anomeric C of the other (C1 in ketose sugars and C2 in aldose sugars). Glycosidic bonds give rise to stereoisomers at the anomeric C, which form either α- or β-linkages. The two linkages result in very different polymer structure and functions, the classical example being starch and cellulose as polymers composed of repeating glucose moieties, however starch is α1-4 linked and cellulose is β1-4 linked.

Peptide bonds are used to link the amino acids, namely, asparagine, serine, threonine and tyrosine, while glycosidic bonds link these sidechains with the glycan. Monosaccharides can simultaneously form two glycosidic bonds; this gives rise to branching of the oligosaccharide, which is an important mechanism for adding structural diversity to the glycan. The chair structure is rigid,
Figure 2.1 Chair conformations of commonly occurring mammalian sugars.
while the torsion angles around glycosidic bonds can vary, making glycans flexible, with the ability to occupy multiple conformations in solution. The simultaneous structural flexibility in the linkers and rigidity in the rings contribute to the biological functioning of the glycan.

Monosaccharides can be labelled D- or L-, this is determined by the stereocenter furthest away from the carbonyl group. Most vertebrate monosaccharides have a D- conformation, with the exception of fucose and iduronic acid, which take on an L- conformation. The names of the monosaccharides are generally abbreviated; glucose (Glc), N-Acetyl-glucosamine (GlcNAc), D-Galactose (Gal), N-Acetyl-Galactosamine (GalNAc), D-Mannose (Man), N-Acetyl-Neuraminic acid (NeuNAc), and L-fucose (Fuc) (Figure 2.1) (Varki et al., 2009).

2.1.3 N-linked glycosylation

N-linked glycans, are so named, in reference to their binding to the nitrogen, N, from an asparagine residue, with the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid residue except for proline. However, glycosylation of the Asn residues with this sequence is not guaranteed, approximately 70% of these sites are glycosylated. In eukaryotes, the asparagine is always linked to a GlcNAc moiety, however in prokaryotes greater diversity in the Asn-linked moiety is allowed.

The core sequence of all eukaryote N-linked glycans contains, Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1–4GlcNAcβ1-Asn, where Manα1-3 and Manα1-6 are each a branch of the main core. There are three main glycan structures in eukaryotes; oligomannose, where the core is extended by Man moieties, complex, where GlcNAc moieties extend the core referred to antennae, and hybrid, where Man moieties extends the Manα1-6 arm of the core, and GlcNAc moieties extend from the Manα1-3 arm (Figure 2.2) (Varki et al., 2009).
2.1.3.1 Biosynthesis of the N-linked Glycan

Figure 2.2 The three predominant N-linked oligosaccharide structures occurring in eukaryotes.

The three oligosaccharide types are:

A. Oligomannose
B. Complex
C. Hybrid
The pathway for N-linked glycan synthesis can be divided into three stages, the first is the formation of the lipid-linked precursor oligosaccharide, followed by the *en bloc* transfer of the oligosaccharide onto the polypeptide, and finally the processing of the oligosaccharide (Figure 2.4). The first two processes occur at the endoplasmic reticulum (ER), and the final process occurs in the Golgi apparatus.

In the first step the lipid-linked precursor oligosaccharide is a Glc$_3$Man$_9$GlcNAc$_2$ molecule attached via a pyrophosphate linkage to dolichol (Dol), a lipid, composed of five-carbon isoprene units. The dolichol is a long molecule, longer than the fatty acid tails of the lipid bilayer. The dolichol inserts into the membrane of the ER with the phosphate head group facing the cytoplasmic side of the ER (Taylor *et al.*, 2011). The glycan is transferred onto the phosphate head group of the dolichol group in two steps. For glycan synthesis the sugar must be activated to a high-energy donor, achieved by a nucleoside triphosphate, either uridine triphosphate (UTP) or guanosine triphosphate (GTP). A kinase can then catalyse the sugar activation resulting in uridine diphosphate (UDP), UDP-GlcNAc, or guanosine diphosphate (GDP), giving GDP-Man. Activated sugars can also be synthesized by previously activated sugars (Varki *et al.*, 2009) (Figure 2.3).

**Figure 2.3 Schematics illustrating sugar activation.**

A. Sugar activation by kinases. B. Sugar activation by previously synthesized activated nucleotide sugar. Adapted from (Varki *et al.*, 2009).

To start the process, GlcNAc-1-P from UDP-GlcNAc is attached to the Dol molecule by a GlcNAc-1-phosphotransferase to form Dol-P-P-GlcNAc. Subsequent glycan additions are catalysed by glycosyltransferases that transfer
the UDP/GDP linked glycan to the growing oligosaccharide chain by forming a glycosidic bond. Glycosyltransferases have a high specificity for a nucleotide sugar donor and an acceptor. The glycosyltransferase catalyses the attachment of two GlcNAc and five Man moieties (Varki et al., 2009). The resulting Man5GlcNAc2-P-P-Dol is translocated across the membrane into the luminal side of the ER, in a process that is still poorly understood due to the difficulty in identifying the scramblase responsible for the dolichol-linked glycan translocation (Verchère et al., 2021). When the dolichol glycan is on the luminal side four more Man moieties are added followed three Glu moieties, resulting in the mature N-glycan precursor Glc3Man9GlcNAc2-P-P-Dol.

Finally, the dolichol-linked glycan is transferred *en bulk* to asparagine residues of polypeptides catalysed by an oligosaccharyltransferase in the lumen of the ER, by cleaving the GlcNAc-P bond, which releases the Dol-P-P. In order for this transfer to occur the polypeptide must meet several criteria. This includes the requirement that the Asn must be found within an invariable consequence sequence, Asn-X-Ser or Asn-X-Thr, where X can be an amino acid except proline. The protein needs to have an appropriate 3D structure as glycans are polar and are generally bound at the surface of proteins. Glycans are transferred onto the polypeptide nascent chain, meaning that the protein has not yet undergone folding. Therefore, glycans can influence the folding. However, in cases of rapid folding the N-linked consensus sequence may not be available to the OST.

Once the glycan is attached to the polypeptide it undergoes processing, through the removal and addition of glycan moieties by glycosidases and glycosyltransferases, respectively. The first processing step results in the sequential removal of the three Glc residues. This first step plays a key role in the regulation of protein folding, as the Glc moiety acts as marker for the regulatory proteins calreticulin and calnexin, and any misfolded proteins are targeted for ER degradation (Tannous et al., 2015). Once the Glc moieties are all removed, the glycoprotein ER α-mannosidase I will remove the terminal Manα1-2 to give ManβGlcNAc2, and this will proceed onto the Golgi apparatus.

More Man moieties are trimmed by mannosidases in the *cis*-Golgi to give Man6GlcNAc2. Further processing of the glycan to yield hybrid or complex oligosaccharides occurs in the medial-Golgi. An N-acetylglucosaminytransferase
Figure 2.4 Diagram illustrating the biosynthesis of the N-glycan.

Diagram depicting the biosynthesis of the full N-linked glycan, illustrating the different compartments in which the moieties are added. The process is described in Section 2.1.3.1. Adapted from (Varki et al., 2009; Taylor et al., 2011).
adds a GlcNAc moiety to the C2 of the Manα1-3 moiety of the Man5GlcNAc2 core. This is followed by the trimming of the terminal Manα1-3 and Manα1-6 moieties forming GlcNAcManα3GlcNAc2. Once both Man moieties are removed, a second GlcNAc is added to the C2 of Manα1-6. This gives the precursor to all biantennary complex N-glycans.

It is possible for the biantennary N-glycan to undergo further branching by the addition of another GlcNAc moiety at the C-4 of the core Manα1-3 and C-6 of the core at Manα1-6, generating tri-antennary and tetra-antennary N-glycans. A bisecting GlcNAc moiety attached to the β-Man of the core may be found in complex and hybrid N-glycans.

Hybrid and complex N-glycans can undergo even further processing by the addition of more sugar moieties to the core. For example, an Fucα1-6 can be added to the Asn-linked GlcNAc in the N-glycan core. The N-glycans can also have extended branches through the addition of Gal to the GlcNAc moieties at the branches. The glycan structures can be capped by the addition of a number of sugars, including NeuAc, Fuc, Gal, and GlcNAc. These sugars are α-linked, meaning that they protrude away from the β-linked Gal-GlcNAc. Processing of the N-linked glycan results in diverse glycan structures and in a heterogeneous mix of glycan structures in each sample. The mechanisms behind glycan heterogeneity are still not fully understood but could be a result of protein conformations that affect substrate availability for the glycosidases and glycotransferases. Heterogeneity could be due to the nucleotide sugar metabolism, impacting the availability of sugars for use in processing, or could be due to the rate at which the glycan moves from the ER through the Golgi. Finally, heterogeneity could be due to the distribution of glucosidases and glucotransferases within the Golgi apparatus network (Varki et al., 2009).

### 2.1.4 O-linked glycosylation

O-linked glycans are GalNAc attached to the hydroxyl group of Ser or Thr residues. Proteins such as the mucins have a large number of O-GalNAc glycans, however some proteins contain only a few scattered O-glycans. O-glycans can be composed of GalNAc, Gal, GlcNAc, Fuc and sialic acid moieties. The length
of the O-linked glycan is variable as it can have just one glycan moieties to over twenty moieties.

2.1.4.1 Biosynthesis of the O-linked Glycan

There are several differences between the biosynthesis of O-glycans and N-glycan, firstly all sugars are added one at a time, as opposed to *en bulk* transfer. Secondly, there is no defined consensus sequence for O-glycan addition. This lack of consensus sequence is probably due to the existence of several glycosyltransferases that can transfer the first GalNAc onto the Ser/Thr residue, but there is only one oligosaccharyltransferase in N-linked glycosylation. However, extensive analysis into O-glycan placement has uncovered several trends, including the observation that glycosylation sites tend to be in a region rich with proline and alanine (Taylor *et al.*, 2011). The O-GalNAc glycans are attached to the protein in the Golgi apparatus by type II transmembrane glycosyltransferases, with a catalytic side facing the lumen of the Golgi. The first step in this biosynthesis pathway is the transfer of UDP-GalNAc to Ser/Thr with an α-linkage by polypeptide N-acetylglucosaminyltransferase. There are four common O-glycans core structures (*Figure 2.5*). The first core structure includes a Galβ1-3 linked to the GalNAc, catalysed by the enzyme Core 1 β1-3 galactosyltransferase. Core 2 type O-glycans, involves an addition of another GlcNAc moiety onto the GalNAc of Core 1 type structure, with a β1-6 linkage. The Core 2 β1-3 N-acetylglucosaminyltransferase uses the Core 1 structure as a substrate. Core 2 can become monoantennary or biantennary elongated, with the addition of multiple Gal moieties. The Core 3 O-glycan production is controlled by Core 3 β1-3 N-acetylglucosaminyltransferase and uses the GalNAcα-Ser/Thr as a substrate, to add a GlcNAc β1-3 linkage. Like Core 1 β1-3 galactosyltransferase, Core 3 β1-3 N-acetylglucosaminyltransferase uses the same GalNAcα-Ser/Thr as a substrate, therefore some competition exists between the Core 1 and Core 3 enzymes. The formation of Core 3 inhibits the formation of Core 2. Core 4 is formed by the addition of a GlcNAc with a β1-6 linkage (Varki *et al.*, 2009).
Figure 2.5 Diagram of O-glycan core structures.
A. The O-glycan precursor. B. Core 1 with a Gal moiety attached to the precursor.
C. Core 2 with Gal and GlcNAc moiety attached to the precursor. D. Core 3 with a GlcNAc moiety attached to the precursor. E. Core 4 with two GlcNAc moieties attached to the precursor.
2.1.5 Immunoglobulin G N-glycosylation

All antibodies of the IgG subclass have a conserved N-glycan at Asn\(^{297}\) (EU numbering), at the \(\text{C}_{\text{H}2}\) domain of the IgG-Fc. There are several high-resolution structures of the human IgG-Fc region. In all of these the N-glycan is found in the central cavity flanked by two \(\text{C}_{\text{H}2}\) domains (Figure 2.6). The PDB IDs for IgG1-Fc are 4W4N, 4KU1, 4BM7, 3AVE, 4Q74, 4BYH, 1H3X (Krapp et al., 2003; Matsumiya et al., 2007; Crispin et al., 2013; Yu et al., 2013; Ahmed et al., 2014; Frank et al., 2014; Kiyoshi et al., 2015). The PDB IDs for IgG2-Fc are 4HAF, 4HAG, 4L4J (Teplyakov et al., 2013; Vafa et al., 2014). The PDB ID for IgG3-Fc is 5W38 (Shah et al., 2017). The PDB IDs for IgG4-Fc are 4C54, 4C55, 5LG1, 5W5M and 5W5N (Davies et al., 2014; Davies et al., 2017; Tam et al., 2017). There are fewer full-length IgG structures available however most of these confirm that the position of the Fc-glycan lies within the central cavity, PDB IDs: 1HZH, 6GFE (Saphire et al., 2001; Blech et al., 2019). The exception is a higher resolution structure of an IgG4 antibody, Pembrolizumab, with a Ser\(^{228}\)Pro mutation, that showed that the one N-glycan does not sit within this cavity, rather it is rotated by 120° to be completely solvent exposed (PDB ID: 5DK3). This shift was suggested to be due to the shorter hinge of this antibody (Scapin et al., 2015).

There can be large diversity in the glycoforms present given that, between 0-16 possible glycan chains are possible, thus when considering that there are two glycosylation sites per IgG, that amounts to over 128 possible glycoform combinations. The glycoforms present in the sample are dependent on the cell expressing the protein, and the culture method and conditions (Jefferis, 2009).

2.1.5.1 Role of the Immunoglobulin N-glycan

The N-glycan has been implicated in the ability of the IgG antibodies to bind to the FcRs and C1q. Several studies of deglycosylated IgG have shown abrogated binding to the FcγRs and C1q (Nose et al., 1983; Walker et al., 1989; Mimura et al., 2001; Krapp et al., 2003; Kang et al., 2020). High resolution structures of the IgG1-Fc co-crystallised with the FcγRs are available (Figure 1.10). High-resolution structures of the Fc region in complex with the FcγRI and FcγRIII receptors showed binding in a similar region in the upper part of the Fc
IgG-Fc structures with C_{H2} N-glycan

Figure 2.6 IgG-Fc structures with C_{H2} N-glycan

IgG1-Fc from PDB ID: 4W4N (Kiyoshi et al., 2015), IgG2-Fc from PDB ID: 4L4J (Vafa et al., 2014), IgG3-Fc from PDB ID: 5W38 (Shah et al., 2017) and IgG4-Fc from PDB ID: 4C54 (Davies et al., 2014). The N-glycans are shown in green sticks.
region with many conserved contacts, despite their varying affinities to the Fc
region (Sondermann et al., 2000; Radaev, Motyka, et al., 2001; Kiyoshi et al.,
2015). In these structures, very few contacts are seen between the Fc-glycan and
the FcγRs. There are no high-resolution structures of full-length IgG in complex
with the FcγRs. This is limiting as it is not clear from the available structures IgG-
Fc structures what role the glycan has in stabilising the hinge. The deglycosylated
IgG1-Fc structure has a more compact conformation in comparison to the
glycosylated IgG1-Fc. This indicates that Fc-glycan may play a role in stabilising
the Fc architecture (Krapp et al., 2003). SAXS studies of glycosylated and
deglycosylated IgG1-Fc suggested that the deglycosylated Fc had a larger R<sub>G</sub>
compared to the glycosylated Fc (Borrok et al., 2012). Additionally, removal of
the glycan may have more local implications such as making the C'E loop (Gln<sup>293</sup>-
Phe<sup>303</sup>) in the C<sub>H2</sub> domain, the region where the FcγRs bind more disordered
(Borrok et al., 2012).

Studies of deglycosylated IgG4-Fc (PDB ID: 4D2N) demonstrated that the
C<sub>H2</sub>-C<sub>H2</sub> domain interactions partially bury the C<sub>H2</sub> surface, which in the
presence of the glycan would be solvent exposed. Moreover, the conformation of
the DE loop to which the IgG4-Fc glycan is altered, possibly further providing an
explanation the lack of FcγRs binding (Davies et al., 2015). There is a paucity of
literature investigating the effect of the Fc-glycan on the structure/function of
antibodies from subclasses other than IgG1. There are a limited number of
studies investigating IgG4-Fc, however there are distinct lack of structural studies
for the deglycosylated structures of IgG2 and IgG3. Therefore, it is difficult to
understand the role of the Fc-glycan in all of the IgG subclasses.

Some studies have shown that deglycosylation of the IgG1-Fc has an
impact on the thermostability of the IgG1 and IgG2 molecule, when compared to
the glycosylated IgG1-Fc (Mimura et al., 2000; Latypov et al., 2012). The
limitation of these studies is that they focused on the Fc, rather than the full-length
IgG molecule. A study investigating the thermostability properties of different
glycoforms demonstrated that the successive enzymatic removal of N-glycan
moieties down to the GlcNAc moiety results in a marked decrease in the melting
temperature of the IgG1 antibody (Wada et al., 2019). This correlated with recent
thermodynamic studies of glycosylated IgG1 which demonstrate that aggregation
is driven by the Fc region of the IgG molecule (Zhang et al., 2020). The role of
that the Fc-glycan plays on the thermodynamics of IgG1 and IgG4 is explored in Chapter 6 of this thesis.

To complicate matters, different glycan structure can confer different effector functions. For example removal of the core fucosylation leads to higher binding affinity FcγRIIIa (Niwa et al., 2005; Matsumiya et al., 2007; Pereira et al., 2018). Terminal sialic acids impairs complement-dependent cytotoxicity (Quast et al., 2015). IgG1 binding of C1q was found to be increased in its affinity when the terminal Gal moieties and one NeuAc are present in the N-glycan, when compared to when just the core GlcNAc moieties are present in the N-glycan (Wada et al., 2019). The same study found that the presence or absence of the core Fuc made no difference on C1q binding. A similar study found that deglycosylated IgG1 had a very significantly lower melting temperature, -5.2 ± 1.0°C than its glycosylated counterpart. The combination of the lack of structural data and heterogenous glycan structures that can illicit different effector functions makes unravelling the role of the Fc-glycan a very complicated task.
Thesis aims and objectives

This first aim of this thesis is to address the knowledge gap in understanding the role of the Fc-glycosylation, for antibodies of the IgG subclass, namely IgG1 (Chapter 4) and IgG4 (Chapter 5) which are widely exploited as therapeutics and IgG3 (Chapter 7) which suffers from a paucity of a research into its structure and function. To achieve this, a combined structural approach is employed including analytical ultracentrifugation (AUC), small angle X-ray and neutron scattering (SAXS and SANS) and atomistic modelling (Chapter 3).

Combining several in solution techniques offers numerous advantages over other more static structural techniques such as X-ray crystallography and electron microscopy as this approach allows for the interpretation of dynamics when SAS data is paired with atomistic modelling of the possible different conformers that can exist within the solution. This is especially pertinent considering that IgG molecules are large multi-domain proteins that have inherent flexibility conferred by the hinge region. This approach starts to allow one to probe the effect of structural elements such as glycosylation, typically a difficult post translational modification to study, on the whole protein rather than simply examining the local role of the glycan.

Another aim is to understand the influence of the Fc glycan on the thermal stability of IgG1 and IgG4 was also investigated (Chapter 6), using intrinsic fluorescence and static light scattering (SLS), measured on the UNit (Chapter 3) to find effect of deglycosylation on melting and aggregation temperatures. By using SAS to study the effects of thermal melting it is possible to scrutinize parts of the structure that start to unfold first and thus might have less stability. In this thesis, the aim is to merge the insights into the structure of IgG1 and IgG4 from Chapters 4 and 5 and build up a more significant understanding of the stabilising role of the Fc-glycan in Chapter 6.

Studying IgG3 is another aim of this thesis, an antibody with an elongated hinge region. Investigating IgG3 allows for benchmarking against the other IgGs in the class to understand the role of the hinge region in more detail, and this can be related this to their relative function and ability to activate different receptors.
Chapter 3 Methods for studying the solution structure, stability and function of immunoglobulins with and without glycans
3.0 Introduction

To evaluate the role of the Fc glycan in the structure of IgG1 (Chapter 4), IgG3 (Chapter 7) and IgG4 (Chapter 5) a combined approach was implemented in this thesis, consisting of analytical ultracentrifugation (AUC), small angle X-ray and neutron scattering (SAXS/SANS), and atomistic modelling. These methods will be described in Sections 3.1, 3.2 and 3.5, respectively. To assess the effect of the Fc glycan on the thermostability of IgG1 and IgG4 (Chapter 6), two industrially relevant antibodies, thermal unfolding of the antibodies was followed by fluorescence and static light scattering (SLS) described in Section 3.4, and by SAXS and SANS (Section 3.2).

3.1 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is an in solution, molecular mass and shape determination technique that was initially developed by Theodor Svedberg (Svedberg et al., 1923, 1926). It can be used to study macromolecules in their near native state, in a range of concentrations and buffer formulations. AUC is also non-destructive so samples can be used for other experiments. The macromolecules in a solution will sediment when subjected to a high centrifugal force, and by AUC this sedimentation can be followed in real time to determine molecular weights, conformation, sample homogeneity, and molecular interactions (Ralston, 1993; Cole et al., 2008).

The analytical ultracentrifuge instrumentation allows for two main types of experiments; sedimentation velocity experiments, where high speeds and long sample cells are used, giving information on the size and shape of the molecule or sedimentation equilibrium experiments, using low speeds with shorter sample cells, providing information on the solution including, molecular masses, stoichiometries, association constants and solution nonideality. More recent advancements in instrumentation allow the user to follow several wavelengths simultaneously, which can be used to probe molecular interactions more extensively. In this thesis, primarily sedimentation velocity experiments are used to probe the structure of glycosylated and deglycosylated antibodies in Chapters 4, 5 and 7. AUC is complementary to other in solution techniques employed in this thesis, such as small angle scattering. Additionally, AUC can provide an extra
experimental restraint for computational modelling, as theoretical sedimentation parameters can be generated for modelled proteins, discussed in Section 3.5.7.

3.1.1 Principles of analytical ultracentrifugation

The rate at which a protein's sediment is dependent on protein-specific factors, including its density, shape and mass. Proteins that have a more compact structure, approaching a globular shape, and have a larger mass, sediment more quickly, compared to proteins that are more elongated, converging to a rod-like structure and have a smaller mass. In an AUC experiment sedimentation can also be influenced by experimental parameters such as rotor speed and buffer density and viscosity. This is summarised in the Svedberg equation.

As the protein is sedimenting it is subjected to several forces (Figure 3.1), including the downward centrifugal force \( F_{sed} \), frictional force \( F_f \) and the buoyancy force \( F_b \). The latter is the up thrust equal to the weight of the displaced solvent. A concentration gradient forms as the protein is sedimenting, generating diffusion effects, which also oppose sedimentation (Ralston, 1993).

The centrifugal force \( F_{sed} \) is proportional to the square of the rotor speed:

\[
F_{sed} = m \omega^2 r
\]

where \( m \) is mass of the solute (in grams, g), \( \omega^2 r \) is the centrifugal acceleration, \( \omega \) is the rotor angular velocity (in radians per second) and \( r \) is the distance of the solute from the centre of rotation.

The buoyancy force \( F_b \) is based on Archimedes’ principle, in which the buoyancy force is proportional to the weight of the solvent displaced:

\[
F_b = -m_s \omega^2 r
\]

Where \( m_s \) is the mass of the solvent displaced by the solute given by:

\[
m_s = m \bar{v} \rho
\]
The main forces acting on a sedimenting molecule include the centrifugal force $F_{\text{sed}}$, frictional force $F_f$, buoyancy force $F_b$, and velocity $v$, redrawn from (Ralston, 1993).
where the solute partial specific volume is \( \bar{\rho} \) (volume in mL for each gram of the solute occupies in solution), which is determined in this thesis from the residue and glycan composition of the macromolecule using the program SLUV (Perkins, 1986). \( \rho \) is the density of solvent (in g/ml). In this thesis this parameter was experimentally measured using a density meter.

The sedimenting solute will be subject to a frictional force \( F_f \) expressed by:

\[
F_f = -f \nu
\]

Equation 3.4

where \( f \) is the translational frictional coefficient. This is dependent on the size and shape of the solute, where \( \nu \) is the velocity of the migrating solute. \( F_f \) is the difference between \( F_{sed} \) and \( F_b \), when equilibrium has been established.

Following sedimentation over time allows for the calculation of the sedimentation coefficient, also referred to as the \( s \) value, given in Svedberg units (S), where 1 S = 10^{-13} sec. The Svedberg equation incorporates experimental and molecular parameters for the determination of the sedimentation coefficient:

\[
s = \frac{\nu}{\omega^2 r}
\]

Equation 3.5

where \( \nu \) is the velocity of the particle, \( \omega^2 r \) is the gravitational acceleration.

When the three central forces, \( F_{sed} \), \( F_b \) and \( F_f \) are balanced the sedimentation coefficient \( s \) can be found by the Svedberg equation:

\[
s = \frac{m(1-\bar{\rho})}{f}
\]

Equation 3.6

The sedimentation coefficient is proportional to the molecular mass \( m \) and inversely proportional to the frictional coefficient \( f \). The frictional coefficient is expressed by Stokes’ law and the Stokes radius \( R_s \):

\[
f = 6\pi \eta R_s
\]

Equation 3.7
where \( \eta \) is the solvent viscosity. In this thesis viscosity is measured experimentally using a viscosity meter (Anton Paar AMVn Automated microviscometer). \( R_s \) is the radius of a representative sphere which has the same frictional coefficient as the macromolecule. Stokes’ law can also be used to determine the frictional coefficient of a smooth, compact spherical protein. The sphere has the same mass and density as the macromolecule with a radius \( R_0 \) and corresponds to translational friction coefficient \( f_0 \). The frictional ratio \( \frac{f}{f_0} \) is equal to \( \frac{R_s}{R_0} \), which describes the friction that the macromolecule has relative to the sphere resulting the mass rearrangement to its most compact form. The frictional ratio is a measure of shape asymmetry (Lebowitz et al., 2009). Taken together the sedimentation coefficient can be expressed as:

\[
S = \frac{m(1-\bar{\rho})}{6\pi\eta R_0(f/f_0)}
\]

Equation 3.8

Sedimentation coefficients derived from different experiments, under different conditions, need to be adjusted to reflect standard conditions of water at 20°C:

\[
S_{20,w} = S \frac{\eta(1-\bar{\rho}\rho_0)}{\eta_0(1-\bar{\rho})}
\]

Equation 3.9

where the density \( \rho_0 \) is 0.9982 g/ml and the viscosity \( \eta_0 \) is 0.01002 poise (IAPWS, 2008).

3.1.2 Instrumentation

An analytical ultracentrifuge is composed of a motor and rotor, which enable the “spinning” of samples with speeds reaching up to 60,000 rpm (262,000 x g) (Ralston, 1993). The rotor includes holes for the AUC “cells”, which are precisely aligned prior to starting an AUC experiment. An AUC cell has two sector-shaped chambers, one for the sample solution and the other for the matching reference buffer. The sector shape is critical as it moves the sedimenting molecules in a radial direction, while preventing convection disturbances caused by sedimenting molecules at the periphery as they collide.
with chamber walls, which would be the case if the chamber was rectangular. The two cell chambers are enveloped in between transparent windows made of sapphire or quartz that allow light to be detected through the sample and buffer. The difference in the absorbance between the reference buffer versus the sample is measured by the optical system (Figure 3.2), therefore exact buffer matching, achieved by a long dialysis, prior to an AUC experiment, is required to ensure high quality data is acquired.

An analytical ultracentrifuge comes equipped with detection systems in order to follow the sedimentation of the sample as it is being centrifuged, by taking radial concentration distributions, or simply, scans, as the cell passes through the optical detector’s path, at regular intervals ranging from seconds in velocity sedimentation experiments to hours in equilibrium sedimentation experiments. An analytical ultracentrifuge comes equipped with three main detection systems, absorbance optics, Rayleigh interference optics or fluorescence detection system. Absorbance and interference optics can be used simultaneously, providing coupled datasets for the same sample, making data analysis more rigorous. In this thesis data are presented from absorbance and interference optics. All AUC experiments in this thesis were performed on either a Beckman Coulter ProteomeLab XL-A/XL-I (Furst, 1997) or the new Beckman Coulter Optima (Chadee-Luciano et al., no date). The exterior of the analytical ultracentrifuge is a temperature-controlled and evacuated armoured-chamber.

### 3.1.3 Absorbance detection system

The absorbance optical system (Figure 3.2, A) is highly sensitive allowing for very dilute samples, with concentrations in the μg/ml range, to be studied with ease. However, there is an upper absorbance limit of 1.5 optical density units (OD) (Cole et al., 2008), which may limit the use of this detection system when working with more concentrated samples. This is due to the loss of the linear relationship of the Beer-Lambert law at OD values of above 1.5 (Equation 3.10). The system produces scans through a Xenon flash lamp, a monochromatic light source, which enable the use of wavelengths ranging from 190-800 nm. The light passes through the cell chambers towards the detector, a photomultiplier, that radially scans the transmitted light. The raw intensity data collected from these
scans are used to calculate the absorbance values following the Beer-Lambert law:

\[ A = \varepsilon l c \]  

where \( A \) is absorbance, \( \varepsilon \) is the molar extinction coefficient per cm, \( c \) is the molar concentration and \( l \) is the optical pathlength which is determined by the centre piece type used for the optical system, which is 1.2 cm in this thesis.

A limitation to the absorbance detection system is it is comparatively slower (about 90 sec/cell on the older AUC and less than 20 sec/cell on the new Optima AUC) than the other detection systems (interference is about 5 sec/scan). This means that fewer absorbance scans are collected in the same experimental timeframe (Chadee-Luciano et al., no date).

### 3.1.4 Rayleigh interference detection system

The governing principle behind Rayleigh interference optics, is that the velocity of light is decreased as it passes through a higher refractive index. The refractive index is the ratio of the speed of light in a vacuum to the velocity of light in the solution. Monochromatic light passes through two parallel slits, that are located below each of the cell chambers. As the light waves pass through the solutions in the cells, they undergo interference producing a band of alternating light and dark “fringes” given as \( \Delta \Upsilon \). This interference pattern is detected by a charge-coupled device camera (CCD camera) (Figure 3.2, B) (Cole et al., 2008).

The refractive index of the sample is generally higher than that of the reference buffer, due to the sample having a higher solute concentration, therefore the sample wave is slower than the reference buffer wave. Thus, the positions of the fringes shift vertically proportionally to the concentration difference, which will change as the solute is sedimenting, altering the concentration distribution. The fringes have equal horizontal spacing. The difference in refractive index between the two solution is given as \( \Delta n \) and therefore fringes can be calculated by:

\[ \Delta \Upsilon = \frac{\Delta n l}{A} \]
Figure 3.2 Absorbance and interference optical systems used in an analytical ultracentrifuge

A is a schematic of the absorbance optical system and B is a schematic of the interference optical system. The dashed line indicates the path along which the light travels. Adapted from (Ralston, 1993; Planken et al., 2008).
where $l$ is the optical path length and $\lambda$ is the wavelength of the light source. For macromolecules in solution the refractive index increment is:

\[
\Delta n = c \frac{dn}{dc}
\]

where $\frac{dn}{dc}$ is the refractive increment of the macromolecule and $c$ is the concentration of that macromolecule. A sample with a concentration of 1 mg/ml approximately amounts to 3.25 fringes.

Interference optics can be applied to more systems, as it has infinite signal linearity so there are no theoretical limitations on the concentration of the sample. Moreover, it has a high signal-to-noise ratio of approximately 1000 to 1, meaning that samples lower in concentration be easily be studied (Cole et al., 2008). However, the interference system is more susceptible to time-independent noise caused by any optical imperfections such as dust, scratches on lens and mirrors or dirt. This is owing to interference optics not relying on signal from chromophores, thus signal comes from everything in the system. Due to this source of noise, a good agreement between the reference buffer and sample buffer is advised and can be achieved through prolonged sample dialysis with several buffer changes. Another source of noise is radially independent, as fringes cannot be tracked through certain features, such as menisci. Both time-independent and radially independent noise should be removed before analysis of the data and can be done so easily using analysis suites such as SEDFIT (Schuck, 2000; Schuck et al., 2002).

3.1.5 Sedimentation velocity experiments

A sedimentation velocity experiment is the main type of experiment run on an analytical ultracentrifuge. This experiment starts with a sample with uniform dispersity when a high angular velocity is applied, this causes rapid sedimentation of the solute toward the bottom of the chamber. Thus, over time the concentration of the solute near the meniscus is depleted as the solute’s concentration near the bottom of the chamber is increased, resulting in a sharp boundary (Figure 3.3) (Ralston, 1993). It is not possible to probe the velocity of individual molecules, however the rate at which this boundary moves toward the
bottom of the chamber can be measured, leading to the determination of the diffusion coefficient ($D$):

**Equation 3.13**

$$D = \frac{RT}{Nf}$$

where $R$ is the gas constant (8.3145 $J \text{ mol}^{-1} K^{-1}$) and $T$ the absolute temperature (in Kelvin). $N$ is Avogadro’s number (6.023 x 10^{23}). The diffusion coefficient is dependent on the effective size of the molecules. The ratio of the sedimentation to diffusion coefficient gives the molecular weight:

**Equation 3.14**

$$M = \frac{sRT}{D(1-\psi \rho)}$$

where $M$ is the molar weight of the solute. This equation can be rearranged to give the Svedberg equation. A sedimentation velocity can be defined by the Lamm equation (Gilbert *et al.*, 1959), that describes the sedimentation and diffusion of the solute in the chambers under a centrifugal force:

**Equation 3.15**

$$\frac{\partial c}{\partial t} = D \left[ \frac{\partial^2 c}{\partial r^2} + \frac{1}{r} \frac{\partial c}{\partial r} \right] - s\omega^2 \left[ r \frac{\partial c}{\partial r} + 2c \right]$$

This equation forms the basis for the analysis of sedimentation velocity experiments on programs such as SEDFIT.

### 3.1.6 Analysis of sedimentation velocity data

In this thesis all sedimentation velocity data is analysed using the program SEDFIT (Schuck, 1998, 2000; Schuck *et al.*, 2002). The Lamm equation is employed to fit the sedimentation boundaries, and the output is then converted to give a continuous size-distribution $c(s)$ plot, in which each peak represents a different species, and the peak positions define the sedimentation coefficients of the species that are present. Multiple peaks indicate that multiple species are present. These peaks can be integrated in order to give the peaks’ relative abundance. The broadness of the peak is of interest as a broad peak can indicate different conformations, multiple species in fast equilibrium with each other or
Figure 3.3 AUC cell and sedimentation velocity experiment

**A** Diagram of the cell used in AUC experiments, of note is the sector shape of the reference buffer and sample chambers.

**B** Diagram of the data collected in a sedimentation velocity experiment, showing the progression over three timepoints, with boundaries moving in radial position, mirroring the movement of the solute boundary down the sector chamber. Adapted from (Ralston, 1993).
small modifications to the main molecule are present. This is particularly pertinent for glycosylated system, in which differences in the presence or absence of glycosylation within the same sample are expected. Goodness-of-fit parameters, including residual plots, boundary fits, and root mean square deviation (RMSD), are inspected to ensure that an adequate fit has been reached.

3.2 Small angle scattering

Small angle scattering (SAS) in this thesis refers to both small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS). SAS techniques facilitate the study of molecules and assemblies in solution, where they are often assumed to have random orientations. This will give a low-resolution structure of the molecule, with an approximate real space resolution of about 2-4 nm (Wright et al., 2015). SAS allows users from diverse scientific disciplines to probe molecular structures within an extensive size range of a few kilo Daltons (kDa) to giga Daltons (GDa) (Kikhney et al., 2015). SAS allows samples to be probed under physiologically and industrially relevant conditions. However, due to the low-resolution nature of these techniques, interpreting the data is aided by drawing in as much a priori information about the system as possible (Lipfert et al., 2007; Putnam et al., 2007; Rambo and Tainer, 2013; Perkins et al., 2016). The primary basic parameters that can be investigated using SAXS and SANS with respect to bio-macromolecules include global dimensions, shape and flexibility. In this thesis, SAS data is combined with data from both AUC experiments and atomistic modelling. This is shown to be a powerful approach, that allows for the filtering and analysis of functionally relevant three-dimensional biological structures, which can result in a clearer understanding of structure-function relationship of the molecule.

3.2.1 Principles of small angle scattering

X-ray and neutron scattering techniques generally probe the same regions of reciprocal space (within instrumental limitations) and use similar wavelengths. The primary difference is the nature of the radiation used. Specifically, they differ in the subatomic particles from which the scattering occurs; in SAXS this occurs
Figure 3.4 Scattering from a particle

Diagram of two scattering domains (labelled A and B) in a macromolecule. The incident radiation $k_i$ is scattered ($k_s$) from the particle. The scattered beam is in phase with the incident radiation by a factor of $1\lambda$ at an angle of $2\theta$. Adapted from (Perkins et al., 2008). Macromolecule in the background is an IgG1 antibody, PDB ID: 1HZH (Saphire et al., 2001).
from the electrons and in SANS from the atomic nuclei. This diffraction is described by Bragg’s Law:

**Equation 3.16**  \[ \lambda = 2d \sin \theta \]

where \( \lambda \) is the wavelength of the incident beam, \( d \) is the diffraction spacing, and \( 2\theta \) is the scattering angle between the incident beam and the scattering beam.

Thus, scattering can be parameterised by:

**Equation 3.17**  \[ Q = \frac{4\pi \sin \theta}{\lambda} \]

where \( Q \) is the scattering vector which is described by:

**Equation 3.18**  \[ Q = k_i - k_s \]

where \( k_i \) is the incident wave vector and \( k_s \) is the scattered wave vector. Low \( Q \) values correspond to small angles and larger distances, which represent the macromolecules’ global structure and size, and high \( Q \) values correlate to large angles and shorter distances that capture the hydration layer/solvent interactions. The \( Q \) range for SAS experiments in this thesis is approximately 0.03-4.40 nm\(^{-1}\) for SAXS measurements and 0.06-5.00 nm\(^{-1}\) for SANS measurements. SAS beamlines are typically optimised such that the direct, unscattered beam of radiation is well collimated onto the detector and is as small as possible in size using, for example the pinhole camera principle or focussing devices. This allows the smaller angles to be measured with good signal to noise ratio. This is important in biological macromolecules as we often wish to obtain excellent data of the Guinier region at low \( Q \). Scattering waves are a composite of both coherent scattering, where the wave travels in a single wave in one direction, and incoherent scattering, which is random scattering that occurs from individual atoms in the system. Incoherent scattering intensity is \( Q \)-independent and can be manifested as flat background noise. The sources of this incoherent scattering can be instrumental (Hollamby, 2013).

In SAXS every electron scatters the incident X-ray beam with the same amplitude (given in femtometres, fm), and the scattering waves interfere
constructively therefore the scattering amplitude of an atom is proportional to the atomic number. In the neutron case, different isotopes of the same atom have different scattering amplitudes. The terms, scattering amplitude and scattering lengths can be used interchangeably. Contributions from all the scattering lengths combine to give the scattering curve (Table 3.1) (Perkins, 1988).

3.2.3 Debye equation

The Debye scattering equation describes the geometrical relationship between individual atoms within the macromolecule, by taking into account the spherically averaged intensities of the sum of all scattering atoms with the scattering lengths in the macromolecule (Debye, 1915):

\[ I(Q) = \sum_a \sum_b f_a f_b \frac{\sin(r_{ab}Q)}{r_{ab}Q} \]

where \( f_a \) and \( f_b \) correspond to X-rays, which are substituted for \( b_a \) and \( b_b \) for neutrons, and are the scattering lengths of the electrons or nuclei at points A and B in the macromolecule. \( r \) describes the distance between A and B (Figure 3.4). In this equation, \( Q \) and \( 2\theta \neq 0 \). Theoretically, if \( 2\theta \) is zero, the scattering is in phase and scattering intensity \( I \) is at its maximum. At zero scattering angle, the intensity of scattering, given as \( I(0) \), is a direct measure of molecular mass (Perkins, 1988; Putnam et al., 2013; Skou et al., 2014).

3.2.4 Contrast variation

The scattering density \( \rho \) of the solute or solvent is given by the sum of all scattering lengths divided by the molecular volume. The difference between the density of the solute and solvent gives rise to the density contrast \( \Delta \rho \), given by:

\[ \Delta \rho = \rho_v - \rho_s \]

where \( \rho_v \) is the mean scattering density of the solute and \( \rho_s \) is the mean scattering density of the solvent (Perkins, 1988).

In SANS experiments, the majority of scattering lengths are positive and similar between different nuclei, thus neutrons are able to strongly interact with
atoms of a similar scattering length. The scattering lengths of $^1\text{H}$ (protium) is -3.742 fm, and $^2\text{H}$ (deuterium) is 6.671 fm. In biological systems, in general, composed of proteins, carbohydrates, nucleic acids and lipids, each contains different proportions of $^1\text{H}$ atoms, and thus have different scattering densities (Table 3.2).

In a SANS experiments the ratio of $\text{H}_2\text{O}$ to $^2\text{H}_2\text{O}$ in the buffer can be altered to vary $\Delta \rho$, owing to $^1\text{H}$ and $^2\text{H}$ exchange. At a $\Delta \rho = 0$ there will be no scattering, consequently said part of the molecule will be “invisible” (Perkins, 1988; Mahieu et al., 2020). This makes SANS a powerful technique, as it allows the scattering from various components in a system to be deconvolved (Clifton et al., 2012; Lapinaite et al., 2020; Papachristodoulou et al., 2020). In this thesis, SANS samples are dialysed into deuterated buffer in order to eliminate incoherent scattering contributions from the buffer.

### 3.2.5 Hydration shell

The hydration shell is principally made up of a single layer of water molecules that is strongly hydrogen bonded to solvent exposed residues. Beyond this hydration shell, the water is considered “bulk” water, where water molecules experience weaker hydrogen bonding interactions between themselves (Svergun et al., 1998). The electron density of the hydration shell is comparable to that of the protein rather than to that of the bulk water. This is due to the bound water molecule occupying a volume that is approximately 20% smaller than that of bulk water (Perkins, 1986). This hydration shell is detected by SAXS. This shell can be made invisible in SANS experiments when dialysing into $^2\text{H}_2\text{O}$ buffer, forcing the exchange of $^1\text{H}$ to $^2\text{H}$ in the shell (Perkins, 2001), where the hydration shell has approximately the same scattering density as bulk water.

### 3.2.7 Small-angle X-ray scattering instruments

All small angle X-ray scattering data were collected at the Diamond Light Source, Didcot, Oxfordshire, United Kingdom (Figure 3.5). The SAXS beamline used was B21 (Figure 3.6). Diamond is a high brilliance synchrotron, which is

Table 3.1 Scattering lengths elements.
Scattering lengths are given in femtometres (fm, $10^{-15}$ m). Adapted from (Perkins, 1988).

<table>
<thead>
<tr>
<th>Element</th>
<th>Atomic Number</th>
<th>$f (2\theta = 0^\circ C)$ (fm)</th>
<th>$b$ (fm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrogen</td>
<td>$^1\text{H}$</td>
<td>1</td>
<td>-3.74</td>
</tr>
<tr>
<td></td>
<td>$^2\text{H}$</td>
<td>1</td>
<td>6.67</td>
</tr>
<tr>
<td>Carbon</td>
<td>$^{12}\text{C}$</td>
<td>6</td>
<td>16.9</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>$^{14}\text{N}$</td>
<td>7</td>
<td>19.7</td>
</tr>
<tr>
<td>Oxygen</td>
<td>$^{16}\text{O}$</td>
<td>8</td>
<td>22.5</td>
</tr>
<tr>
<td>Phosphorous</td>
<td>$^{31}\text{P}$</td>
<td>15</td>
<td>42.3</td>
</tr>
</tbody>
</table>

**Table 3.2 Scattering densities of biological macromolecules.**

Scattering densities for X-rays is given in electron density (e.nm$^{-3}$), for neutrons scattering densities are given as % out of 100% $^2\text{H}_2\text{O}$. Adapted from (Perkins, 1988)

<table>
<thead>
<tr>
<th>Macromolecule and Solvent</th>
<th>X-rays (e.nm$^{-3}$)</th>
<th>Neutrons (% $^2\text{H}_2\text{O}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$O</td>
<td>334</td>
<td>0</td>
</tr>
<tr>
<td>$^2\text{H}_2\text{O}$</td>
<td>334</td>
<td>100</td>
</tr>
<tr>
<td>Proteins</td>
<td>410 - 450</td>
<td>40 - 45</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>490</td>
<td>47</td>
</tr>
</tbody>
</table>
essential for collecting high quality scattering data for relatively dilute biological samples. SAXS beamlines have undergone a process of modernisation, making data collection very easy, allowing for sample mail-in, automatic sample changers, and size-exclusion chromatography directly prior to being subjected to the beam, which eliminates any aggregates and results in near perfect buffer matching (Round et al., 2015; Jordan et al., 2016; Brennich et al., 2017; Cowieson et al., 2020).

Diamond achieves high flux by acceleration of particles (in this case electrons) in a synchrotron. This process starts in a linear accelerator (also known as a linac), where an electron gun creates an electron beam, that is packaged in bunches and is accelerated, reaching energies of 100 MeV at Diamond. Once these energies have been reached, these bunches enter the booster synchrotron, where magnets are used to change the direction of the electrons and to follow the bend of the ring, causing the electrons to undergo further acceleration to reaching energies of 3 GeV. The booster ring will then fill the storage ring. In the storage ring these electrons circulate at a constant speed almost at the speed of light, which is maintained using vacuum and radiofrequency cavity units. The storage ring is not a perfect circle but a pentacontagon, a 50-sided polygon, with a circumference of 561.6 metres. The ring is composed of bending magnets and insertion devices (also known as undulators). As the electrons move around the storage ring they encounter undulators which force a change in direction of travel, and synchrotron light is emitted. This light spans the whole electromagnetic spectrum, including X-rays, which are then allowed to pass through apertures into the beamlines. The wavelength of selected range is achieved by a monochromator. Bending magnets force the bending of the electron beam, that also forces a loss of energy in the emission of light (Diamond Light Source, no date).

3.2.8 B21 beamline

SAXS instruments have undergone upgrades in the last decade that mean that data collection is now largely automated for low viscosity solutions, allowing the beamlines to become ‘high throughput’. This has been achieved at several beamlines by upgrades in sample environments primarily using autosamplers,
Figure 3.5 Beamline plan of Diamond Light Source.
Diagram showing beamline geography at Diamond Light Source, the beamline utilised in this thesis is B21, a soft condensed matter (in burgundy) beamline for high throughput SAXS. Taken from (Diamond Annual Review 2019, 2019).
such as the Arinax BioSAXS sample handling robot (Pernot *et al.*, 2013; Round *et al.*, 2015), which is also used on B21 at Diamond. BioSAXS allows for “batch-mode”, through the use of 96-well plates, facilitating the exploration of many sample conditions with minimal effort. Alternatively, 200 μl PCR strips or 1.5 ml microfuge tubes can be handled by a robot. In batch mode small sample volumes of about 30-100 μl are required.

B21 is also equipped with an in-line Agilent 1260 HPLC and a size exclusion column (several are available and chosen based on the sample being run) through which samples can be passed just prior to coming in contact with the X-ray beam, in a procedure known as SEC-SAXS. SEC-SAXS allows for removal of non-specific aggregates, mitigates heterogeneity, and can allow for the investigation of different species. The resulting SAS curves undergo buffer subtraction, where the buffer scattering curve is subtracted from that of the sample scattering curve. For optimal data, exact buffer matching is required, achieved through extensive dialysis. However, in SEC-SAXS the buffer matching is even better than can be achieved through prolonged dialysis, as the buffer subtracted is taken from the flat line at the beginning of the HPLC trace for each sample. However, SEC-SAXS requires higher sample concentrations due to dilution, and additionally, the sample may interact with the column resin to potentially alter macromolecular conformations. The column pressure may also induce changes to the macromolecular conformations. This may be problematic for certain users, and in such instances it may be more appropriate that SEC-SAXS is avoided (Rambo, 2017). The sample exposure unit is temperature controlled within a temperature range of 2°C to 60°C, this temperature range was availed of in Chapter 6 when investigating the thermostability of IgG1 and IgG4.

The B21 beamline (Figure 3.6) uses a bending magnet source, which delivers approximately 3 x 10^{12} photons.s^{-1}. The beam enters from the bending magnet source and passes through white-beam slits which define the size of the beam which hits the monochromator crystals / multilayers. The monochromator on B21 is a double multilayer model (Axilon, Germany), which allows for wider X-ray bandwidth than say a Si(111) monochromator and thus delivers higher flux. The beam then goes through beam-defining slits and is then focussed vertically and horizontally via a toroidal mirror to deliver suitable collimation. The beam then goes through a final set anti-scatter slits to remove parasitic scattering, thus
minimising background noise and allowing access to low $Q$ values. The guard slits are the final pair of slits encountered prior to the beam meeting the sample.

The sample is loaded from the SEC column, 96-well plate, microtube or PCR tube, into an *in vacuo* quartz capillary. The sample within the capillary is continually moved during data collection in order to minimise radiation damage. Data can be collected in static mode; this requires a smaller sample volume of approximately 15 μl. Data can be collected for about one second to several minutes. In this thesis, SAXS data for each sample were collected for approximately 30 seconds to one minute (*Cowieson et al.*, 2020). Data are collected as frames in intervals of one frame per second. These frames are averaged to give a collective curve. Averaging and buffer subtraction has been done using the SAS data analysis suite SCATTER (*Rambo*, 2021) or in later data collections data was automatically subtracted and averaged using an inhouse (at B21) automated pipeline. As the frames are collected over time, it is possible to investigate each frame and exclude any frames with evidence of radiation damage from the final average. Data subtraction will be considered in more detail in Section 3.2.13. Due to the relative stability of the IgGs studied in this thesis there was no need to discard any of the collected frames. Between each sample run the sample capillary is quickly cleaned and dried, using a surfactant and ethanol mixture, with water rinse and air drying, generally within a minute, making high throughput data collection possible. On average, in batch mode, one sample can be measured in about three minutes.

### 3.2.9 Small angle neutron scattering

All small angle neutron scattering data presented in this thesis was collected at the ISIS Neutron and Muon Source, Didcot, Oxfordshire, United Kingdom (*Figure 3.7*). Two small angle scattering beamlines were used, SANS2d and Zoom (*Figure 3.8*). ISIS is a spallation neutron and muon source, divided into two target stations, target station 1 (TS1) and target station 2 (TS2), the latter being the newer target station built in 2003, and at which SANS2d and Zoom are located (*Thomason*, 2019). At ISIS neutrons are produced by a spallation process. This process starts with the production of negative hydrogen ions, H⁻
Table 3.3 Characteristics of beamlines B21 (Diamond), Sans2d and Zoom (ISIS).

Data collated for B21 from (Cowieson et al., 2020), for Sans2d from (ISIS Neutron and Muon Source, no date; Heenan et al., 2011) and updated by Dr James Doutch, and for Zoom data provided courtesy of Dr James Doutch.

<table>
<thead>
<tr>
<th></th>
<th>B21</th>
<th>Sans2d</th>
<th>Zoom</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wavelength range</strong></td>
<td>0.089 – 0.13 nm</td>
<td>0.175 - 1.65 nm</td>
<td>0.175 – 1.65 nm</td>
</tr>
<tr>
<td><strong>Q-range</strong></td>
<td>0.026 – 3.4 nm(^{-1})</td>
<td>0.02 – 20 nm(^{-1})</td>
<td>0.02 - 5 nm(^{-1})</td>
</tr>
<tr>
<td><strong>Beam size at sample</strong></td>
<td>1102 × 240 (\text{µm})</td>
<td>6 - 14 mm diameter (circular apertures/pinholes)</td>
<td>4 x 4 mm to 15 x 15 mm (square/ rectangular apertures)</td>
</tr>
<tr>
<td><strong>Flux at sample</strong></td>
<td>10(^{12}) photons/second</td>
<td>&gt;10(^6) neutrons/cm(^2)/s</td>
<td>&gt;10(^6) neutrons/cm(^2)/s</td>
</tr>
<tr>
<td><strong>Sample-detector distance</strong></td>
<td>3.7 m static</td>
<td>4 - 12 m variable</td>
<td>4 - 10 m</td>
</tr>
</tbody>
</table>
ions, made up of a proton and two electrons, which are produced using an electric discharge in the ion source. These ions are accelerated and divided into bunches by a Radio Frequency Quadrupole accelerator. These bunches are then further accelerated in a linac, which consists of four 10 metre tanks containing copper drift tube electrodes. These ions continue to be accelerated in the 800 MeV synchrotron ring, which has a circumference of 163 metres. As ions enter the ring, they encounter thin alumina foil that strips away the electrons, leaving behind a beam of protons. Radio-frequency electric fields located in ten accelerating cavities further accelerate the proton beam. Each proton will undergo about 10,000 revolutions until it reaches approximately 84% of the speed of light. At this point the protons are separated into two large bunches, each of which is ejected from the synchrotron by fast kicker magnets that rise to 5000 amps in 100 nanoseconds.

The proton bunches continue to travel and collide with a tungsten target, roughly the size of a house-brick. This collision is what generates neutrons through a process called spallation. As the high energy protons collide with the nuclei of the target atoms, neutrons are released. These neutrons are slowed down by an array of hydrogen moderators around the target. Slower, ‘cold’ neutrons are more amendable to condensed matter experiments, such as in this thesis. Finally, neutrons are allowed to passed through apertures to the neutron instruments, i.e., the beamlines (ISIS Neutron and Muon Source, no date). At ISIS the proton beam is pulsed, each one producing a beam of “white” neutrons, containing a full spectrum of neutron wavelengths. Since each neutron wavelength has a distinct speed, the time taken from the neutrons to go from target to detector can be measured, making this a time-of-flight technique (Rogers et al., 2009).

3.2.10 Sans2d and Zoom beamlines

Sans2d (Heenan et al., 2011) (Figure 3.8A) and Zoom (ISIS Neutron and Muon Source, no date) (Figure 3.8B) are part of a suite of three small angle scattering beamlines found at TS2, the other being Larmor which was not used in this work. The two beamlines are comparable, with Zoom being the more recently commissioned of the two. The first component on SANS2d and Zoom is
Figure 3.6 Components of the B21 beamline
Provide courtesy of Dr Robert Rambo and Dr Nathan Cowieson.
Figure 3.7 Beamline plan at ISIS Neutron and Muon Source.

Diagram of ISIS Neutron and Muon Source, showing target station (TS1) and target station 2 (TS2) and the arrangement of the beamlines. In this thesis the beamlines used are Sans2d and Zoom, at TS2, both of which are small angle scattering beamlines coloured in green. Taken from (Thomason, 2019).
the supermirror bender. This acts to filter high energy neutrons out, and ensures
the sample position is not in direct line of sight with the source. The bender is
followed by a counter-rotating double disc chopper. This allows tuning of the
wavelength range used, and, for a 4m camera length, is often set to give a beam
with neutron wavelengths within a range of 0.175 nm to 1.65 nm. Five removable
neutron guides then direct and collimate the beam towards the sample position.
On SANS2d the guides are 2 m each and on Zoom the one guide is fixed at 4 m.
Scattering intensity is measured using two moveable $^3$He multiwire proportional
counters (STFC Technology, Harwell, UK), which together have an area of
approximately 1 m$^2$ and can be independently positioned (Heenan et al., 2011).
For standard biological experiments, the detectors are positioned so that one
detector is at 4 m and the other is at 2.5 m (Perkins et al., 2008). In Zoom the
detector is composed of an array of 192 cells, each of 8 mm diameter, and 1
metre long, as a $^3$He gas tube (STFC Technology, Harwell, UK). Adjusting the
detector distances, a feature not available on B21, make it possible to study
structures of different sizes. The optimal detector distance is determined by
considering the $Q$ range:

$$ Q_{\text{minimum}} L \leq \pi $$

where $Q_{\text{minimum}}$ is the minimum value required for data analysis, and $L$ is the
maximum length of the molecule (Perkins, 1988).

In a SANS experiment samples are placed in cylindrical quartz cells often
termed “banjo” cells, which are housed in a sample changer that allows for up to
~40 such cells. The cells are manufactured in a variety of a path lengths;
generally 1-mm thick (generally used for 0-30% $^2$H$_2$O) or 2-mm (for use with 40-
100% $^2$H$_2$O) (Perkins et al., 2008). Thinner cells are used for high H2O content
to minimise incoherent scattering. In this thesis, 2-mm quartz cells were used for
all SANS data collection. For high D$_2$O fractions a 2-mm cell will give a higher
count rate from a dilute sample, but for some protein system expressing enough
protein to fill a 2-mm may be challenging. The sample changer temperature can
be set by the user; this functionality was used for data collection in Chapter 6.

Sample volumes of about 600 $\mu$l are required to fill the “banjo” cells. Prior
to experiments, the samples are subjected to extensive dialysis with $^2$H$_2$O buffer,
in order to eliminate scattering from the buffer background. As in SAXS, scattering from the buffer is required to be collected for buffer subtraction. Care should be taken to seal the cell openings properly to prevent $^1$H and $^2$H exchange. The banjo cells should be clean, ensuring there are no surface fingerprints or scratches that may interfere with scattering. A challenge often encountered in SANS experiments is the formation of small air bubbles, generally triggered by sudden changes in temperature, for example when taking samples out of a fridge. Bubbles in the sample can increase transmission due to lower effective scattering mass. These bubbles can be removed by careful agitation in extreme cases with a gel pipette tip. Bubbles can interfere with scattering and should therefore be removed.

The advantage of SANS over SAXS is that samples can generally be recovered as neutrons are non-destructive, owing to the beam having a lower energy (Table 3.3). In SAXS, the high energy X-ray photons lead to the generation of free radical species, which most often cause proteins to aggregate in the beam or otherwise trigger bond lysis which can disturb the chemical structure of the protein. However, same samples may become aggregated in $^2$H$_2$O buffer (Perkins et al., 2008), especially as, unlike SAXS where data collection happens in a matter of seconds, protein samples are often dilute and require longer data collection times, on the magnitude of hours per sample. This means that samples are left within the sample changer for extended periods of time, and if they are sensitive to aggregation samples will not be useable for later experiments. In this thesis, all samples were sufficiently stable in $^2$H$_2$O buffer and can be left for several hours without danger of oligomerisation. This problem can in future be surmounted by the addition of in-line SEC systems to SANS beamlines (Jordan et al., 2016).

Prior to data collection the instrument is readied by aligning the beam stop, using a thin Teflon strip, a strong isotropic scatterer, to reveal the beam stop on the detector. Cadmium is used as a measurement of ‘darkness’ as it absorbs neutrons (Perkins et al., 2008), this is done to check for background scattering on some SANS instruments, such as on SANS2d, and this run is subtracted from experimental data during data reduction. On Zoom cadmium is not usual measured. Each sample’s transmission is measured from time to time during the
experiment, to ensure that dialysis in $^2$H$_2$O to complete, as $^2$H$_2$O has a higher neutron transmission than $^1$H$_2$O.

3.2.11 Buffer subtraction

Following data collection, frames from SAXS experiments are averaged and data from SANS experiments are merged and reduced. In both SANS and SAXS, experimental data for the sample and its accompanying buffer are collected separately, for the purpose of buffer subtraction. The final scattering curve is obtained by subtracting the averaged buffer profile from the averaged protein-solution profile (Skou et al., 2014). For SAXS this buffer subtraction can be achieved using SCÅTTER (Rambo, 2021), or done automatically using an automatic pipeline at the beamline which performs the same procedure on a computing cluster during the measurements. For the SANS case data merging, reduction, and subtracting is performed using MANTID (Arnold et al., 2014). For both SAXS and SANS the data are collected on 2-D detectors, which are radially averaged in the reduction software such as MANTID (Arnold et al., 2014) on SANS2D, and Data Analysis WorkbeNch (DAWN) (Basham et al., 2015) on B21.

3.2.12 Small angle scattering Guinier analysis

Once the final 1D scattering curve is obtained, one-dimensional $I(Q)$ analysis by Guinier analysis allow the derivation of parameters including the radius of gyration ($R_G$), cross-sectional radii of gyration (in IgGs two cross-sectional radii can be calculated, $R_{XS-1}$ and $R_{XS-2}$) and the forward scattered intensity, $I(0)$. The $R_G$ values are determined from the slope of the linear curve fit of the Guinier plot, $\ln I \propto Q^2$ and $I(0)$ is determined by the y-intercept of the curve extrapolation. The $R_G$ is a measure of the structure’s elongation and is analogous to the sedimentation coefficient derived from AUC experiments. The $R_G$ and $I(0)$ are observed at low $Q$ values on a Guinier plot, of $\ln I(Q)$ against $Q^2$ (nm$^{-2}$) (Figure 3.9) (Glatter et al., 1982):

Equation 3.22

$$\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}$$
Figure 3.8 SAS beamlines at ISIS Neutron and Muon Source

A and B are diagrams of Sans2d and Zoom respectively. Provided courtesy of Dr James Doutch
Figure 3.9 Analysis of IgG scattering curves using Guinier, distance distribution $P(r)$ and dimensionless Kratky (legend in overleaf).
Figure 3.9 Analysis of IgG scattering curves using Guinier, distance distribution $P(r)$ and dimensionless Kratky (continued).

A is a typical scattering curve for IgG, from which the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values can be extracted from the Guinier region. B is a typical distance distribution function IgG1 and IgG4 that is shown in black with two peak maxima, $M_1$ and $M_2$. In red a typical distance distribution curve for IgG3, is shown with three peak maxima $M_1$-$M_3$ discernible. C is a typical IgG dimensionless Kratky plot. Adapted from (Hui, 2019). D is a visual key to the parameters generated from Guinier and distance distribution ($P(r)$) analysis for IgG1, IgG3 and IgG4. The $R_G$ is a measure of the structure’s elongation. The $R_{XS-1}$ and $R_{XS-2}$ parameters, are available for elongated proteins such as antibodies. In an antibody, the $R_{XS-1}$ value describes the cross-sectional distances between pairs of Fab-Fc and Fab-Fab regions. $R_{XS-2}$ gives the cross-sectional distance of a single Fab or Fc region. For IgG1 and IgG4 $M_1$ corresponds to the distances within a Fab or Fc domain and $M_2$ corresponds to distances between pairs of Fabs or Fab-Fc. For IgG3 $M_1$ corresponds to distances within Fab and Fc domains, $M_2$ corresponds to distances between Fab pairs and finally $M_3$ corresponds to the Fab-Fc distances.
In this equation the $Q.R_G$ values should be between 0.5 and 1.5 (Perkins et al., 2011). Due to the physical beamstop, the experimentally derived $Q$ ranges do not start at 0. Therefore the fitted $I(0)$ is a good approximation. Concentration dependence, i.e. the effect of increased protein concentration on aggregation, can be accessed from the $I(0)$ values divided by the sample concentration (in mg/ml) to give $I(0)/c$. For successful Guinier analysis, good quality low $Q$ data is an imperative. This means that low $Q$ data should be as linear as possible. Any increased non-linearity in this region can be indicative of poor data subtraction, or aggregation caused by radiation damage (Skou et al., 2014). A downturned low $Q$ region is indicative of repulsive interparticle interference, alternatively an upturned curve at low $Q$ indicates aggregation (Blanchet et al., 2013).

The $R_{XS-1}$ and $R_{XS-2}$ parameters, are available for elongated proteins such as antibodies. In an antibody, the $R_{XS-1}$ value describes the cross-sectional distances between pairs of Fab-Fc and Fab-Fab regions. $R_{XS-2}$ gives the cross-sectional distance of a single Fab or Fc region. These parameters can be derived by finding a linear fit at large $Q$ values, in a $\ln I(Q).Q$ against $Q^2$ (nm$^{-2}$).

The larger $Q$ values used in fitting for $R_{XS}$ values should not overlap with the smaller $Q$ values used to fit $R_G$. The $R_{XS}$ value is given by:

$$\text{Equation 3.23} \quad \ln[I(Q)Q] = [I(Q)Q]_{Q \to 0} - \frac{R_{XS}^2 Q^2}{2}$$

### 3.2.13 SAS distance distribution function analysis

The SAS curves can be converted from reciprocal space in (nm$^{-1}$) by an indirect Fourier transform to real space (in nm) to give the distance distribution function $P(r)$:

$$\text{Equation 3.24} \quad P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q)Q r \sin(Qr) \, dQ$$

The $P(r)$ curve gives the distributions of all of the interatomic distances $r$ between all of the volume elements within the macromolecule. Structural parameters can be extracted from this curve, analogous to those derived from Guinier analysis, including the $R_G$ and $I(0)$. It should be noted that, although the
$R_G$ derived from the $P(r)$, calculated using GNOM is similar to the one derived by Guinier analysis, there are often slight difference owing to the full scattering curve being used to calculate $R_G$ in GNOM, whereas in Guinier analysis, only a limited range of low $Q$ values are used for linear fitting. For clarity, both $R_G$ values are presented in this thesis. Significant deviation between the two $R_G$ values can be indicative of problems with either the Fourier transform or the quality of the experimental data.

In this thesis, $P(r)$ were calculated using GNOM (Semenyuk et al., 1991). The program requires user inputs, such as the maximum and minimum $Q$ values, and a $D_{\text{max}}$, which describes the maximum dimension of the molecule. An acceptable $D_{\text{max}}$ is one which allows for the curve to be stable, i.e., there are no significant changes to the peak(s) present, if it is increased beyond $L$ (which also describes the maximum dimension of the molecule) (Figure 3.9, B). For an adequate $D_{\text{max}}$ one should expect to see a clean tailing off of the $P(r)$ curve, rather than a sharp end. The $L$ of the molecule is given when the $P(r)$ value (on the $y$-axis) for the largest $r$ value (in nm) reaches zero (Perkins et al., 2008).

$P(r)$ curves generally result in one, or more peaks, the maxima of which are designated $M$. The $r$ (nm) for $M$ describes commonly occurring distance found between two volume elements in the molecule. For IgGs, large multidomain molecules, one can expect to see two peaks (for IgG1 and IgG4) or three peaks for IgG3. These are designated $M_1$, $M_2$ and $M_3$. For IgG1 and IgG4, the $M_1$ and $M_2$ values are comparable to the $R_{\text{XS}-1}$ and $R_{\text{XS}-2}$ values derived from Guinier analysis. Whereby $M_1$ corresponds to the distances within a Fab or Fc domain and $M_2$ corresponds to distances between pairs of Fabs or Fab-Fc. For IgG3 there are typically three $M$ peaks, and this is due to its elongated hinge (Chapter 7). Therefore, $M_1$ corresponds to distances within Fab and Fc domains, $M_2$ corresponds to distances between Fab pairs and finally $M_3$ corresponds to the Fab-Fc distances.

### 3.2.14 Dimensionless Kratky Plot

The dimensionless Kratky plot, where $(Q.R_G)^2.I(Q)/I(0)$ is plotted against $Q.R_G$ and is derived from the wide angle scattering curve. This can be used to interpret folded state and flexibility. The shape of the dimensionless Kratky plot
varies depending on the folded state and flexibility of the protein. For globular proteins, the dimensionless Kratky plot will have a bell-shaped profile with a clear maximum value at $Q R_G = \sqrt{3}$ of approximately 1.1. In a $Q$ range of $0 < Q R_G < 3$ nearly all curves will be superimposable. For a random chain, such a disordered protein, the curve will have an increasing angle and a plateau is reached at between 1.5 and 2 on the $y$-axis (Figure 3.10). (Receveur-Brechot et al., 2012).

For IgG1, IgG2 and IgG4, dimensionless Kratky plots typically have two bell-shaped peaks followed by a plateau (Figure 3.9C).

### 3.3 Thermostability investigations using the UNit

Thermostability of IgG1 and IgG4 with and without the Fc-glycans was investigated in Chapter 6. A device called the UNit (Unchained labs, California, United States), probes protein stability, by giving information on melting ($T_m$) and aggregation ($T_{agg}$) temperatures. The UNit requires very small sample sizes, 9 $\mu$l, with concentrations of between 0.1-150 mg/ml (for IgG), with the possibility of running up to 48 samples at a go, and results being achieved in under two hours. It simultaneously measures full spectrum fluorescence (250-720 nm) and static light scattering (SLS) at two wavelengths 266 and 473 nm.

The fluorescence data allows the derivation of $T_m$ and SLS data following aggregation and gives $T_{agg}$. Following both these parameters simultaneously can provide insight into when unfolding events result in aggregation (Figure 3.11). SLS is proportional to the mean solute particle mass, when the signal is very low it indicates that the proteins are in monomeric state, as signal increases it indicates that aggregates are present. The samples are loaded into a bespoke sample holder, referred to as the “UNi”, which are sealed in an airtight clasp, to prevent sample evaporation (Unchained labs, 2016). The UNit has a sample temperature range of 15-95°C, with a heating rate of 0.01-10°C per minute.

Two main types of experiments can be run on the UNit, isothermal stability experiments or thermal recovery experiments. In this thesis, all data were generated in an isothermal stability experiment. An isothermal stability run studied in a gradual temperature ramp from a set minimum temperature to a set maximum temperature, with temperature increasing in predetermined intervals.

For experiments described in this thesis, experimental conditions were set to a
minimum temperature of 20°C to a maximum temperature of 95°C with intervals of 1°C per minute and a 30 second equilibration at each temperature.

In order to analyse SLS data, the temperature versus SLS intensity is plotted and fitted to a Boltzmann equation on OriginPro (Version 2020b, OriginLab Corporation, Northampton, MA, USA) in order to extract $T_{agg}$. **Equation 3.25** for Boltzmann equation fitting is given by OriginPro (*Boltzmann, no date*):

**Equation 3.25**

$$ y = A_2 + \frac{(A_1-A_2)}{1+e^{(x-x_0)/d_x}} $$

where $A_1$ is the $y$-axis minimum, $A_2$ is the $y$-axis maximum, $x_0$ is the 50% threshold where $\frac{A_1-A_2}{2}$, and $d_x$ is a temperature constant. The $T_{agg}$ value is derived from the point closest to the onset of aggregation at a signal change of 10%, therefore this point was found from the fitted Boltzmann curve by:

**Equation 3.26**

$$ T_{agg} = dx \left[ \ln \left( \frac{A_1-A_2}{10\% - A_2} - 1 \right) \right] + x_0 $$

Fluorescence data are converted to the barycentric mean (BCM) data by the instrument and are then plotted against temperature (converted to absolute temperature for curve fitting). This data were fitted in parts to a two state (**Equation 3.27**) or three state unfolding model depending on the number of transitions observed in the curve (*Zhang et al., 2020*). Due to the number of transitions present in the experimental data collected for this thesis, it became easier to adequately fit the data in segments, so for all curves the first segment encompassed the first transition ($T_{m1,app}$) and values between approximately 46.9-72.9°C, and the second transition ($T_{m2,app}$) is fitted to values of approximately 61.9-93.9°C, with variation of a few degrees between curves. $T_m$ values are presented as apparent $T_{m,app}$ as it assumes that unfolding is convoluted with irreversible aggregation from the unfolding protein for a given temperature ramp rate.

The two-state unfolding model is given by:

**Equation 3.27**

$$ I_T = \frac{(I_N+aT) + (I_D+bT) K}{1+K} $$

where $T$ is experimental temperature, $T_{m,app}$ is the temperature at which half the protein is denatured. $I_N$ and $I_D$ are spectroscopic signals of the protein at each temperature ($T$), at the native, and fully denatured state, respectively. $a$ and $b$ are
Figure 3.10 Dimensionless Kratky plots for disordered and globular proteins.

In red a typical dimensionless Kratky plot for a globular protein and in blue for a disordered very flexible protein. Adapted from (Skou et al., 2014).
Figure 3.11 Deriving apparent melting and aggregation temperature from fluorescence and SLS curves

In light blue, normalised barycentric mean (BCM) signal given as a percentage plotted against temperature in Celsius (°C), $T_{m1,app}$ and $T_{m2,app}$ derived from a three-state unfolding model fitting are labelled. In dark blue, normalised SLS signal plotted against temperature in °C, $T_{agg}$ derived from fitting of the Boltzmann equation is labelled. Adapted from (Unchained labs, 2016).
the baseline slopes of the native and denatured region of the curve. K is an equilibrium constant for the transition between native and denatured state given by:

\[
K = \exp \left( \frac{\Delta H_{vH}}{R} \left( \frac{1}{T_{m,app}} - \frac{1}{T} \right) \right)
\]

where \( \Delta H_{vH} \) is the van’t Hoff enthalpy and \( R \) is the gas constant. All temperatures terms in these equations are absolute temperatures in Kelvin (K).

The three-state unfolding model is given by:

\[
I_T = \frac{(I_N + aT) + (I_D + bT) K_1 + (I_D + cT) K_2}{1 + K_1 + K_2}
\]

where \( K_1 \) is the equilibrium constant for the transition between the native and intermediate state and \( K_2 \) is the equilibrium constant between the intermediate and denatured states given by:

\[
\begin{align*}
K_1 &= \exp \left( \frac{\Delta H_{vH1}}{R} \left( \frac{1}{T_{m,app1}} - \frac{1}{T} \right) \right) \\
K_2 &= \exp \left( \frac{\Delta H_{vH2}}{R} \left( \frac{1}{T_{m,app2}} - \frac{1}{T} \right) \right)
\end{align*}
\]

where \( T_{m,app1} \) and \( T_{m,app2} \) are the temperatures at which the protein is half denatured for the transition of native to intermediate and intermediate to denatured states, respectively. \( a, b \) and \( c \) are the baseline slopes of the native, intermediate and denatured region of the curve. \( \Delta H_{vH1} \) and \( \Delta H_{vH2} \) are the van’t Hoff enthalpy for both transitions.

Prior to analysis, BCM or SLS data points are normalised to the highest data point (making it 100%) for each sample dataset given by:

\[
y \text{ normalised} = \frac{y - y_{minimum}}{y_{maximum} - y_{minimum}} \times 100
\]

3.4 Modelling

3.4.1 SASSIE

To realise the full potential of SAS data, it can be coupled to atomistic modelling. Advancements in computational capacity has made it possible to
model large macromolecules, such as antibodies, generating large conformer libraries using atomistic methods, such as Torsion Angle Monte Carlo simulations. The SASSIE project, developed by the Collaborative Computational Project for Small-Angle Scattering (CCP-SAS), has integrated several features (Figure 3.12), into one web-based graphical user interface (GUI), in order to make the methodology accessible to non-experts and aims to render the production and analysis of potentially millions of models highly intuitive (Curtis et al., 2012; Perkins et al., 2016). Given SASSIE’s web-based nature (website link: https://sassie-web.chem.utk.edu/sassie2/), the user does not need to install several software packages, also data is stored on a centrally managed server, so the user does not need large storage space available on their own machines. As SAS data takes into account random orientations present in solution, the ability to examine thousands of conformations, recognises the true potential of the technique.

SASSIE requires each user to create an account, accessed by login in credentials, such as username and password. Inside SASSIE there are several modules, including “Tools”, “Build”, “Interact”, “Simulate”, “Calculate”, “Analyze”, “Beta” and “Retired”. Beta and Retired modules can be accessed in the first instance by going through user settings. With each module there are several programs that one can run. For example, the Beta module houses the program Torsion Angle Monte Carlo (TAMC), which is used to create a library of structures, the Calculate module has SasCalc, which is used to generate the theoretical scattering curves for each model and Analyze has the Chi-Square program where one can compare each theoretical curve to an experimental curve.

Data generated using each program are saved and can be accessed via the “Filing cabinet”. As jobs are running, they can be managed via “Job Manager”. All data are saved and can be used between programs, so for example the models generated in TAMC will be saved so when later running SASCALC one simply needs to select the model set from TAMC via the filing cabinet. The modules and programs used in this thesis are elaborated on in the next sections, and an overview of the whole modelling workflow used for all computational modelling projects in this thesis is graphically outlined in Figure 3.12.
3.4.2 Building starting structure

To start a modelling project an adequate starting structure must be obtained. The guiding principle here is the well-known saying “garbage in garbage out”. Accordingly, efforts are made to create a starting structure that matches the system as closely as possible. To ensure this, two things are required; firstly high-resolution atomic coordinate data of the structure, taken from the Protein Data Bank (PDB), and secondly an amino acid sequence that matches that of the experimental system. The PDB has several high-resolution Fab and Fc structures, which were used to generate full-length IgG1, IgG4 and IgG3 in this thesis. The hinge that connects the Fab and Fc regions, with a specific amino acid sequence for each IgG subclass was generated by a PyMOL script build_seq (PyMOL Script Repository, Queen’s University, Ontario, Canada). Full-length IgG1 and IgG4 structures are also available on the PDB, however these structures had a few residues missing in their hinge and Fc regions. This is not unusual as it is often difficult to resolve flexible residues by X-ray crystallography.

When building IgG1 A33 and IgG4 A33, the amino acid sequences for the samples used experimentally were kindly provided by UCB (who also provided the IgG1 and IgG4 antibodies used in this thesis). Therefore a homology model could be created by inserting the amino acid sequence of these antibodies into the known crystal structures using a program called Modeller (Version 9.19) (Webb et al., 2016).

For IgG3, samples were purified from human myeloma samples, by Dr Margaret Goodall (University of Birmingham) and thus there is no known sequence for the IgG3 samples used in this thesis. There are some IgG3 Fc structures in the PDB however there were no available Fabs. A structure was built using the available Fc structure, IgG4 A33 Fabs and a hinge was created using the aforementioned PyMOL script.

An oligosaccharide was attached to each C-H2 domain of the IgGs at Asn297. The oligosaccharides were approximated as complex-type biantennary oligosaccharide structures with a Man3GlcNAc2 core and two NeuNAc.Gal.GlcNAc antennae. A glycan structural template was taken from the GitHub repository (website link: https://github.com/dww100) and energy
This diagram illustrates the workflow employed to generate and analyse model ensembles. “Input”, in orange box, are generated on the user’s machine. “SASSIE”, in blue box, is a web-based user interface for simulating motions in the start structure, generating theoretical scattering curves and comparing these curves to experimental data. “Output”, pink box, illustrates the files that can be extracted from SASSIE for inspection on user’s local machine. In the grey box, good and bad fits are depicted. Adapted from (Perkins et al., 2016)
minimised for one nanosecond prior to attachment, to achieve a fully relaxed structure. This glycan was added to the Fc region by bringing the C1 atom in the first GlcNAc residue to within 0.14 nm to the N sidechain atom of Asn\textsuperscript{297} in the C\textsubscript{12} domain of IgG1, while ensuring no steric clashes between the Fc residues and the glycan chain. The PDB file was then opened on Discovery Studio (Dassault Systèmes BIOVIA, San Diego) where “CONECT” records were created for these glycosidic bonds.

Following this, the structures were uploaded to the “Glycan Reader and Modeler” input generator on CHARMM-GUI (Chemistry at Harvard Molecular Mechanics) (Jo et al., 2011) (website link: http://www.charmm-gui.org/?doc=input/glycan). CHARMM-GUI is a powerful and ever evolving online input generator that allows users to edit structures, add bond information such as disulphide bonds, and build post-translational modifications such as glycosylation. Newer updates allow users to generate input compatible with the CHARMM forcefield (Raman et al., 2010a; Huang, 2014) for molecular dynamics simulations, including solvating the system (i.e. add water molecules), and designating periodic boundary conditions. This input can be generated for a number of different simulation engines including NAMD (Phillips et al., 2005; Lee et al., 2016). This is an incredibly powerful and useful tool, for both experienced and novice users. CHARMM-GUI is used throughout this thesis for the generation of input files, such as protein structure files (PSF) and parameter files. For IgG3 CHARMM-GUI was used to create input files for running a molecular dynamics simulation on the hinge region, in order to try to capture some of the dynamics in the long polyproline hinge.

### 3.4.3 Energy minimisation and molecular dynamics

After all the input files have been created, the start structures are subjected to energy minimisation in order to remove any steric clashes and to adjust any physically unrealistic bond angles. This is achieved using the CHARMM36 forcefield on the simulation engine NAMD. During energy minimisation the potential energy of the system is reduced to a local minimum point. A system is energy minimised prior to running molecular dynamics (Serdyuk et al., 2007).
In a molecular dynamics simulation, the motions of all atoms in the system are simulated over time by solving the Newtonian equations of motion. A forcefield, such as CHARMM36, is a collection of equations for potential energies, forces, and how atoms interact with respect to one another, parameters such as bond lengths and angles (Brooks et al., 2009). A time step is given as an interval for which the forces displacing atoms are calculated to give the atoms new position. Molecular dynamics simulations in this thesis use UCL’s high performance computing (HPC) clusters to achieve longer time scales (up to 100 nanoseconds).

3.4.4 Torsion angle Monte Carlo simulations

Torsion angle Monte Carlo (TAMC) simulations differ from molecular dynamics simulations in that most of the atoms are held rigid, and only short, assigned linkers are allowed to be varied. This allows for the rapid generation of random conformations, that survey a large conformational space. This is different to molecular dynamics where the time-dependent nature of the simulation means that it will require incredibly long runs, which are computationally expensive, to cover the same extensive conformational landscape as one could with a Monte Carlo simulation. Monte Carlo methods samples a system through randomly generated states (Curtis et al., 2012). In a Monte Carlo simulation for each state, a random number is used to create changes in the protein backbone. The energy of each state is calculated, and subsequently each state is either accepted or discarded depending on the Metropolis criterion (Metropolis et al., 1953). This criterion calculates the energy change between the last accepted state and the new state and the current ‘temperature’ of the system in the simulation. It allows steps that decreases the energy of the system, but it applies a probabilistic rejection to steps that increase the energy of the system (Putnam et al., 2007).

In the TAMC simulations outlined in this thesis the linker regions that are varied are found between the Fab region and the hinge, and between the hinge and the Fc region. This is to mimic the movement of the antibody as one would expect there to be the greatest flexibility in these regions. As the linkers are varied by sampling the rotations in torsion angles, it results in the movement of the Fab or Fc regions. Torsion angles are angles of bonds between N and Cα (ψ), Cα
and C (ψ), and N and C (ω) in the polypeptide main chain. As the linker is varied the system checks for steric clashes. The basis on which steric clashes are checked can be by heavy atoms, all, background, or atom name. In this thesis, steric clashes were checked on the basis heavy atoms. The maximum angle, at which the atoms in the variable regions are varied, can be adjusted by the user. Any models that have steric clashes, are discarded, models that are free of steric clashes are saved. TAMC simulations often result in slightly stretched or broken linkers, which are corrected by a short energy minimization.

### 3.4.5 Generating theoretical scattering curves of models using golden vector-scattering curve calculations

In order to generate theoretical scattering curves for each model in the library of models created by TAMC, golden vector scattering curve calculations (Watson et al., 2013) are used. This is achieved through the SasCalc program in the Calculate module on SASSIE. Calculating a theoretical scattering curve can be in principle a simple task, calculated by the Debye equation (Equation 3.19), however doing this for a large system on an atomistic level is incredibly computational expensive due to the double summation of atom pairs $O(N^2)$. SasCalc, reduces computation load by employing exact $O(N)$ expression for the intensity $I(Q)$ in several scattering directions. The $Q$ vectors describing each direction are generated by the golden ratio and are isotopically distributed.

To calculate the scattering curve Euler's formula is employed:

\[
I(Q) = |\sum_j^N b_j \cos(Qr_j)|^2 + |\sum_j^N b_j \sin(Qr_j)|^2
\]

where $b_j$ is the generalised scattering length and $r_j$ is the atomic position.

All macromolecules in solution adopt random conformations. To take this into account, the rotationally averaged intensity $I(Q)$ is calculated, by averaging $I(Q)$ in all scattering directions. These scattering directions are drawn from a quasi-uniform lattice on a sphere. To construct the lattice, where the number of grid points $n$ can be finely adjusted, is given by Equation 3.33 where $n$ can be any odd integer:
Equation 3.33

\[
Q_x^{(k)} = Q \cos \left( \sin^{-1} \left( \frac{2k}{n} \right) \right) \cos \left( \frac{2\pi k}{\phi} \right),
\]

\[
Q_y^{(k)} = Q \cos \left( \sin^{-1} \left( \frac{2k}{n} \right) \right) \sin \left( \frac{2\pi k}{\phi} \right),
\]

\[
Q_z^{(k)} = \frac{2kQ}{n}
\]

where \( k \) runs over \( \left\{ -\frac{(n-1)}{2}, \ldots, 0, \ldots, \frac{(n-1)}{2} \right\} \) and \( \phi = \frac{1 + \sqrt{5}}{2} \) is the golden ratio. The golden ratio is used to find the angle of turn between two consecutive points in the spiral lattice for them to be distributed on the sphere in a quasi-uniform manner (González, 2010).

The rotationally averaged intensity is then approximated by:

Equation 3.34

\[
I(Q) = \frac{1}{n} \left\{ \sum_{k=-\frac{(n-1)}{2}}^{\frac{(n-1)}{2}} I[Q^{(k)}] \right\}
\]

The SasCalc program can adjust for contrast variation, used in SANS experiments. The user is also prompted to specify \( I(0) \), the number of \( Q \) values and the maximum \( Q \) value. In this thesis, \( I(0) \) was set to 1, in order to more easily achieve the same number of points with the same \( Q \) spacing for comparison of theoretical and experimental curves that were collected across different concentrations. The number of \( Q \) values is dependent on the number of points in the experimental curve. For example SAXS curves will have hundreds more points than a SANS curve, as the detector has a smaller pixel size, and therefore smaller \( Q \) intervals can be taken resulting in more points. Finally, the maximum \( Q \) value is in all cases set to 1.5 nm\(^{-1}\), \( Q \) values larger than 1.5 nm\(^{-1}\) often have a high signal to noise ratio and thus fitting to these points may result in overfitting. As the number of theoretical \( Q \) values generated are evenly spaced, interpolation of the experimental curve is required in order to compare this to the theoretical curves. Interpolation is where the experimental points are adjusted to have an even \( Q \) spacing. This enables the direct comparison of experimental and theoretical curves. SasCalc is more advanced than older theoretical curve determination programs such as SCT (Wright et al., 2015), as it considers all atom contributions to the curve rather than employing a simplified coarse-grained bead model of the molecules. This advancement is consistent with improvements and better access to computational power, i.e. HPCs.
3.4.6 Curve comparison and filtering the best-fit models.

As a final step, the theoretical curves for every model in the ensemble is compared to the experimental scattering curve. In this thesis the program Chi-Square filter was used, found in the Analyze module of SASSIE. First the degree of similarity was calculated by the \( R \)-factor option, given by:

\[
R - factor = \frac{\sum |I_{\text{experimental}}(Q) - \eta I_{\text{theoretical}}(Q)|}{\sum |I_{\text{Experimental}}(Q)|} \times 100
\]

where \( \eta \) is the scaling factor. Lower \( R \)-factors indicates better fits. A fit of less than 10\% is considered satisfactory, with less than 5\% being considered good (Perkins et al., 2011). The theoretical and experimental scattering curves must have the same number of points. Next, the Chi-Squared program will generate an \( R_G \) versus \( R \)-factor distribution. Generally, this distribution should be approaching a somewhat symmetrical shape, indicating that \( R_G \) values on either side of the experimental \( R_G \) have been surveyed. This implies that both extremes of the conformational space have been explored. Here the best-fit structures are those that are found in the minimum of this distribution i.e., have the lowest \( R \)-factors and match the experimental \( R_G \).

In this thesis the best-fitting 100 models with the lowest \( R \)-factors were extracted for further analysis, including Guinier analysis. The theoretical dimensionless Kratky plot and \( P(r) \) curves of the best fitting model are also compared to the experimentally derived curves to ensure that their shape is consistent. Finally, in order to understand the functional significance of the best-fitting conformations, they are visually examined on VMD (Humphrey et al., 1996). In order to appreciate the relative motions of each domain independently, for example to deconvolve the motions of the Fc region, the Fabs or the hinge will be aligned. These aligned models are then used to create density plots to visually show the conformational space that is explored by each of these regions.

Other methods for extracting models have been employed in the scattering community namely Ensemble Optimisation Method (EOM) which fits an ensemble of conformations to the SAS data, and is primarily used to fit systems that have significant inherent flexibility such as intrinsically disordered proteins and multi-domain proteins with unstructured regions (Bernadó et al., 2007; Tria
et al., 2015). Here an ensemble of conformers that survey a wide conformational space are created and sub-ensembles of conformers are selected to fit the experimental data. The methods presented in this thesis are comparable in that a large ensemble of conformers are also generated however, rather than fitting sub-ensembles only the 100 best-fit conformers are analysed. The EOM is very interesting and probably represents a step in the right direction for fitting of SAS data as it allows for a better understanding of the populations that are present in the sample.

3.4.7 Hydrodynamic property calculations

The models’ theoretical hydrodynamic properties, namely the sedimentation coefficient, analogous to those derived from AUC experiments, can be calculated using HullRad (Fleming et al., 2018). This is a new freely available Python program that has some advantages when compared to earlier methods, such as HYDROPRO (Ortega et al., 2011), predominantly due to it being much faster and easier to use, allowing for calculations on large libraries of models. HullRad uses a convex hull to model the hydrodynamic volume of the molecular structure. A convex hull is three-dimensional mathematical construct, that is defined as the smallest convex envelope that contains a set of points. The convex hull of a molecular model captures several parameters that would influence its hydrodynamic properties including, the overall volume of the molecule, the shape asymmetry and surface roughness.

To run the program, one must have the Hullrad script, which is accessible from (website link: http://52.14.70.9/Code.html), and all the models are saved as PDB files. There is generally no need to modify the script, however caution is advised when dealing with glycosylated structures. Glycan nomenclature has not as yet been properly standardised and may mean that adjustments need to be made to three-letter glycan names listed in the HullRad script to reflect those found in the molecule PDB. When running the script several hydrodynamic properties are calculated, however for the purpose of this thesis, the only property extracted and discussed is the sedimentation coefficient. This is calculated by the program at 20°C in water, using a modified version of the Svedberg equation:
where $\rho_{20,w}$ is the water density at 20°C (g mL$^{-1}$), $N_A$ is Avogadro’s number, $\eta_0$ is water viscosity at 20°C (poise), the $10^8$ term converts $R_{T,calc}$ from Å to cm to convert the answer to the correct order of magnitude for the sedimentation coefficient. $R_{T,calc}$ (Å) is the translational hydrodynamic radius calculated from a sphere with the equivalent volume $V_{TH}$ as the expanded convex hull for a particular molecule and is corrected by the translational shape factors $F_T$. This is calculated by the program prior to calculating the sedimentation coefficient:

$$R_{T,calc} = F_T \times \frac{3 V_{TH}}{4 \pi}$$

where $F_T$ is the translational diffusion shape factor, calculated as, the major axis ($a$) of a prolate ellipsoid of revolution is estimated from the maximum dimensions of the initial convex hull of a molecule. Assuming the volume of a prolate ellipsoid is $4\pi ab^2/3$, the minor axis is calculated by:

$$b = \sqrt[3]{\frac{3V_{H}}{4\pi a}}$$

where $V_H$ is the equivalent volume of the initial convex hull for a particular molecule. From the axial ratio, $a/b$, a translational shape factor based on Perrin’s equation (Perrin, 1936) is obtained. A prolate ellipsoid of revolution is used to model the nonsphericity of all molecules.

The vertices of the convex hull are placed at the outermost atom centres, and thus the side chain conformations encoded by the PDB file will influence the convex hull for a particular protein. This is not representative of side chain motions that occur in solution. To average surface-exposed side chains, their conformation is represented by a single, unified pseudoatom placed along the $C\alpha$ – $C\beta$ axis. This pseudoatom is pushed outward by a distance equal to the radius of a sphere corresponding to respective side chain volume.

Another consideration is that, in solution a structure is hydrated, however the input models used are anhydrous. To mitigate this, the initial convex hull
volume is increased by expanding the planes of the hull outward. This expansion is empirically determined by optimising the agreement of the calculated parameters with experimental hydrodynamic data.
Chapter 4 Solution structure of deglycosylated human IgG1 shows the role of C_{H2} glycans in its conformation.
4.1 Introduction

Immunoglobulins are an important class of humoral (adaptive) glycoproteins, composing of 82-96% protein and 4-18% carbohydrate (Vidarsson et al., 2014). The most abundant immunoglobulin class in human serum is IgG, a Y-shaped molecule that is found as four subclasses, namely IgG1, IgG2, IgG3 and IgG4, which occur at 8.0 mg/ml, 4.0 mg/ml, 0.8 mg/ml and 0.4 mg/ml in serum (Hamilton, 1987). The structure of IgG1 is formed as two Fab regions, which bind with high specificity and affinity to a specific antigen, and an Fc region which binds to Fcγ receptors (FcγRs) on the surface of immune cells, and to complement C1q to initiate the classical pathway of activation (Figure 4.1A). In IgG1, the Fabs and the Fc regions are connected by a 15-residue hinge held together by two disulphide bonds at Cys226 and Cys229 in the hinge (Vidarsson et al., 2014). As well as being the most abundant class in serum, IgG1 is the predominant class used in therapeutic antibodies, where 66 IgG1 monoclonal antibodies are commercially available out of a total of 89 antibody-based products in a multibillion-dollar industry (Reichert, 2021).

The conserved N-linked glycosylation in the Fc region plays a key functional role in all four IgG subclasses (Figure 4.1B). A complex-type biantennary glycan with a Man3GlcNAc2 core and two NeuNac.Gal.GlcNAc antennae is attached at Asn297 on each CH2 domain (Deisenhofer, 1981) (Figure 4.2). However, the glycan structure is chemically heterogeneous (Wuhrer et al., 2007). The Fc glycans modulate several IgG-Fc effector functions (Wada et al., 2019). Glycoengineering is becoming increasingly important in order to elicit desired responses. For example, afucosylated IgG1 is able to activate a natural killer antibody-dependent cellular cytotoxicity response more effectively for reason of its increased affinity for FcγRIIIa (Shields et al., 2002; Pereira et al., 2018). Deglycosylated antibodies may be good candidates for therapeutics if a lower propensity to activate inflammatory cascades is desirable, because the removal of glycan reduces the IgG interactions with Fc receptors (Jefferis, 2007). Thus, aglycosylated (when protein is expressed without glycosylation) and deglycosylated IgG1 (when glycan is removed after protein expression) have an abrogated or reduced propensity for binding to the FcγRs and C1q but does not affect Fab antigen binding.
Figure 4.1 The human IgG1 domains and its glycosylation.

**A** The heavy chains are comprised of VH, CH1, CH2 and CH3 domains, and the light chains are comprised of VL and CL domains. The heavy chains are connected by two Cys-Cys disulphide bridges at Cys226 and Cys229. An N-linked oligosaccharide at Asn297 is present on each of the CH2 domains. The hinge region connecting the Fab and Fc regions was constructed from 23 residues 216EPKSCDKTTHCPCPAPELLGGP238.

**B** At the left, the glycosylation of IgG1 Fc at two Asn297 residues in the Fc region is shown as stick models. The three hinge tripeptides that were conformationally varied in the TAMC searches are in red and circled in red. The central schematic shows the glycosylation pattern used in this study (N-acetyl glucosamine, GlcNAc; mannose, Man; galactose, Gal; N-acetyl neuraminic acid, NeuNAc. At the right, the detailed view of a single CH2 domain with its glycan chain is shown, with the glycan colours coordinated with those in the schematic.
Figure 4.2 Sequence alignment of human IgG4 (legend overleaf).
Figure 4.2. Sequence alignment of human IgG1 (continued). (A-G) The IgG1 A33 sequence was kindly provided by UCB. The IgG1 6a and 19a sequences were taken from Rayner et al. (2015). The IgG1 b12 sequence was taken from the crystal structure (PDB ID: 1HZH). The Fc sequence was taken from the crystal structure 4W4N (PDB ID: 4W4N). A, B, the V\textsubscript{L} and C\textsubscript{L} domains; C-E, the V\textsubscript{H} and C\textsubscript{H}1 domains and the hinge, with the TAMC-varied tripeptides identified in green; F, G, the C\textsubscript{H}2 and C\textsubscript{H}3 domains, with Asn\textsuperscript{297} in blue. EU sequence numbering was used, and the complementarity-determining regions (CDRs) sequences were identified in red. Beneath the alignments, consensus symbols indicated the degree of conservation, where (*) indicates full conservation, (:/) indicates conservation between groups of strongly similar properties based on the Gonnet PAM 250 matrix, (.--) indicates conservation between groups of weakly similar properties, and a space indicates no conservation.
Deglycosylated antibodies have been of interest to treat autoimmune disorders (Collin et al., 2008; Crispin, 2013). The deglycosylation of pathogen neuromyelitis optica anti-aquaporin-4 IgG in patients reduced its complement-dependent and antibody-dependent cell-mediated cytotoxicity with a reduction in antigen binding, giving it therapeutic potential (Tradtrantip et al., 2013).

Structural studies of IgG antibodies are crucial to understand their function. While many crystal structures are known for Fab regions, only one crystal structure is known for the full-length human IgG1 (PDB ID: 1HZH) (Saphire et al., 2007), together with other full-length structures for two murine IgG subclasses (Harris et al., 1995). These crystal structures only provide a static view of full-length IgG1, and do not take into proper account a mobile and flexible hinge region, that allows for the independent movement of the Fab and Fc regions in solution (Jefferis, 2012). Crystal structures for glycosylated human Fc regions revealed the two glycans to be found at the centre of the Fc region in contact with each other (PDB IDs: 4W4N, 4KU1, 4BM7, 3AVE, 4Q74, 4BYH, 1H3X) (Krapp et al., 2003; Matsumiya et al., 2007; Crispin et al., 2013; Yu et al., 2013; Ahmed et al., 2014; Frank et al., 2014; Kiyoshi et al., 2015). Crystal structures of the Fc region in complexes with the FcγRI and FcγRIII receptors showed similar modes of receptor binding to the upper part of the Fc region with many conserved contacts, despite their varying affinities to the Fc region (Sondermann et al., 2000; Radaev et al., 2001; Kiyoshi et al., 2015). In these structures, very few contacts were found between the glycans and the FcγR, and there is no information on the full IgG1 structure after deglycosylation, making it unclear what role the glycans have. Nonetheless, crystal structures for deglycosylated Fc regions showed a more compact conformation of the Cn2 domains compared to the glycosylated Fc region, indicating that the glycans stabilise the Fc regions (Krapp et al., 2003). In the deglycosylated Fc structures, the C′E loop (Gln293-Phe303) of the Cn2 domain that is involved in FcγR binding is more disordered (Borrok et al., 2012). Previous solution studies of glycosylated and deglycosylated human Fc gave a larger radius of gyration Rg for deglycosylated Fc than that of glycosylated Fc (Borrok et al., 2012). These studies suggest that Asn297 glycosylation is important to stabilise the open conformation of the Cn2 domains.
The effect of the Fc glycan chains on the full IgG structure is not well understood. To address this question, small angle X-ray and neutron scattering (SAXS, SANS) and analytical ultracentrifugation (AUC) were jointly applied to intact IgG1 as powerful solution structural techniques for studying biological macromolecules. SAXS provides data sets measured in high positive solute-solvent contrast, in which the contribution of the hydrophilic surface regions of the glycoprotein are accentuated, while SANS measured with heavy water buffers provides data sets measured in high negative solute-solvent contrast, in which the contribution of the buried hydrophobic core of the glycoprotein is accentuated (Perkins, 1986, 2001; Svergun et al., 1998). For the dense hydration shell surrounding the protein, X-rays are able to see this, however neutrons measured in D2O buffers do not detect the denser hydration shell which becomes less visible, so the protein becomes seen without most of its hydration shell visible. The reproducibility of the two data sets validates the individual SAXS and SANS data sets because radiation effects in SAXS and aggregation in heavy water, for SANS may perturb the output of either method. Their utility is much enhanced by the development of atomistic modelling of the SAXS and SANS data sets using molecular dynamics and Monte Carlo methods (Perkins et al., 2016). Previous atomistic scattering modelling with glycosylated IgG1 revealed that IgG1 is conformationally stable, even in different buffer conditions, and exhibited an asymmetric IgG1 structure in which the arrangement of the Fab regions permitted the Fc region to bind to its FcγR and C1q ligands with no steric clashes (Rayner et al., 2015; Wright et al., 2019). Here, this joint SAXS-SANS-AUC approach is applied together with atomistic modelling to show that deglycosylation does in fact result in a more flexible Fc structure within IgG1, in turn affecting the receptor-binding function of IgG1.

4.2 Materials and Methods

4.2.1 Purification and composition of IgG1

IgG1 A33 (148 kDa) was generously supplied by Dr John O’Hara and Dr Berni Sweeney (UCB). Its enzymatic deglycosylation was performed using peptide:N-glycosidase F (PNGase F) (35.5 kDa; New England Biolabs, Massachusetts, USA) for reason of its ability to remove glycans completely from
glycosylated Asn residues (Plummer et al., 1984). To digest the glycans, 3.7 μl PNGase F (1850 activity units) was used to deglycosylate 150 μl IgG1 A33 (16.3 mg/ml). Native IgG1 was incubated at 37°C for time points of 1 hour (TP1), 6 hours (TP6) and 10 hours (TP10). Each deglycosylated IgG1 sample was filtered, through three successive dilutions using Amicon Ultra-0.5 ml centrifugal filters (100 kDa cut-off), which simultaneously allowed the PNGase F to pass through the membrane, while concentrating the deglycosylated IgG1 sample. Immediately before SAXS, SANS or AUC measurements, glycosylated and deglycosylated IgG1 were purified by gel filtration to remove any non-specific aggregates using a Superose 6 Increase 10/300 GL column (Cytiva, Amersham, UK), then concentrated using Amicon Ultra-15 ml spin concentrators (100 kDa cut-off) and dialyzed at 4°C into 20 mM L-histidine, 138 mM NaCl, and 2.6 mM KCl buffer, pH 6.0. This histidine buffer was found to increase the stability of IgG1. The sequence of IgG1 A33 was aligned against those for IgG1 19a, 6a and b12 (PDB ID: 1HZH), and an IgG1 Fc structure (PDB ID: 4W4N) (Saphire et al., 2007; Kiyoshi et al., 2015; Rayner et al., 2015) (Figure 4.2). The N-linked glycans at Asn297 on the CH2 domains were assumed to be complex-type biantennary structures with a Man3GlcNAc2 core and two NeuNAc.Gal.GlcNAc antennae (Deisenhofer, 1981). From this sequence, the molecular mass of glycosylated IgG1 A33 was calculated to be 148.4 kDa, its unhydrated volume was 191.4 nm³, its hydrated volume was 252.0 nm³ (based on 0.3g of water/g of glycoprotein and an electrostricted volume of 0.0245 nm³ per bound water molecule), its partial specific volume \( \bar{\nu} \) was 0.731 ml/g and its absorption coefficient was 14.0 cm⁻¹ (1%, 1-cm pathlength, 280 nm) (Perkins, 1986). The molecular mass of deglycosylated IgG1 A33 was 144.0 kDa, its unhydrated volume was 186.7 nm³, its hydrated volume was 245.4 nm³, its partial specific volume \( \bar{\nu} \) was 0.733 ml/g and its absorption coefficient was 14.4 cm⁻¹ (Perkins, 1986). The X-ray and neutron scattering densities of glycan residues are similar to those for hydrophilic (polar) amino acid residues, these being slightly higher than those for hydrophobic (non-polar) amino acid residues (Perkins, 1986). The buffer density was measured on an Anton Paar DMA 5000 density meter at 20°C to be 1.00578 g/ml in light water. In heavy water, the density was 1.11106 g/ml. Buffer viscosities were measured on an Anton Paar AMVn Automated microviscometer.
at 20°C. The viscosity for the L-histidine buffer in light water, pH 6.0, was 0.010190 poise.

The completeness of deglycosylation was verified by Superose 6 gel filtration, SDS-PAGE, and mass spectrometry. In the Mass Spectrometry Facility at the Chemistry Department, University College London, the antibodies were analysed on an Agilent 6510 Quadrupole time-of-flight liquid chromatography mass spectrometry system (Agilent, UK). Ten µL of each sample was injected onto a PLRP-S, 1000A, 8 µM, 150 mm × 2.1 mm column, which was maintained at 60°C at a flow of 0.3 ml/min. The separation was achieved using mobile phase A (water with 0.1% formic acid) and B (acetonitrile, with 0.1% formic acid) using a gradient elution. The column effluent was continuously electrosprayed into the capillary electrospray ionization source of the Agilent 6510 QTOF mass spectrometer and electrospray ionization mass spectra were acquired in positive electrospray ionisation mode using the m/z range 1,000−3200 in profile mode. The raw data was converted to zero charge mass spectra using the maximum entropy deconvolution algorithm in the MassHunter software (version B.07.00). The glycan masses were calculated by subtracting the mass of the full glycosylated IgG1 from the partially deglycosylated glycoform giving the mass of a single glycan chain. The single glycan mass was also found by subtracting the mass of fully glycosylated IgG1 from that for deglycosylated IgG1 and halving this mass.

4.2.2 Sedimentation velocity data and analysis for IgG1

Analytical ultracentrifugation data for native and deglycosylated IgG1 in light water at timepoints TP1, TP6 and TP10 were obtained on two Beckman XL-I instruments equipped with AnTi50 rotors. Data were collected at 20°C at a rotor speed of 40,000 rpm in two-sector cells with column heights of 12 mm for approximately 6 hours. Sedimentation analyses were performed using direct boundary Lamm fits of up to 896 scans using SEDFIT (version 15.01b) (Schuck, 1998, 2000). SEDFIT resulted in size-distribution analyses c(s), for which the algorithm assumes that all species have the same frictional ratio f/f0. The final SEDFIT analyses (Table 4.1) used a fixed resolution of 200 and optimized the
Table 4.1 Experimental data by X-ray and neutron scattering and analytical ultracentrifugation for glycosylated and deglycosylated IgG1

<table>
<thead>
<tr>
<th></th>
<th>Concentration (mg/ml)</th>
<th>Rg (nm)</th>
<th>RxS-1 (nm)</th>
<th>RxS-2 (nm)</th>
<th>L (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG1 Glycosylated</td>
<td>3.60</td>
<td>5.13 ± 0.27</td>
<td>2.50 ± 0.17</td>
<td>1.42 ± 0.12</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>3.19</td>
<td>5.10 ± 0.28</td>
<td>2.50 ± 0.18</td>
<td>1.41 ± 0.14</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>2.05</td>
<td>5.07 ± 0.34</td>
<td>2.49 ± 0.20</td>
<td>1.39 ± 0.14</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>1.36</td>
<td>5.02 ± 0.37</td>
<td>2.46 ± 0.22</td>
<td>1.35 ± 0.16</td>
<td>17</td>
</tr>
<tr>
<td>IgG1 TP1</td>
<td>4.89</td>
<td>5.16 ± 0.30</td>
<td>2.50 ± 0.18</td>
<td>1.43 ± 0.13</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>3.25</td>
<td>5.13 ± 0.28</td>
<td>2.50 ± 0.19</td>
<td>1.42 ± 0.13</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>2.01</td>
<td>5.08 ± 0.31</td>
<td>2.50 ± 0.19</td>
<td>1.41 ± 0.14</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>1.23</td>
<td>5.06 ± 0.41</td>
<td>2.49 ± 0.20</td>
<td>1.41 ± 0.16</td>
<td>17</td>
</tr>
<tr>
<td>IgG1 TP6</td>
<td>3.85</td>
<td>5.15 ± 0.29</td>
<td>2.51 ± 0.19</td>
<td>1.43 ± 0.13</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>2.94</td>
<td>5.12 ± 0.32</td>
<td>2.51 ± 0.18</td>
<td>1.41 ± 0.13</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>1.61</td>
<td>5.08 ± 0.35</td>
<td>2.51 ± 0.20</td>
<td>1.39 ± 0.14</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>5.03 ± 0.41</td>
<td>2.49 ± 0.25</td>
<td>1.38 ± 0.17</td>
<td>17</td>
</tr>
<tr>
<td>IgG1 TP10</td>
<td>4.29</td>
<td>5.19 ± 0.28</td>
<td>2.52 ± 0.18</td>
<td>1.43 ± 0.13</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>3.19</td>
<td>5.15 ± 0.31</td>
<td>2.51 ± 0.18</td>
<td>1.43 ± 0.13</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>1.34</td>
<td>5.08 ± 0.34</td>
<td>2.50 ± 0.21</td>
<td>1.42 ± 0.16</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>1.02</td>
<td>5.06 ± 0.41</td>
<td>2.50 ± 0.23</td>
<td>1.39 ± 0.18</td>
<td>17</td>
</tr>
<tr>
<td>Neutron data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG1 Glycosylated</td>
<td>2.60</td>
<td>5.27 ± 0.28</td>
<td>2.35 ± 0.18</td>
<td>1.18 ± 0.14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.38</td>
<td>5.26 ± 0.34</td>
<td>2.36 ± 0.20</td>
<td>1.10 ± 0.14</td>
<td>16</td>
</tr>
<tr>
<td>IgG1 TP1</td>
<td>4.78</td>
<td>5.34 ± 0.25</td>
<td>2.44 ± 0.20</td>
<td>1.19 ± 0.14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2.32</td>
<td>5.22 ± 0.63</td>
<td>2.42 ± 0.35</td>
<td>1.18 ± 0.25</td>
<td>16</td>
</tr>
<tr>
<td>IgG1 TP6</td>
<td>3.71</td>
<td>5.28 ± 0.74</td>
<td>2.42 ± 0.34</td>
<td>1.17 ± 0.19</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.78</td>
<td>5.27 ± 0.74</td>
<td>2.43 ± 0.36</td>
<td>1.16 ± 0.26</td>
<td>16</td>
</tr>
<tr>
<td>IgG1 TP10</td>
<td>2.73</td>
<td>5.31 ± 0.77</td>
<td>2.39 ± 0.32</td>
<td>1.12 ± 0.21</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>5.19 ± 0.89</td>
<td>2.38 ± 0.45</td>
<td>1.17 ± 0.35</td>
<td>16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analytical ultracentrifugation data</th>
<th>s20,w(S)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG1 Glycosylated</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>IgG1 TP1</td>
<td>2.32</td>
</tr>
<tr>
<td></td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td>0.59</td>
</tr>
<tr>
<td>IgG1 TP6</td>
<td>6.01</td>
</tr>
<tr>
<td></td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
</tr>
<tr>
<td>IgG1 TP10</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
</tr>
</tbody>
</table>

*Standard error of the mean for s20,w data as follows; ± 0.05 S for glycosylated, ± 0.02 S for TP1, ± 0.04 S for TP6 and ± 0.01 S for TP10.
c(s) fits by floating f/f₀ and the baseline until the overall root mean square deviations and visual appearance of the fits were satisfactory. The percentage of oligomers in the total loading concentration was derived using the c(s) integration function. The observed s values were normalised to s_{20,w} by:

\[ s_{20,w} = s_{T,B} \left( \frac{\eta_{T,B}}{\eta_{20,w}} \right) \left( 1 - \bar{\nu} \rho \right)_{20,w} \left( 1 - \nu \rho \right)_{T,B} \]

where s is the sedimentation coefficient, the subscripts \( T,B \) refers to the temperature of the buffer. \( s_{20,w} \) refers to water at 20°C. \( \rho \) is the solvent density, \( \eta \) is the solvent viscosity and \( \bar{\nu} \) is the protein partial specific volume.

### 4.2.3 X-ray and neutron scattering data and analyses for IgG1

X-ray scattering data were obtained during one beam session (October 2017) on Instrument B21 (Cowieson et al., 2020) at the Diamond Light Source at the Rutherford Appleton Laboratory (Didcot, UK), operating with a ring energy of 3 GeV, and an operational energy of 12.4 keV. A PILATUS 2M detector with a resolution of 1475 × 1679 pixels (pixel size of 172 × 172 µm) was used with a sample-to-detector distance of 4.01 m giving a Q range from 0.04 nm\(^{-1}\) to 4 nm\(^{-1}\) (where \( Q = 4 \pi \sin \theta / \lambda \); \( 2 \theta \) = scattering angle; \( \lambda \) = wavelength). The glycosylated IgG1 (1.4-5.4 mg/ml) and the TP1 (0.7-4.9 mg/ml), TP6 (1.0-3.9 mg/ml) and TP10 (1.0-4.3 mg/ml) samples in light water were loaded onto a 96 well plate and placed into an EMBL Arinax sample holder (Pernot et al., 2013; Round et al., 2015). This condition showed the antibody molecule as a hydrated structure in a high positive solute-solvent contrast (Perkins, 1986). An automatic sampler injected 30 µl of sample from the well plate into a temperature-controlled quartz cell capillary with a diameter of 1.5 mm. Data sets of 30 frames with a frame exposure time of 1 second each were acquired in duplicate as a control of reproducibility. Checks during data acquisition confirmed the absence of radiation damage. ScÂtter (version 3.0) was used for buffer subtraction and data reduction, in which the 30 frames were averaged (Rambo, 2021).

Neutron scattering data on glycosylated IgG1 (2.60-1.38 mg/ml) and the TP1 (4.78-2.32 mg/ml), TP6 (3.71-1.78 mg/ml) and TP10 (2.73-0.90 mg/ml)
samples in heavy water were obtained in two sessions (March, October 2017) on instrument SANS2D at the ISIS pulsed neutron source at the Rutherford Appleton Laboratory (Didcot, UK) (Heenan et al., 2011). This condition showed the antibody structure in a high negative solute-solvent contrast (Perkins, 1986). No conformational differences in the antibody between light and heavy water were detected in this study or previously (Wright et al., 2019). A pulsed neutron beam was derived from proton beam currents of \( \sim 40 \) μA. SANS2D data were recorded with 4 m of collimation, a 4 m sample-to-detector distance, a 12 mm sample aperture, and a wavelength range of 0.175-1.65 nm made available for time of flight. This gave a \( Q \) range from 0.05 nm\(^{-1}\) to 4 nm\(^{-1}\). The data were acquired using a two-dimensional \(^3\)He detector with 512 \( \times \) 512 pixels of 7.5 \( \times \) 7.5 mm\(^2\) in size. Samples of volume 1 ml were measured in 2 mm path length circular banjo cells for 1-7 h in a thermostated sample rack at 20\(^\circ\)C. Data were reduced using MANTID software (Arnold et al., 2014). The MANTID data reduction steps include corrections for the \( Q \) resolution, i.e. beam divergence effects and smearing from the shape and size of the slits, as well as the wavelength overlap in each pulse (Arnold et al., 2014). Using SASview software, the Guinier analyses (below) were found to be almost unaffected if the smearing was turned on or off.

Guinier analyses of the scattering data gave information of the radius of gyration \( R_G \), cross-sectional radius (\( R_{XS} \)) and molecular mass. The scattering curve \( I(Q) \) intensities at low \( Q \) are defined by the \( R_G \) value which is the averaged distance of each scattering point from the centre of scattering. In a given solute-solvent contrast, the radius of gyration \( R_G \) is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low \( Q \) gives the \( R_G \) value and the extrapolated forward scattering at zero angle \( I(0) \) (Glatter et al., 1982):

\[
\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}
\]

For antibodies, this expression is valid in a \( Q,R_G \) range up to 1.5, and was used in previous studies (Rayner et al., 2015; Wright et al., 2019), although the usual upper range reported in the literature is 1.0-1.3. If the structure is elongated, the cross-sectional radius of gyration, \( R_{XS} \), and corresponding intensity extrapolated to zero angle \( [I(Q)]_\infty \rightarrow _0 \) is obtained from (Pilz et al., 1970):
For immunoglobulins, it has been long recognised that the cross-sectional plot exhibits two regions, a steeper innermost one and a flatter outermost one (Pilz et al., 1973) and the two analyses are denoted by $R_{XS-1}$ and $R_{XS-2}$ respectively. The $R_{XS-1}$ parameter represents the averaged overall spatial separation of the Fab and Fc regions, while the $R_{XS-2}$ parameter represents the averaged spatial cross-section of the two Fab and one Fc region. The $R_G$ and $R_{XS}$ analyses were performed using SCT (Table 4.1) (Wright et al., 2015). The $Q$ ranges for the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values were 0.10–0.22, 0.29–0.52, and 0.66–1.05 nm$^{-1}$, respectively, as previously (Rayner et al., 2015; Wright et al., 2019).

Indirect transformation of the scattering data $I(Q)$ in reciprocal space into real space to give the distance distribution function $P(r)$, was carried out using GNOM (version 4.6) (Semenyuk et al., 1991; Svergun, 1992).

$$P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q)Qr \sin(Qr) dQ$$

$P(r)$ corresponds to the distribution of distances $r$ between the volume elements in the macromolecule. This yields the maximum dimension of the macromolecule $L$ and its most commonly occurring distance vector $M$ in real space. For this $P(r)$ analysis, the X-ray $I(Q)$ curve utilized up to 755 data points in the $Q$ range between 0.032 and 1.70 nm$^{-1}$ for both glycosylated and deglycosylated IgG1. The neutron $P(r)$ curve utilized up to 155 $I(Q)$ data points in the $Q$ range between 0.055 and 1.60 nm$^{-1}$ for both glycosylated and deglycosylated IgG1.

4.2.4 Atomistic modelling of IgG1

Starting structures were created for each of glycosylated and deglycosylated IgG1 A33 based on the A33 sequence provided by UCB. The latter was aligned with the sequences of IgG1 6a, IgG1 19a and IgG1 b12 (Saphire et al., 2007; Rayner et al., 2015) (Figure 4.2). This multiple sequence alignment was generated using Clustal Omega software (EMBL-EBI) (Russell, 2014). The Fab structure (Figure 4.1) was based on that found in the human IgG1 b12 crystal structure (PDB ID: 1HZH) (Saphire et al., 2007) and the Fc structure was based on that for the rituximab IgG1 antibody Fc crystal structure (PDB ID: 1HZH).
4W4N) (Kiyoshi et al., 2015) which is unchanged from that of human IgG1 but is structurally complete compared to the IgG1 b12 structure where its Fc region showed gaps. Modeller (version 9.19) (Webb et al., 2016) was used to generate the human IgG1 structure. In this, the IgG1 hinge was built using a PyMOL script build_seq (PyMOL Script Repository, Queen’s University, Ontario, Canada), based on the sequence of $^{216}$EPKSCDKTHTCPPCPAPELGGP$^{238}$. The two N-linked oligosaccharides at Asn$^{297}$ on the C$_{H2}$ domains were assumed to be complex-type biantennary oligosaccharide structures with a Man$_3$GlcNAc$_2$ core and two NeuNAc.Gal.GlcNAc antennae (Deisenhofer, 1981). The glycan template was taken from the GitHub repository (https://github.com/dww100), which was energy minimized using NAMD (Phillips et al., 2020) for 1 nanosecond to achieve a relaxed structure. This glycan was added to the Fc region by bringing the C1 atom in the first GlcNAc residue to within 0.14 nm to the N sidechain atom of Asn$^{297}$ in the C$_{H2}$ domain of IgG1, while ensuring no steric clashes with the Fc residues and the glycan chain. The PDB file was then opened on Discovery Studio (Dassault Systèmes BIOVIA, San Diego) where “CONECT” records were created for these glycosidic bonds. The CHARMM force field parameters and PSF, including those for the disulphide bridges and glycans were generated using the CHARMM-Gui GlycanReader tool (Miller et al., 2008; Jo et al., 2011; Lee et al., 2016) in order to be compatible with the CHARMM36 forcefield (MacKerell et al., 1998; Mackerell et al., 2004; Guvench et al., 2009; Best et al., 2012). To relax this structure, the full IgG1 structure with and without glycans were then energy minimised for 10,000 steps using the simulation engine NAMD (version 2.9) with the CHARMM36 forcefield.

For the Monte Carlo simulations to generate trial structures, the starting IgG1 structure was renumbered and its naming nomenclature was adjusted to match the required format for the Torsion Angle Monte Carlo (TAMC) module on SASSIE-web (Zhang et al., 2017). For TAMC to work, the IgG1 residue numbering was changed to be continuous for two segments, one segment corresponding to the first Fab region, its hinge and the Fc region, and the other segment to only the second Fab region and the hinge connected to this. A library of physically realistic glycosylated and deglycosylated structural conformations was generated by subjecting the starting structures to the TAMC module in SASSIE-web (Zhang et al., 2017). The flexible regions were assigned within the
hinge, namely $^{216}$EPK$^{218}$ and $^{231}$APE$^{233}$ on one side of IgG1, and just $^{216}$EPK$^{218}$ on the other side (Figures 1B, 2E). These tripeptides corresponded to surface-accessible structures outside the structurally-defined Fab and Fc regions and the disulphide-linked hinge core. These tripeptides could be structurally varied to create the required IgG1 conformers for testing against the scattering curve. The rest of the IgG1 structure was held rigid. Making $^{216}$EPK$^{218}$ flexible on both sides of IgG1 rendered both Fab regions to be conformationally mobile, and making $^{231}$APE$^{233}$ flexible made the Fc region mobile. For each of these nine linker residues, the backbone phi ($\phi$) and psi ($\psi$) torsion angles were varied in steps of up to either 30° or 180°, the degree to which the torsion angle were varied was decided based on which angle yielded the highest model acceptance rate, therefore minimising computational time. In the Monte Carlo simulation, many attempted moves will be physically unrealistic and were therefore discarded. For the glycosylated IgG1 simulations, 2,500,000 moves were attempted of which 123,284 were accepted. For the deglycosylated simulations, in which the glycan chains were omitted, 2,600,000 moves were attempted, of which 119,191 models were accepted.

For each of the 123,284 and 119,191 models, a scattering curve was generated using the SasCalc module in SASSIE-web. SasCalc calculates the scattering curve $I(Q)$ using an all-atom expression for the scattering intensity in which the orientations of the $Q$ vectors are taken from a quasi-uniform spherical grid generated by the golden ratio (Watson et al., 2013). For X-ray modelling, consideration of the hydration shell would require the explicit addition of a monolayer of water molecules to the protein surface before calculating $I(Q)$, and would require much computational effort (Watson et al., 2013). Thus, the hydration shell was not considered here for X-rays, and was not required for neutrons. These scattering curves were compared to the X-ray and neutron experimental scattering curves, using the R-factor function in SASSIE-web. This function calculates the difference between the modelled curve $I_{\text{Model}}(Q_i)$ and the experimental curves $I_{\text{Expt}}(Q_i)$, this function being analogous to that used in protein crystallography:

$$R = \frac{\sum \left| I_{\text{Expt}}(Q_i) - \eta I_{\text{Model}}(Q_i) \right|}{\sum |I_{\text{Expt}}(Q_i)|} \times 100$$
where $Q_i$ is the $Q$ value of the $i^{th}$ data point, $l_{\text{Expt}}(Q_i)$ is the experimental scattering intensity and $l_{\text{Model}}(Q_i)$ is the theoretical modelled scattering intensity, and $\eta$ is a scaling factor used to match the theoretical curve to the experimental $I(Q)$ \cite{wright2015}. Lower R-factor values represent better fits. An iterative search to minimize the R factor was used to determine $\eta$ \cite{wright2015}. In the extrapolated experimental scattering curves, the lowest $Q$ values in the range before the fitted Guinier $R_G$ region were interpolated to zero $Q$ using MATLAB in order to satisfy the input requirement for the SasCalc module in SASSIE-web. Interpolation makes the $Q$ spacing uniform between the data points, and extrapolation extends the full $I(Q)$ curve to zero $Q$. The resulting 680 and 72 $I(Q)$ values in the $Q$ range of 0.0-1.5 nm$^{-1}$ were utilised for the X-ray and neutron curve fits respectively, and defined the $Q$ spacing for SasCalc and the R-factor values. The use of $\chi^2$ analyses to evaluate the fits was not possible because this requires the experimental data points to have errors associated with them, which were not available when interpolating the curve. For the neutron curve fits, no correction was required for a flat incoherent background because the IgG1 concentrations were relatively low and dialyses had sufficiently reduced the proton content in the buffers. The 123,284 glycosylated and 119,191 deglycosylated models gave an $R$-factor vs. $R_G$ distribution that encompassed the experimental extrapolated $R_G$ value. This $R$-factor analysis was repeated for four experimental X-ray scattering curves at different concentrations for each of glycosylated and deglycosylated IgG1 (Table 4.2). The same analysis was repeated for two neutron scattering curves at different concentrations, for each of glycosylated and deglycosylated IgG1 (Table 4.4). For each concentration, the best-fit 100 models with the smallest $R$-factors were accepted.

Principal component analysis (PCA) provided by the Bio3d package in R \cite{grant2006} was used to identify the main classes of best-fit IgG1 conformations found in the 800 best-fit glycosylated and deglycosylated models from X-ray scattering (Table 4.3). A separate analysis of the 400 best-fit models from neutron scattering was performed. To remove any bias in the PCA clustering of coordinate sets caused by the presence or absence of the glycans, the glycan coordinates were removed from the best-fit glycosylated models prior to generating the PCA. The X-ray and neutron models were assessed through two
Table 4.2 Modelling fits for the X-ray scattering and analytical ultracentrifugation data in light water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Model</th>
<th>$R_g$ before minimization (nm)</th>
<th>$R_g$ after minimization (nm)</th>
<th>$R_{XS-1}$ (nm)</th>
<th>$R_{XS-2}$ (nm)</th>
<th>$L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library of glycosylated models</td>
<td>123284</td>
<td>4.55-5.59</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.70-8.43</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 3.60 mg/ml</td>
<td>100</td>
<td>5.09-5.21</td>
<td>5.09-5.21</td>
<td>2.48-2.63</td>
<td>1.31-1.61</td>
<td>NA</td>
<td>0.70-1.13</td>
<td>0.71-1.21</td>
<td>6.57-6.73</td>
</tr>
<tr>
<td>Best fit at 3.60 mg/ml</td>
<td>1</td>
<td>5.14</td>
<td>5.14</td>
<td>2.53</td>
<td>1.50</td>
<td>17</td>
<td>0.70</td>
<td>0.71</td>
<td>6.64</td>
</tr>
<tr>
<td>Top 100 at 3.19 mg/ml</td>
<td>100</td>
<td>5.08-5.19</td>
<td>5.07-5.19</td>
<td>2.48-2.63</td>
<td>1.24-1.55</td>
<td>NA</td>
<td>0.75-1.18</td>
<td>0.75-1.23</td>
<td>6.59-6.73</td>
</tr>
<tr>
<td>Best fit at 3.19 mg/ml</td>
<td>1</td>
<td>5.14</td>
<td>5.14</td>
<td>2.53</td>
<td>1.49</td>
<td>17</td>
<td>0.75</td>
<td>0.75</td>
<td>6.70</td>
</tr>
<tr>
<td>Top 100 at 2.05 mg/ml</td>
<td>100</td>
<td>5.02-5.18</td>
<td>5.02-5.17</td>
<td>2.44-2.63</td>
<td>1.20-1.54</td>
<td>NA</td>
<td>1.05-1.54</td>
<td>1.04-1.55</td>
<td>6.59-6.73</td>
</tr>
<tr>
<td>Best fit at 2.05 mg/ml</td>
<td>1</td>
<td>5.15</td>
<td>5.15</td>
<td>2.48</td>
<td>1.37</td>
<td>17</td>
<td>1.05</td>
<td>1.04</td>
<td>6.71</td>
</tr>
<tr>
<td>Top 100 at 1.36 mg/ml</td>
<td>100</td>
<td>5.00-5.18</td>
<td>5.00-5.18</td>
<td>2.32-2.67</td>
<td>1.20-1.54</td>
<td>NA</td>
<td>0.94-2.18</td>
<td>0.96-2.29</td>
<td>6.62-6.77</td>
</tr>
<tr>
<td>Best fit at 1.36 mg/ml</td>
<td>1</td>
<td>5.08</td>
<td>5.08</td>
<td>2.48</td>
<td>1.32</td>
<td>17</td>
<td>0.94</td>
<td>0.96</td>
<td>6.68</td>
</tr>
<tr>
<td>Library of deglycosylated models</td>
<td>119191</td>
<td>4.63-5.67</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.87-8.60</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 4.29 mg/ml</td>
<td>100</td>
<td>5.10-5.28</td>
<td>5.10-5.28</td>
<td>2.49-2.68</td>
<td>1.30-1.62</td>
<td>NA</td>
<td>0.80-1.67</td>
<td>0.81-1.67</td>
<td>6.24-6.45</td>
</tr>
<tr>
<td>Best fit at 4.29 mg/ml</td>
<td>1</td>
<td>5.21</td>
<td>5.21</td>
<td>2.54</td>
<td>1.52</td>
<td>17</td>
<td>0.80</td>
<td>0.81</td>
<td>6.27</td>
</tr>
<tr>
<td>Top 100 at 3.19 mg/ml</td>
<td>100</td>
<td>5.05-5.27</td>
<td>5.05-5.27</td>
<td>2.49-2.69</td>
<td>1.26-1.62</td>
<td>NA</td>
<td>0.87-1.65</td>
<td>0.85-1.65</td>
<td>6.28-6.45</td>
</tr>
<tr>
<td>Best fit at 3.19 mg/ml</td>
<td>1</td>
<td>5.17</td>
<td>5.17</td>
<td>2.59</td>
<td>1.38</td>
<td>17</td>
<td>0.87</td>
<td>0.85</td>
<td>6.36</td>
</tr>
<tr>
<td>Top 100 at 1.34 mg/ml</td>
<td>100</td>
<td>4.98-5.25</td>
<td>4.98-5.25</td>
<td>2.42-2.71</td>
<td>1.25-1.56</td>
<td>NA</td>
<td>0.93-1.77</td>
<td>0.94-1.79</td>
<td>6.28-6.50</td>
</tr>
<tr>
<td>Best fit at 1.34 mg/ml</td>
<td>1</td>
<td>5.11</td>
<td>5.10</td>
<td>2.57</td>
<td>1.31</td>
<td>17</td>
<td>0.93</td>
<td>0.94</td>
<td>6.45</td>
</tr>
<tr>
<td>Top 100 at 1.02 mg/ml</td>
<td>100</td>
<td>4.94-5.19</td>
<td>4.93-5.19</td>
<td>2.42-2.73</td>
<td>1.25-1.56</td>
<td>NA</td>
<td>1.06-1.93</td>
<td>1.03-1.95</td>
<td>6.28-6.50</td>
</tr>
<tr>
<td>Best fit at 1.02 mg/ml</td>
<td>1</td>
<td>5.11</td>
<td>5.11</td>
<td>2.57</td>
<td>1.31</td>
<td>17</td>
<td>1.06</td>
<td>1.03</td>
<td>6.45</td>
</tr>
</tbody>
</table>
Table 4.3 Summary of Principal component analysis for fits of X-ray scattering and analytical ultracentrifugation data in light water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Model</th>
<th>$R_G$ before minimization (nm)</th>
<th>$R_G$ after minimization (nm)</th>
<th>$R_{KS-1}$ (nm)</th>
<th>$R_{KS-2}$ (nm)</th>
<th>$L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA Group 1</td>
<td>217</td>
<td>5.00-5.28</td>
<td>5.00-5.28</td>
<td>2.32-2.63</td>
<td>1.25-1.61</td>
<td>NA</td>
<td>0.87-2.17</td>
<td>0.85-2.29</td>
<td>6.24-6.75</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>175</td>
<td>5.00-5.18</td>
<td>5.00-5.18</td>
<td>2.32-2.63</td>
<td>1.25-1.61</td>
<td>NA</td>
<td>0.91-2.17</td>
<td>0.89-2.29</td>
<td>6.61-6.75</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>42</td>
<td>5.08-5.28</td>
<td>5.08-5.28</td>
<td>2.50-2.63</td>
<td>1.30-1.61</td>
<td>NA</td>
<td>0.87-1.84</td>
<td>0.85-1.82</td>
<td>6.24-6.33</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.07</td>
<td>5.06</td>
<td>2.32</td>
<td>1.43</td>
<td>17</td>
<td>2.15</td>
<td>2.29</td>
<td>6.74</td>
</tr>
<tr>
<td>PCA Group 2</td>
<td>50</td>
<td>5.06-5.28</td>
<td>5.06-5.28</td>
<td>2.42-2.58</td>
<td>1.50-1.62</td>
<td>NA</td>
<td>0.94-2.18</td>
<td>0.96-2.21</td>
<td>6.25-6.72</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>46</td>
<td>5.06-5.18</td>
<td>5.06-5.18</td>
<td>2.42-2.58</td>
<td>1.20-1.54</td>
<td>NA</td>
<td>0.94-2.18</td>
<td>0.96-2.21</td>
<td>6.62-6.72</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>1</td>
<td>5.12-5.28</td>
<td>5.12-5.28</td>
<td>2.50-2.56</td>
<td>1.34-1.62</td>
<td>NA</td>
<td>1.27-1.62</td>
<td>1.24-1.67</td>
<td>6.25-6.38</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.10</td>
<td>5.10</td>
<td>2.51</td>
<td>1.20</td>
<td>17</td>
<td>1.48</td>
<td>1.48</td>
<td>6.64</td>
</tr>
<tr>
<td>PCA Group 3</td>
<td>234</td>
<td>5.02-5.28</td>
<td>5.02-5.28</td>
<td>2.46-2.61</td>
<td>1.20-1.59</td>
<td>NA</td>
<td>0.70-2.18</td>
<td>0.71-2.12</td>
<td>6.24-6.77</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>176</td>
<td>5.02-5.21</td>
<td>5.02-5.21</td>
<td>2.46-2.59</td>
<td>1.20-1.59</td>
<td>NA</td>
<td>0.70-2.18</td>
<td>0.71-2.12</td>
<td>6.57-6.77</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>58</td>
<td>5.09-5.28</td>
<td>5.09-5.28</td>
<td>2.52-2.61</td>
<td>1.30-1.58</td>
<td>NA</td>
<td>0.87-1.92</td>
<td>0.87-1.91</td>
<td>6.24-6.38</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.19</td>
<td>5.19</td>
<td>2.50</td>
<td>1.53</td>
<td>17</td>
<td>1.10</td>
<td>1.09</td>
<td>6.65</td>
</tr>
<tr>
<td>PCA Group 4</td>
<td>153</td>
<td>5.11-5.26</td>
<td>5.10-5.26</td>
<td>2.42-2.61</td>
<td>1.33-1.56</td>
<td>NA</td>
<td>0.80-2.11</td>
<td>0.81-2.10</td>
<td>6.29-6.71</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>2</td>
<td>5.15</td>
<td>5.15</td>
<td>2.53</td>
<td>1.38</td>
<td>NA</td>
<td>2.11</td>
<td>2.10</td>
<td>6.71</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>151</td>
<td>5.11-5.26</td>
<td>5.10-5.26</td>
<td>2.42-2.61</td>
<td>1.33-1.56</td>
<td>NA</td>
<td>0.80-1.93</td>
<td>0.81-1.93</td>
<td>6.29-6.50</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.21</td>
<td>5.21</td>
<td>2.55</td>
<td>1.52</td>
<td>17</td>
<td>1.67</td>
<td>1.67</td>
<td>6.40</td>
</tr>
<tr>
<td>PCA Group 5</td>
<td>146</td>
<td>4.94-5.22</td>
<td>4.93-5.22</td>
<td>2.58-2.73</td>
<td>1.25-1.61</td>
<td>NA</td>
<td>0.92-1.93</td>
<td>0.93-2.18</td>
<td>6.28-6.67</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>1</td>
<td>5.02</td>
<td>5.02</td>
<td>2.67</td>
<td>1.31</td>
<td>NA</td>
<td>2.17</td>
<td>2.18</td>
<td>6.67</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>145</td>
<td>4.94-5.22</td>
<td>4.93-5.22</td>
<td>2.58-2.73</td>
<td>1.25-1.61</td>
<td>NA</td>
<td>0.92-1.93</td>
<td>0.93-1.95</td>
<td>6.28-6.46</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.16</td>
<td>5.16</td>
<td>2.61</td>
<td>1.59</td>
<td>17</td>
<td>1.28</td>
<td>1.29</td>
<td>6.45</td>
</tr>
</tbody>
</table>
Table 4.4 Modelling fits for the neutron scattering and analytical ultracentrifugation data in heavy water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Model</th>
<th>$R_G$ before minimization (nm)</th>
<th>$R_G$ after minimization (nm)</th>
<th>$R_{x\beta-1}$ (nm)</th>
<th>$R_{x\beta-2}$ (nm)</th>
<th>$L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library of glycosylated models</td>
<td>123284</td>
<td>4.51-5.55</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.68-11.46</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 2.60 mg/ml</td>
<td>100</td>
<td>5.16-5.33</td>
<td>5.15-5.32</td>
<td>2.32-2.53</td>
<td>1.03-1.32</td>
<td>NA</td>
<td>1.68-2.19</td>
<td>1.74-2.32</td>
<td>6.53-6.71</td>
</tr>
<tr>
<td>Best fit at 2.60 mg/ml</td>
<td>1</td>
<td>5.24</td>
<td>5.23</td>
<td>2.47</td>
<td>1.19</td>
<td>16</td>
<td>1.68</td>
<td>1.74</td>
<td>6.64</td>
</tr>
<tr>
<td>Top 100 at 1.38 mg/ml</td>
<td>100</td>
<td>5.19-5.35</td>
<td>5.17-5.34</td>
<td>2.32-2.49</td>
<td>1.10-1.25</td>
<td>NA</td>
<td>1.98-2.38</td>
<td>1.99-2.43</td>
<td>6.57-6.71</td>
</tr>
<tr>
<td>Best fit at 1.38 mg/ml</td>
<td>1</td>
<td>5.24</td>
<td>5.23</td>
<td>2.36</td>
<td>1.22</td>
<td>16</td>
<td>1.98</td>
<td>2.04</td>
<td>6.70</td>
</tr>
<tr>
<td>Library of deglycosylated models</td>
<td>119191</td>
<td>4.57-5.61</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2.22-12.04</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 2.73 mg/ml</td>
<td>100</td>
<td>5.20-5.34</td>
<td>5.20-5.34</td>
<td>2.46-2.57</td>
<td>1.11-1.27</td>
<td>NA</td>
<td>2.28-2.54</td>
<td>2.24-2.50</td>
<td>6.26-6.35</td>
</tr>
<tr>
<td>Best fit at 2.73 mg/ml</td>
<td>1</td>
<td>5.31</td>
<td>5.31</td>
<td>2.47</td>
<td>1.19</td>
<td>16</td>
<td>2.28</td>
<td>2.24</td>
<td>6.28</td>
</tr>
<tr>
<td>Top 100 at 0.90 mg/ml</td>
<td>100</td>
<td>5.08-5.22</td>
<td>5.07-5.21</td>
<td>2.42-2.57</td>
<td>1.15-1.39</td>
<td>NA</td>
<td>2.23-2.66</td>
<td>2.22-2.62</td>
<td>6.29-6.44</td>
</tr>
<tr>
<td>Best fit at 0.90 mg/ml</td>
<td>1</td>
<td>5.15</td>
<td>5.14</td>
<td>2.50</td>
<td>1.26</td>
<td>16</td>
<td>2.22</td>
<td>2.22</td>
<td>6.32</td>
</tr>
</tbody>
</table>
Table 4.5 Summary of Principal component analysis for fits of neutron scattering and analytical ultracentrifugation data in heavy water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Model</th>
<th>$R_g$ before minimization (nm)</th>
<th>$R_g$ after minimization (nm)</th>
<th>$R_{xx-1}$ (nm)</th>
<th>$R_{xx-2}$ (nm)</th>
<th>$L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA Group 1</td>
<td>185</td>
<td>5.14-5.33</td>
<td>5.13-5.32</td>
<td>2.32-2.54</td>
<td>1.03-1.35</td>
<td>NA</td>
<td>1.87-2.63</td>
<td>1.95-2.60</td>
<td>6.28-6.71</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>167</td>
<td>5.16-5.33</td>
<td>5.15-5.32</td>
<td>2.32-2.53</td>
<td>1.03-1.27</td>
<td>NA</td>
<td>1.87-2.19</td>
<td>1.95-2.32</td>
<td>6.60-6.71</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>18</td>
<td>5.14-5.29</td>
<td>5.13-5.29</td>
<td>2.48-2.54</td>
<td>1.16-1.35</td>
<td>NA</td>
<td>2.23-2.63</td>
<td>2.22-2.60</td>
<td>6.28-6.34</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.14</td>
<td>5.13</td>
<td>2.51</td>
<td>1.34</td>
<td>16</td>
<td>2.52</td>
<td>2.51</td>
<td>6.32</td>
</tr>
<tr>
<td>PCA Group 2</td>
<td>18</td>
<td>5.09-5.28</td>
<td>5.08-5.27</td>
<td>3.37-2.57</td>
<td>1.16-1.39</td>
<td>NA</td>
<td>1.84-2.65</td>
<td>1.87-2.62</td>
<td>6.24-6.66</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>9</td>
<td>5.22-5.28</td>
<td>5.21-5.27</td>
<td>2.37-2.50</td>
<td>1.18-1.29</td>
<td>NA</td>
<td>1.84-2.09</td>
<td>1.87-2.13</td>
<td>6.53-6.66</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>9</td>
<td>5.09-5.21</td>
<td>5.08-5.20</td>
<td>2.42-2.57</td>
<td>1.16-1.39</td>
<td>NA</td>
<td>2.26-2.65</td>
<td>2.25-2.62</td>
<td>6.32-6.44</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.15</td>
<td>5.14</td>
<td>2.5</td>
<td>1.35</td>
<td>16</td>
<td>2.38</td>
<td>2.38</td>
<td>6.36</td>
</tr>
<tr>
<td>PCA Group 3</td>
<td>16</td>
<td>5.08-5.25</td>
<td>5.07-5.24</td>
<td>2.45-2.50</td>
<td>1.20-1.26</td>
<td>NA</td>
<td>1.86-2.48</td>
<td>2.00-2.56</td>
<td>6.25-6.62</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>15</td>
<td>5.18-5.25</td>
<td>5.17-5.24</td>
<td>2.45-2.50</td>
<td>1.20-1.26</td>
<td>NA</td>
<td>1.86-2.14</td>
<td>2.00-2.28</td>
<td>6.54-6.62</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>1</td>
<td>5.08</td>
<td>5.07</td>
<td>2.46</td>
<td>1.26</td>
<td>NA</td>
<td>2.48</td>
<td>2.56</td>
<td>6.38</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.25</td>
<td>5.24</td>
<td>2.49</td>
<td>1.21</td>
<td>16</td>
<td>2.13</td>
<td>2.16</td>
<td>6.32</td>
</tr>
<tr>
<td>PCA Group 4</td>
<td>143</td>
<td>5.12-5.35</td>
<td>5.11-5.34</td>
<td>2.41-2.53</td>
<td>1.11-1.39</td>
<td>NA</td>
<td>1.68-2.66</td>
<td>1.74-2.61</td>
<td>6.26-6.67</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>9</td>
<td>5.19-5.35</td>
<td>5.18-5.30</td>
<td>2.41-2.50</td>
<td>1.19-1.32</td>
<td>NA</td>
<td>1.68-2.14</td>
<td>1.74-2.24</td>
<td>6.63-6.67</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>134</td>
<td>5.12-5.34</td>
<td>5.11-5.34</td>
<td>2.42-2.53</td>
<td>1.11-1.39</td>
<td>NA</td>
<td>2.28-2.66</td>
<td>2.23-2.61</td>
<td>6.26-6.37</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.22</td>
<td>5.22</td>
<td>2.47</td>
<td>1.23</td>
<td>16</td>
<td>2.43</td>
<td>2.4</td>
<td>6.54</td>
</tr>
<tr>
<td>PCA Group 5</td>
<td>1</td>
<td>5.1</td>
<td>5.09</td>
<td>2.54</td>
<td>1.24</td>
<td>NA</td>
<td>2.61</td>
<td>2.61</td>
<td>6.29</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>1</td>
<td>5.1</td>
<td>5.09</td>
<td>2.54</td>
<td>1.24</td>
<td>NA</td>
<td>2.61</td>
<td>2.61</td>
<td>6.29</td>
</tr>
<tr>
<td>Centroid</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
separate PCA. The average structure for each PCA group was identified using a centroid model computed using R.

In order to model AUC parameters the theoretical $s_{20,w}$ values were generated for the best-fit 800 and 400 glycosylated and deglycosylated IgG1 models using HullRad Version 7 (Fleming et al., 2018). HullRad includes glycan residues when relevant, however there are inconsistencies in the Protein Database and CHARMM-GUI nomenclature for glycans. The nomenclature in the Hullrad script was thus modified to ensure that the IgG1 glycosylation was correctly incorporated in the $s_{20,w}$ calculation.

4.3 Results

4.3.1 Purification and characterisation of glycosylated and deglycosylated IgG1

A protocol for the deglycosylation of the monoclonal human IgG1 A33 antibody was set up using peptide:N-glycosidase F (PNGase F) digests according to the manufacturer’s protocol (section 4.3.1). The completeness of deglycosylation was verified by a combination of routine gel filtration, SDS-PAGE and mass spectrometry, and also by analytical ultracentrifugation (see below):

(i) At the timepoints of one hour, six hours and ten hours after the start of the digests (TP1, TP6 and TP10), the elution of the IgG1 digested products from a gel filtration column preceded that for native glycosylated IgG1 (Figure 4.3A). Both glycosylated and deglycosylated IgG1 eluted as a main symmetrical peak at 17.88 ml, 17.84 ml, 17.82 ml and 17.86 ml for glycosylated IgG1, and deglycosylated IgG1 at timepoints of TP1, TP6 and TP10, respectively (Figure 4.3A). This process ensured that the IgG1 sample was monodisperse with no aggregates present immediately before analytical ultracentrifugation or scattering experiments.

(ii) When the IgG1 samples were submitted to non-reducing and reducing SDS-PAGE analyses at equimolar concentrations, purified glycosylated and deglycosylated IgG1 showed a single band between 200 and 116kDa on 4-12% Bis Tris NuPage gel under non-reducing conditions, which is consistent with the expected masses of ~147 kDa for IgG1 (Figure 4.3B). Under reducing conditions two bands were present corresponding to the heavy chain (with an apparent
mass of ~50 kDa) and the light chain (with an apparent molecular mass of ~25 kDa) (Figure 4.3B). These apparent molecular masses were as expected from the known sequence.

(iii) Liquid chromatography mass spectrometry measurements showed multiple peaks for glycosylated IgG1 (G) that were assigned to the presence of at least four glycoforms, separated by masses of 160-230 Da that corresponded to single sugar residues (Figure 4.3C) (Hui et al., 2019). The most intense IgG1 glycosylated population had an observed deconvoluted mass of 147,010 Da. After an hour of digest, a partially deglycosylated IgG1 (P) was observed in which the number of glycoforms was diminished, and additional peaks were observed at ~145.4 kDa and 143,958 Da. After six or ten hours, only the single dominant deglycosylated peak (D) was seen at 143,958 Da. The peak at 145.4 kDa was attributed to IgG1 in which one of the two glycans at Asn297 was not present. The mass of each glycan chain was calculated by subtracting the glycosylated and deglycosylated masses and halving the outcome to give 1,526 Da. This glycan mass was also calculated by subtracting the glycosylated and half-deglycosylated masses to give 1613 Da. These values agree well with an assumed glycan composition of Gal$_2$Man$_3$GlcNAc$_4$ that gives a mass of 1622 Da.
Figure 4.3 Purification, SDS-PAGE and mass spectrometry of human glycosylated and deglycosylated IgG1.

A Elution peaks from a Superose 6 Increase 10/300 gel filtration column for four IgG1 samples, these being glycosylated (black) and from the TP1 (blue), TP6 (red) and TP10 (magenta) timepoints. The dashed vertical lines indicate the peak positions.

B Lane 1 and 6, molecular mass markers are denoted in kDa. Lanes 2-5, non-reduced SDS-PAGE of glycosylated IgG1, TP1, TP6 and TP10 after gel filtration. 7-10, reduced SDS-PAGE of glycosylated IgG1, TP1, TP6 and TP10 after gel filtration.

C Mass spectra of glycosylated and deglycosylated IgG1, using the same colour scheme as in A. Peaks labelled G represents glycosylated species, P represents partially glycosylated species, and D represents fully deglycosylated species.
4.3.2 Analytical ultracentrifugation of glycosylated and deglycosylated IgG1

Sedimentation velocity experiments investigated the masses and solution structures of glycosylated and deglycosylated IgG1 at the TP1, TP6 and TP10 timepoints. The SEDFIT analyses of the boundaries involved fits of as many as 896 scans, and the good agreement between the experimental boundary scans and fitted lines is clear (left, Figure 4.4A). In the resulting size distribution analyses \( c(s) \), a monomer peak that monitored the overall IgG1 solution structure was observed at average \( s_{20,w} \) values of 6.25 S for glycosylated IgG1, 6.12 S for IgG1 TP1, 6.15 S for IgG1 TP6, and 6.12 S for IgG1 TP10 (right, Figure 4.4A). For the glycosylated forms, these values agreed well with those of 6.42 S and 6.34 S for IgG1 6a and 19a respectively (Rayner et al., 2015), and with earlier studies (Kilar et al., 1985; Gregory et al., 1987; Phillips et al., 1994). From the \( c(s) \) analyses, the molecular masses of the IgG1 monomer peak were 151 kDa (glycosylated), 147 kDa (TP1), 156 kDa (TP6) and 148 kDa (TP10). These values agree well with the composition-calculated masses of 148.4 kDa and 144.4 kDa for the glycosylated and deglycosylated IgG1 monomers respectively. These also agree well with the values from mass spectrometry of 147,010 Da and 143,958 Da for glycosylated and deglycosylated IgG1 (Figure 4.3C). The 0.2 S reduction (3%) in the \( s_{20,w} \) values on deglycosylation is attributable to the 4 kDa reduction (3%) in the IgG1 mass, according to the Svedberg equation in which \( s_{20,w} \) is proportional to the mass divided by the frictional coefficient. This calculation assumes that the IgG1 conformation (i.e. the frictional coefficient) is unchanged after deglycosylation.

A slight concentration dependence was observed for the monomer \( s_{20,w} \) values for glycosylated and deglycosylated IgG1 (Figure 4.4B), which increased with decreased concentration. In the 2015 study, minor peaks for IgG1 dimers were visible at about 9 S for IgG1 6a and 19a (Rayner et al., 2015). In this work, no dimer peaks were visible for IgG1 A33 (right panel, Figure 4.4A). This difference may result from the use of histidine buffer in this current study, in distinction to the phosphate buffer saline used before. If IgG1 A33 forms dimers, the histidine buffer may have increased the exchange rate between monomer...
A The experimentally observed sedimentation boundaries for a concentration series of glycosylated IgG1 and likewise for deglycosylated IgG1 at the TP1 (blue), TP6 (red) and TP10 (magenta) timepoints in histidine buffer. Scans were recorded at 30,000 rpm and 20°C, from which 34-46 boundaries (black outlines) are shown from totals of up to 896 scans. The SEDFIT fits are shown as white lines. The peaks in the corresponding size distribution analyses c(s) revealed a monomer peak (M) at $s_{20,w}^0$ values of 6.18-6.45 S for glycosylated IgG1 and the three deglycosylation timepoints TP1, TP6 and TP10. B The $s_{20,w}^0$ values for the monomer peaks are shown as a function of concentration for glycosylated IgG1 (●), and the TP1 (●), TP6 (●) and TP10 (●) timepoints.

Figure 4.4 Sedimentation velocity analyses of glycosylated and deglycosylated IgG1.
and dimer such that separate monomer and dimer peaks were no longer seen. Interestingly, the peak width for glycosylated IgG1 is greater than that of deglycosylated IgG1 (Figure 4.4A). The increased width may indicate a mix of monomer and dimer in fast exchange in glycosylated IgG1, which is reduced to monomer only upon deglycosylation.

### 4.3.3 X-ray and neutron scattering of glycosylated and deglycosylated IgG1

The overall solution structures of glycosylated and the three deglycosylated IgG1 samples at the TP1, TP6 and TP10 time points were analysed by X-ray and neutron scattering. The two methods provided slightly different perspectives of the same solution structure. X-rays in light water buffers detect the hydration shell surrounding the protein structure, whereas the effect of the hydration shell is invisible in heavy water buffers for reason of the different solute-solvent contrast in use (Perkins, 1986, 2001; Svergun et al., 1998). The IgG1 X-ray data collection at concentrations between 0.5-5.5 mg/ml used time frame analyses to ensure the absence of radiation damage effects. The resulting $R_G$ and $R_{XS-1}/R_{XS-2}$ values monitor the elongation of the overall IgG1 structure and its approximate cross-sectional structures respectively.

Guinier analyses resulted in high quality linear plots for all four samples and revealed three distinctive regions of the $I(Q)$ curves, as expected for antibodies (Rayner et al., 2014, 2015; Hui et al., 2015). From these, the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values from the individual scattering curves were obtained within satisfactory $Q.R_G$ and $Q.R_{XS}$ limits of 0.5-1.4, 0.7-1.3 and 0.9-1.5 respectively (Figure 4.5A). For both glycosylated and deglycosylated IgG1 a slight concentration dependence was observed in the X-ray $I(0)/c$ values that suggested a small amount of oligomer formation in the concentration series (Figure 4.6A; Table 4.1). This agreed with the AUC data (Figure 4.4B). After extrapolation to zero concentration, the X-ray $R_G$ values that monitor the overall structure for glycosylated IgG1, and the deglycosylated TP1, TP6 and TP10 IgG1 samples were almost unchanged at 5.10 ± 0.13 nm, 5.10 ± 0.20 nm, 5.11 ± 0.13 nm, and 5.13 ± 0.13 nm respectively. These X-ray $R_G$ values for glycosylated
Figure 4.5 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG1 (legend overleaf).
Figure 4.5. X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG1 (continued). A The SAXS curves for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG1 at concentrations of 5.38-0.74 mg/ml. The filled circles between the arrows represent the $Q.R_G$ and $Q.R_{XS}$ fit ranges used to determine the $R_G$ and $R_{XS}$ values. The $Q$ range used for the $R_G$ values was 0.01-0.027 nm$^{-1}$; those for the $R_{XS-1}$ and $R_{XS-2}$ values were 0.029-0.052 nm$^{-1}$ and 0.066-0.105 nm$^{-1}$ respectively. B The SANS curves for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG1 at concentrations of 4.78-0.90 mg/ml. The $Q$ range used for the $R_G$ values was 0.007-0.027 nm$^{-1}$ and those for the $R_{XS-1}$ and $R_{XS-2}$ values were 0.028-0.052 nm$^{-1}$ and 0.066-0.105 nm$^{-1}$ respectively.
IgG1 A33 here agree well with previous $R_G$ values of 5.28 – 5.32 nm for two other human monoclonal IgG1 6a and 19a antibodies (Rayner et al., 2015). The $R_{XS-1}$ values from the individual curves (Figure 4.6A) are an approximate monitor of the cross-sectional structure for glycosylated IgG1 and deglycosylated IgG1. These were extrapolated to zero concentration to show that these were slightly increased from 2.47 ± 0.01 nm to 2.51 ± 0.01 nm respectively, showing some rearrangement between the Fab and Fc regions. The $R_{XS-2}$ values from the individual curves for glycosylated IgG1 and deglycosylated IgG1 at the TP1, TP6 and TP10 timepoints were extrapolated to zero concentration to show that these were unchanged at 1.40 ± 0.07 nm, 1.41 ± 0.05 nm, 1.41 ± 0.05 nm and 1.42 ± 0.04 nm. This showed that the individual Fab and Fc regions were unchanged in structure. In summary, the small increase of 0.04 nm in the extrapolated $R_{XS-1}$ values with increasing deglycosylation suggested that there were small increases of cross-sectional elongation in the IgG1 structure upon removal of the glycan chains.

The corresponding neutron scattering data sets for glycosylated and deglycosylated IgG1 (TP1, TP6, TP10) in 100% $^2$H$_2$O buffer were analysed at concentrations of 0.71–2.73 mg/ml, this concentration range being similar to that used above for SAXS. Again, the Guinier analyses revealed high quality linear fits for the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ parameters (Figure 4.5B). A concentration dependence was not observed for IgG1, this being seen from the $I(0)/c$ values which remained unchanged within error (Figure 4.6B). This difference between the neutron and X-ray data sets is attributable to the fewer data points obtained with neutrons, leading to reduced precision in the data sets. The mean neutron $R_G$ values for glycosylated IgG1 and deglycosylated IgG1 (TP1, TP6 and TP10) were unchanged at 5.27 ± 0.01 nm, 5.28 ± 0.06 nm, 5.28 ± 0.01 nm and 5.25 ± 0.06 nm respectively (Figure 4.6B). The neutron $R_G$ values for glycosylated IgG1 A33 agree well with the $R_G$ values of 5.16–5.18 nm for human IgG1 6a and 19a (Rayner et al., 2015). The mean neutron $R_{XS-1}$ values for glycosylated IgG1 and deglycosylated IgG1 (TP1, TP6 and TP10) were 2.35 ± 0.01 nm, 2.43 ± 0.01 nm, 2.42 ± 0.01 nm and 2.38 ± 0.01 nm respectively, suggesting a small increase in the cross-sectional structure following deglycosylation. The mean neutron $R_{XS-2}$ values for glycosylated IgG1 and deglycosylated IgG1 (TP1, TP6 and TP10) were
unchanged at $1.14 \pm 0.04$ nm, $1.19 \pm 0.01$ nm, $1.17 \pm 0.01$ nm and $1.15 \pm 0.03$ nm respectively. The neutron values confirmed the X-ray analyses.
Figure 4.6 Concentration dependence of the SAXS and SANS Guinier analyses.

The filled symbols show the values determined from the Guinier analyses and the open symbols in the $R_G$ panels indicate those determined from the $P(r)$ analyses. The colours denote the glycosylated IgG1 (black), TP1 (blue), TP6 (red) and TP10 (magenta) timepoints.

A The SAXS $R_G$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for glycosylated (●, ○) and deglycosylated TP1 (●, ○), TP6 (●, ○) and TP10 (●, ○). The solid lines corresponded to linear regression fits of glycosylated IgG1, and the dashed lines to the fits for deglycosylated IgG1.

B The SANS $R_G$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG1, each corresponding to a single measurement in histidine buffer in $^2$H$_2$O. The solid and dashed lines correspond to the mean values for glycosylated and deglycosylated IgG1.
Figure 4.7 SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG1 (legend overleaf).
Figure 4.7. SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG1 (continued). The colours denote the glycosylated IgG1 (black), TP1 (blue), TP6 (red) and TP10 (magenta) timepoints. 

A The peak maxima at $M1$ and $M2$ and the maximum length $L$ are indicated by arrows. The SAXS and SANS $P(r)$ curves for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG1 are shown at concentrations of 5.38-0.74 mg/ml. B The corresponding $P(r)$ curves for the SANS curves for IgG1 4.78-0.90 mg/ml. 

(C,D) The concentration dependence of the peak maxima $M1$ and $M2$ for glycosylated and deglycosylated IgG1 are shown. The fitted lines are the mean values for glycosylated IgG1 (solid line), and for TP1, TP6 and TP10 averaged together (dashed lines).
The distance distribution function $P(r)$ is derived from Fourier transformation of the scattering curve $I(Q)$, and provides structural information in real space on glycosylated and deglycosylated IgG1. The X-ray and neutron $P(r)$ analyses gave $R_G$ values that were similar to those from the X-ray Guinier analyses, showing that the two analyses were self-consistent (open symbols, Figure 4.6). The maximum lengths of glycosylated and deglycosylated IgG1 were determined from the value of $r$ when the $P(r)$ curve intersects zero on the $r$ axis and was 17 nm for all four IgG1 samples. The neutron maximum lengths of glycosylated and deglycosylated IgG1, were 16 nm for all four samples. These were 1 nm smaller when compared to the X-ray value of 17 nm. This can be attributed to the differences in instrumental calibration or smearing, resulting in higher errors in curves derived by SANS and thus slight differences in the maximum length ($L$) of the molecule and/or the reduced contribution of the hydration shell seen by neutron scattering (Figure 4.7B). These reductions in the neutron $R_G$ and $L$ values have been previously seen in earlier joint SAXS and SANS studies of antibodies (Rayner et al., 2015).

The maxima in the $P(r)$ curves corresponded to the most frequently occurring distances between scattering elements within the structures, these being a monitor of the IgG1 structure. For the four IgG1 samples, two peaks, $M_1$ and $M_2$, were visible that are characteristic of antibody-shaped proteins. $M_1$ corresponds primarily to the shorter distances within each Fab and Fc region, and is expected to be almost invariant for this reason. $M_2$ corresponds primarily to the longer distances between pairs of Fab and Fc regions (Figures 4.1 and 4.7). No concentration dependence was observed in the positions of the $M_1$ and $M_2$ peaks, which were measured directly from their maximum values (Figures 4.7B,C). However, the X-ray $M_2$ peak shifted significantly from a mean value of $7.44 \pm 0.03$ nm for glycosylated IgG1 to a mean value of $7.83 \pm 0.02$ nm for deglycosylated IgG1 (Figure 4.7C). The X-ray $M_1$ peak shifted much less from $4.42 \pm 0.02$ nm for glycosylated IgG1 to $4.49 \pm 0.01$ nm for deglycosylated IgG1. The same change was seen in the neutron $P(r)$ curves, when $M_2$ increased from $7.27 \pm 0.22$ nm for glycosylated IgG1 to $7.77 \pm 0.04$ nm for deglycosylated IgG1 (Figure 4.7D). The neutron $M_1$ peak shifted slightly from $4.36 \pm 0.07$ nm for
glycosylated IgG1 to 4.21 ± 0.05 nm for deglycosylated IgG1. Both the X-ray and neutron analyses were consistent with each other.

4.3.4 Atomistic modelling of glycosylated and deglycosylated IgG1

In order to account for the changes seen in the $R_{xs-1}$ and $M2$ scattering parameters in IgG1 following deglycosylation, atomistic modelling simulations of the glycosylated and deglycosylated structures were performed, starting from two high-resolution crystal structures for the human Fab and Fc regions. The sequence in the Fab structure was converted into that for IgG1 A33 using Modeller (Figure 4.2A-D). The Fab and Fc regions were joined by a peptide 216EPKSCDKTHTCPPCPAPELLGGP 238 that included the 15-residue hinge sequence (Figure 4.2E), built using PyMOL. The native glycosylated IgG1 models were created by adding complex-type biantennary glycans to the two Asn297 sidechains in the Fc region (Figure 4.1B). This starting structure was subjected to energy minimisation.

Physically-realistic IgG1 models without steric clashes were created for comparison with the experimental X-ray curves. By varying the torsion angles at three flexible regions at the start and end of the two IgG1 hinges (Section 4.2.4) (Figure 4.1), trial IgG1 structures were created that involved movements of the two Fab and one Fc regions relative to each other. For glycosylated IgG1, 2,500,000 models were generated in 16 Monte Carlo simulations, from which 123,284 models were accepted because these showed no steric clashes between separate residues in the model. For deglycosylated IgG1, 2,600,000 models were generated in 20 Monte Carlo simulations, of which 119,191 models were likewise acceptable. To ensure that no systematic trends were overlooked in the modelling outcome, four X-ray and two neutron scattering curves from up to four concentrations were fitted for each of the four samples in question (Figure 4.8). For both glycosylated and deglycosylated IgG1, comparison of the four experimental X-ray scattering curves at 1.36-3.60 mg/ml with the 123,284 and 119,191 theoretical curves gave a goodness-of-fit $R$-factor vs $R_g$ distribution with
The 123,284 models that were accepted for glycosylated IgG1 and the 119,191 models that were accepted for deglycosylated IgG1 are shown as circles. (A,B) Four experimental X-ray scattering curves for four concentrations of glycosylated and deglycosylated IgG1 are shown. (C,D) Two experimental neutron scattering curves for two concentrations of glycosylated and deglycosylated IgG1 are shown. These experimental curves were fitted to the 123,284 and 119,191 modelled curves. Those models with $R_G$ values closest to the experimental $R_G$ values showed the lowest goodness-of-fit $R$-factor as expected. The top 100 best-fit models (red circles) showed the lowest goodness-of-fit $R$-factors. The experimental $R_G$ is represented by a solid vertical line and the dashed vertical lines represent the ±2% upper and lower boundaries of these $R_G$ values.

Figure 4.8 Atomistic modelling analyses for glycosylated and deglycosylated IgG1.
clear minima in all eight cases (Figure 4.8A,B). The minima agreed with the experimental $R_G$ values (Figure 4.6A). The minima showed that enough trial X-ray best-fit models for each concentration (red, Figure 4.8). The range of the 100 $R$-factors for each of the four concentrations was low at between 0.80-1.93% for ray models had been generated to result in good fits in each case. Filtering of the models to select these with the lowest $R$-factors gave the 100 the best-fit glycosylated models and 0.70-2.18% for the best-fit deglycosylated models (Table 4.2). This indicated good quality X-ray curve fits between the experimental and modelled curves.

The eight sets of 100 best-fit models (Figure 4.8) were examined in order to identify the resulting best-fit IgG1 conformations from the curve fits. For this, PCA were performed (David et al., 2014). The PCA determines the correlated motions of protein residues as linearly uncorrelated variables, each being termed a principal component (David et al., 2014). These “essential motions” were extracted from a covariance matrix of the atomic coordinates of the frames in the selected structure set. The eigenvectors of this matrix each have an associated eigenvalue that characterises the clustering of the models based on structural coordinates (or variance). In order to eliminate bias in the PCA, the glycan chains were removed from the glycosylated IgG1 models before comparison with the deglycosylated models. The PCA confirmed a clear difference between the glycosylated and deglycosylated X-ray IgG1 models (black and magenta respectively, Figure 4.9A-D; Table 4.3). Thus, the distributions of the best-fit 400 glycosylated and 400 deglycosylated X-ray models were each clustered into five distinct groups, with little overlap between glycosylated and deglycosylated groups. The glycosylated models mostly occurred in the PCA Groups 1, 2 and 3, while the deglycosylated models mostly occurred in the PCA Groups 4 and 5. This outcome verified the experimentally-observed changes in the $RXS_1$ and $M2$ parameters before and after deglycosylation (Figures 4.6A, 4.7C). The visually-excellent X-ray curve fits confirmed the validity of the modelling fits (Figures 4.10A,B). Of particular note was the agreement of the experimental and theoretical double peaks in the $P(r)$ curves shown as insets.

The same outcome was found with the theoretical neutron modelling and PCA, thus confirming the reproducibility of the curve fits, although the precision of the neutron scattering curves was reduced. The same 123,284 and 119,191
Figure 4.9 Principal component analysis (PCA) of the best-fit glycosylated and deglycosylated IgG1 models (legend overleaf).
Figure 4.9. Principal component analysis (PCA) of the best-fit glycosylated and deglycosylated IgG1 models (continued). Glycosylated models are represented in black and deglycosylated models are represented in magenta. In this, groups 1, 2, 3, 4 and 5 are represented by ○, Δ, +, × and □ in that order, and the centroid model for each group is represented by large numbers (blue) and a ★. (A-D) The eight sets of 100 best-fit models from the experimental X-ray scattering curves for glycosylated and deglycosylated IgG1 were grouped by PCA into five groups as shown in three panels A, B and C of PC2 vs PC1, PC3 vs PC2 and PC3 vs PC1. D, The first three eigenvalue rankings (PC1, PC2 and PC3) captured 81.9% of the variance in the 800 models. (E-H) The four sets of 100 best-fit models from the experimental neutron scattering curves for glycosylated and deglycosylated IgG1 were grouped by PCA into five groups as shown in E, F and G of PC2 vs PC1, PC3 vs PC2 and PC3 vs PC1. H, The first three eigenvalue rankings (PC1, PC2 and PC3) captured 70.4% of the variance in the 400 models.
Figure 4.10 Scattering curve fits to the experimental data for the best-fit model for each of the glycosylated and deglycosylated IgG1 samples (legend overleaf).
Figure 4.10. Scattering curve fits to the experimental data for the best-fit model for each of the glycosylated and deglycosylated IgG1 samples (continued). The experimental curve is denoted by black circles and the best-fit modelled curves are denoted by solid red lines. The distance distribution curves \( P(r) \) are shown in the top right of each panel. 

A glycosylated and B deglycosylated IgG1 X-ray scattering curve fits for four concentrations each. For the four X-ray fits in A, the glycosylated IgG1 models were taken from PCA group 3 (3.60 mg/ml, 3.19 mg/ml, and 2.06 mg/ml) and group 2 (1.36 mg/ml) in that order (Table 4.3). In B, the deglycosylated IgG1 models used for the fits as shown corresponded to PCA group 4 (4.29 mg/ml) and group 1 (3.19 mg/ml, 1.34 mg/ml and 1.02 mg/ml in that order).

C glycosylated and D deglycosylated IgG1 neutron curve fits shown for two concentrations each. In the four neutron fits, the glycosylated IgG1 models corresponded to PCA group 1 (2.60 mg/ml) and PCA group 2 (1.38 mg/ml) in that order (Table 4.5). Likewise, the deglycosylated IgG1 models corresponded to PCA group 3 (2.73 mg/ml) and PCA group 4 (0.90 mg/ml) in that order.
Figure 4.11 Normalised Kratky plots for the experimental and best fit glycosylated and deglycosylated IgG1 scattering curves.

A X-ray experimental data (solid lines) and model fits (dashed lines) were shown in black for glycosylated IgG1 at 3.60 mg/ml and in magenta for deglycosylated IgG1 at 4.29 mg/ml. B Neutron experimental data (solid lines) and model fits (dashed lines) were shown in black for glycosylated IgG1 at 2.60 mg/ml and in magenta for deglycosylated IgG1 at 2.73 mg/ml
Figure 4.12 Views of representative best-fit X-ray and neutron structures.

The blue ribbon cartoon denotes the protein backbone of the starting glycosylated and deglycosylated IgG1 structures. The Fab regions of the 100 best-fit models were superimposed onto these starting structures, thus focussing on movements in the Fc region. The blue and magenta wireframe envelopes denote the space occupied by the glycosylated and deglycosylated Fc regions, respectively, in the 100 best-fit structures for each of the A four X-ray and B two neutron analyses. In five of the six representations, the magenta wireframes (deglycosylated IgG1) occupy a greater region of space compared to the blue wireframes (glycosylated IgG1). C The cartoon representation based on part (b) of A showed the larger range of Fc conformations as arrowed (magenta) after deglycosylation compared to that for glycosylated IgG1 (blue).
theoretical curves were compared with the neutron scattering curves at 0.90-2.73 mg/ml to show again that 100 best-fit structures could be identified at clear minima in each of the R-factor vs $R_G$ neutron distributions (Figures 8C, 8D). The neutron PCA also indicated clear differences between the glycosylated and deglycosylated neutron IgG1 models (black and magenta respectively, (Figure 4.9E-H; Table 4.5). The distributions of the best-fit 200 glycosylated and 200 deglycosylated neutron models were each again clustered into two major groups 1 and 4, and three more less populated groups, with little overlap between glycosylated and deglycosylated groups. The glycosylated models mostly occurred in the PCA Group 1, while the deglycosylated models mostly occurred in the PCA Group 4. Visually-excellent neutron curve fits were obtained (Figures 4.10C,D).

Further insights into the X-ray and neutron data and their modelling were obtained from the dimensionless Kratky analyses of $(Q.R_G)^2 \cdot I(Q)/I(0)$ vs $Q.R_G$ for the experimental scattering curves at the highest concentrations in use and the scattering curves from the modelled best fit structures. These plots indicate whether the macromolecule in question is globular in its structure or possesses intrinsically disordered regions (Receveur-Brechot et al., 2012). The Kratky plots all demonstrated two clear peaks in both the experimental and theoretical modelled curves. For the X-ray Kratky curves (Figure 4.11A), the $Q.R_G$ values for the experimental glycosylated peaks were 1.96 and 4.05, in good accord with the modelled values of 1.93 and 3.98. The $Q.R_G$ values for the experimental peaks for deglycosylated IgG1 were 2.01 and 4.15 were also in good accord with the modelled deglycosylated peaks of 1.97 and 4.08. It was interesting to note that the second peak showed higher intensities for deglycosylated IgG1 (magenta) than for glycosylated IgG1 (black), suggesting that there was a small increase in antibody disorder after deglycosylation. For the SANS Kratky curves (Figure 4.11B), the $Q.R_G$ values for the experimental peaks for glycosylated IgG1 were 2.45 and 4.68, which were similar to the modelled peaks at 1.99 and 4.43. The $Q.R_G$ values for the experimental peaks for deglycosylated IgG1 were 1.91 and 5.16, but showed less agreement for the modelled peaks at 1.91 and 4.37. Again, the second neutron peak showed higher intensities for deglycosylated IgG1 (magenta) when compared with glycosylated IgG1 (black), suggesting that a greater disorder was present after deglycosylation. This was less apparent in
SAXS Kratky curves, and this could attribute to lack of hydration shell that allows the changes in relative disorder to be better captured.

As another test of the scattering modelling, the $s_{20,w}^0$ values for the eight sets of best-fit 100 glycosylated and deglycosylated models from each X-ray concentration (Figures 4.8, 4.10) were calculated using HullRad (Fleming et al., 2018). This gave an $s_{20,w}^0$ range of 6.57-6.77 S for the four X-ray concentrations for glycosylated IgG1 and 6.24-6.50 S for deglycosylated IgG1 (Table 4.2). These values agreed well with the experimental $s_{20,w}^0$ values of 6.16-6.43 S for glycosylated IgG1 and 6.09-6.15 S for deglycosylated IgG1 (Table 4.1). These agreements corroborated the outcome of the atomistic scattering modelling, given that the mean difference between the modelled and experimental values should typically be ± 0.21 S for related macromolecules (Perkins et al., 2009). This modelling of sedimentation coefficients was however unable to distinguish changes before and after deglycosylation.

4.4 Discussion

This scattering and atomistic modelling study has notably clarified the conformational effect of removing the two glycans chains on the structure of the major IgG1 antibody subclass. Unlike earlier protein structural investigations based on crystallography, nuclear magnetic resonance (NMR) or circular dichroism (CD), the approach used in this work provided a more informative outcome on full-sized IgG1 of the changes accompanying deglycosylation of the two conserved Asn$^{297}$ residues in the Fc region (Figure 4.1B). The complete deglycosylation of IgG1 was validated by a combination of gel filtration, routine mass spectrometry, and AUC. The AUC data showed that the IgG1 samples were monomeric and showed a slight concentration dependence in the $s_{20,w}^0$ values that were extrapolated to give the $s_{20,w}^0$ values. Subsequently, the glycosylated and deglycosylated IgG1 proteins were submitted to abundant SAXS and SANS data collection to establish their Guinier $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values, and their distance distribution curves $P(r)$. Small changes in the $R_{XS-1}$ values and larger changes in the $M_2$ parameter that monitored the mean separation of the Fab and Fc regions in IgG1 were seen on deglycosylation. The advent of atomistic scattering modelling using SASSIE (Perkins et al., 2016) based on molecular
dynamics and Monte Carlo simulations gave excellent curve fits based on large stereochemically-correct trial conformational libraries for IgG1. The display and interpretation of the output was much facilitated by PCA analyses. Two different conformational best-fit structures for glycosylated and deglycosylated IgG1 were identified by PCA. The clearest view of the final result was determined from wireframe representations of the 100 best fit structures (Figure 4.12). Views of the 100 best-fit X-ray models at four concentrations showed that the glycosylated Fc region occupied a smaller volume (blue wireframe, Figure 4.12A) than the notably larger volume occupied by the deglycosylated Fc regions (magenta wireframe, Figure 4.12A). This outcome showed that the Fab and Fc regions formed better defined native glycosylated structures compared to the more dispersed and flexible structures seen after deglycosylation. The 100 best-fit neutron models were limited by the reduced precision of the neutron scattering data, but are consistent with this interpretation (Figure 4.12B). The Kratky plots also suggest greater disorder following deglycosylation (Figure 4.11), in keeping with the flexibility shown in Figure 4.12A.

These results (Figure 4.12A) account for previous functional studies of human IgG1. The Fc region of IgG1 is responsible for interactions with the three classes of human Fcy receptors (FcγRs) and with the globular heads of human C1q. Several studies have reported that both interactions are abrogated following deglycosylation. Thus, the antibody effector functions such as antibody-dependent cytotoxicity and complement dependent cytotoxicity mediated by FcγRs and C1q are impaired for aglycosylated and deglycosylated antibodies (Nose et al., 1983; Mimura et al., 2001b; Krapp et al., 2003). Deglycosylated IgG1 was used as a negative control in surface plasmon resonance experiments in which several FcγRs were tested, and deglycosylated IgG1 failed to bind to these with the exception of the high affinity FcγRI (Thomann et al., 2015). These structural investigations explain these findings by showing that the orientation of the Fc region in IgG1 has become disorganised (Figure 4.12A). The cartoon view (Figure 4.12C) showed that the deglycosylated Fc region occupied more conformational space than the glycosylated one. High-resolution crystal structures of glycosylated Fc regions bound to the FcγRs showed none or little interaction with the Fc glycan chains, and receptor binding generally occurred at
residues in the lower hinge (Sondermann et al., 2000; Mimoto et al., 2013; Kiyoshi et al., 2015; Oganesyan et al., 2015). Accordingly, IgG1 deglycosylation means that the essential presentation of structurally well-defined C\(\text{H2}\) domain surfaces in the Fc region near the hinge peptides is no longer present. It follows from this conclusion that the glycoengineering of human IgG1 based on either the removal or modification of specific glycosylation patterns in the two Fc glycans will influence receptor binding and in turn the effector functions if human IgG1, such as its core fucosylation (Niwa et al., 2005; Matsumiya et al., 2007; Pereira et al., 2018) and terminal sialic acids (Quast et al., 2015).

The advantage of the combined SAXS-SANS-AUC-MC approach is the ability to address the full IgG1 structure, this being the functional native structure. Previous structural studies have focussed on the Fc region alone, because it is difficult to crystallise the full IgG1 antibody compared to the Fc region alone, and NMR and circular dichroism studies are more difficult with the full IgG1 structure because its molecular mass has tripled to over 150 kDa. Previous NMR solution studies of the glycosylated and deglycosylated Fc region (Subedi et al., 2015) showed that the glycan chain at Asn\(^{297}\) stabilises the loop between \(\beta\)-strands C’ and E in the C\(\text{H2}\) domains, and in turn positions the two C\(\text{H2}\) domains into a stable orientation seen in 22 Fc crystal structures to orient the Fc\(\gamma\)R interface on the Fc region for optimal binding affinity with receptors. Previous crystallography studies of the deglycosylated Fc region showed that the C\(\text{H2}\) domains had reorientated themselves to form a more compact structure (Borrok et al., 2012). Circular dichroism solution studies of the glycosylated and deglycosylated Fc region studied the \(\beta\)-sheet secondary structure of its four domains (Figure 4.1) (Alsenaidy et al., 2014). By this, similar \(\beta\)-sheet rich structures showing a minimum at 217 nm were observed for both forms of the Fc region, showing that this \(\beta\)-sheet structure was preserved with or without the glycan chains. The greater mobility of the Fc region following glycan removal, as observed in this study (Figure 4.12A), would not have been observed by circular dichroism. These previous studies on the Fc region alone complement the results showing that the Fc region within intact IgG1 is more flexible after deglycosylation.

The advent of atomistic scattering modelling has resulted in a molecular explanation of the changes induced in IgG1 following glycan removal.
Traditionally solution scattering is a low-resolution method with a resolution of around 2 nm, while protein crystallography routinely achieves resolutions that are over ten times better. IgG1 modelling analysis used "constrained" modelling based on the Fab and Fc crystal structures and joined by a hinge region that was conformationally randomised using molecular dynamics to give 20,000 trial models (Rayner et al., 2015). Seven different fits from X-ray and neutron data out to $Q$ values of 1.5 nm$^{-1}$ for human IgG1 6a and 19a in different NaCl buffers corresponded to clear minima in the $R$-factor vs. $R_G$ graphs. All these revealed an asymmetric solution structure for IgG1, in agreement with the single asymmetric conformation seen in the crystal structure of human IgG1 b12 (Saphire et al., 2007), and that for glycosylated IgG1 from SasCalc (Figure 4.12). The best-fit $R$-factors were 2.8-3.7%. This conformation permitted the Fc region to bind readily to its FcγR and C1q ligands without steric clashes. The follow-up study on human IgG1 using “atomistic” scattering modelling involved further X-ray and neutron data collection out to $Q$ values of 1.5 nm$^{-1}$, and fitting these data to 231,492 trial models produced from full molecular dynamics and rapid Monte Carlo simulations (Wright et al., 2019). The best-fit $R$-factors were 2.9%. This improved method likewise gave an asymmetric IgG1 solution structure similar to that seen in the crystal structure of human IgG1 b12 (Saphire et al., 2007), and also that for glycosylated IgG1 from SasCalc (Figure 4.12). This likewise accounted for the binding of the Fc region to its FcγR and C1q ligands. Of particular interest is that the previous use of the SCT/SCTPL modelling approach (available in SASSIE-web) had explicitly incorporated hydration shells in a coarse-grained approach (Wright et al., 2015). The current atomistic modelling study using SasCalc (Watson et al., 2013) did not include atomistic representation of hydration shells because this is computationally expensive. The final outcomes from all three X-ray modellings were similar, when all three showed asymmetric IgG1 solution structures. In the present study, the simulations of IgG1 A33 were based on X-ray and neutron data that extended out to $Q$ values of 1.5 nm$^{-1}$ and resulted in fits with low $R$-factors of 1% or less (Table 4.2). This $R$-factor improvement is attributed to the improved signal-noise ratio of the scattering curves from the B21 instrument at Diamond. The atomistic modelling approach in combination with high quality SAXS data with little noise
at large $Q$ values has been of great value in studying structural perturbations in IgG antibodies caused by the removal of its two glycans.
Chapter 5 Atomistic solution structures of glycosylated and deglycosylated human IgG4 reveal the stabilising role of its $\text{C}_\text{H}2$ glycans
5.1 Introduction

Immunoglobulin G (IgG) is the most abundant subclass in human serum and is widely exploited as a biotherapeutic due to its high specificity to antigens. There are four IgG subclasses, IgG1, IgG2, IgG3 and IgG4, numbered in order of their concentration in serum which are 8.0 mg/ml, 4.0 mg/ml, 0.8 mg/ml and 0.4 mg/ml, respectively (Hamilton, 1987). IgG is arranged as a characteristic Y-shape, comprised of two Fab regions, which bind with high specificity and affinity to an antigen and a single Fc region, which interacts with Fcγ receptors (FcγRs) on immune cells (Figure 5.1A). The Fab and Fc regions are connected by hinges of varying lengths depending on the subclass. Of the four subclasses, IgG4 has the shortest hinge with 12 residues, held together by two disulphide bonds at Cys^{226} and Cys^{229} (Figure 5.1A) (Vidarsson et al., 2014). IgG4 has several interesting characteristics, including the ability of wild-type IgG4 to undergo a phenomenon termed Fab arm exchange (FAE), whereby the heavy chains dissociate and reassociate in vivo to form bispecific antibodies (Aalberse et al., 1999). In consequence of this, IgG4 functions as though it were monovalent and prevents cross-linking of antibody-antigen complexes, further contributing to its anti-inflammatory nature (van der Zee et al., 1986). This process can be abrogated by a S228P hinge mutation (EU numbering) (Aalberse et al., 1999). IgG4 is also considered an anti-inflammatory antibody due to its inability to activate complement (van der Zee et al., 1986), unlike its other IgG counterparts. Compared to IgG1, IgG4 has a reduced binding affinity for pro-inflammatory FcγRs, including FcγRIIIa, which is implicated in antibody-dependent cell-mediated cytotoxicity (ADCC) (Bruhns et al., 2015). These anti-inflammatory characteristics makes IgG4 a desirable therapeutic treatment for pathologies in which inflammation is problematic. There are 15 approved IgG4 based antibodies available on the market (Reichert, 2021).

Antibodies of the IgG class have a conserved N-linked glycan at the C_{\text{H2}} domain of the Fc region attached to Asn^{297}, which plays a functional role (Figure 5.1B). Each glycan is typically complex-type biantennary with a Man_3GlcNAc_2 core and two NeuNAc.Gal.GlcNAc (Deisenhofer, 1981) (Figure 5.2), however the composition of the two glycans is chemically heterogenous (Wuhrer et al., 2007).
The heavy chains are comprised of $V_H$, $C_{H1}$, $C_{H2}$ and $C_{H3}$ domains, and the light chains are comprised of $V_L$ and $C_L$ domains. The heavy chains are connected by two Cys-Cys disulphide bridges at Cys$^{226}$ and Cys$^{229}$. An N-linked oligosaccharide at Asn$^{297}$ is present on each of the $C_{H2}$ domains. The hinge region connecting the Fab and Fc regions was constructed from 20 residues $^{216}$ESKYGPPCPPCPAPEFLGGP$^{235}$.

At the left, the glycosylation of IgG4 Fc at two Asn$^{297}$ residues in the IgG4 Fc region is shown as stick models (PDB ID: 4C55). The three hinge tripeptides that were conformationally varied in the TAMC searches are shown and circled in green. The central schematic shows the glycosylation pattern used in this study ($N$-acetyl glucosamine, GlcNAc; mannose, Man; galactose, Gal; $N$-acetyl neuraminic acid, NeuNAc. At the right, a single $C_{H2}$ domain with its glycan chain is shown, with the glycan residue colours coordinated with those in the schematic.
Figure 5.2 Sequence alignment of human IgG4 (legend overleaf).
Figure 5.2. Sequence alignment of human IgG4 (continued). (A-G) The IgG4 A33 sequence was kindly provided by UCB. The IgG4 Pro and Ser sequences were taken from (Rayner et al., 2014). The IgG4 b72.3 sequence was taken from the crystal structure (PDB ID: 1BBJ). The Fc sequence was taken from the crystal structure (PDB ID: 4C55). A, B, the V<sub>L</sub> and C<sub>L</sub> domains; C-D, the V<sub>H</sub> and C<sub>H1</sub> domains; E, the hinge with the TAMC-varied tripeptides identified in green, and the S225P mutation identified in pink; F, G, the C<sub>H2</sub> and C<sub>H3</sub> domains, with Asn<sup>297</sup> in blue. EU sequence numbering was used, and the complementarity-determining regions (CDRs) sequences were identified in red. Beneath the alignments, consensus symbols indicated the degree of conservation, where (*) indicates full conservation, (:) indicates conservation between groups of strongly similar properties based on the Gonnet PAM 250 matrix, (.) indicates conservation between groups of weakly similar properties, and a space indicates no conservation.
The IgG-Fc glycan modulates the binding specificity of FcγRs (Wada et al., 2019). The ability to engineer the Fc glycan is of growing interest in biotherapeutics, in which the aim is to modulate IgG4 function by influencing which FcγRs it can bind to. For example, afucosylated IgG4 antibodies are able to elicit a stronger ADCC response through their binding to FcγRIIIa receptors (Gong et al., 2016). Aglycosylated IgG4 was unable to bind to FcγRIIIa, indicating that glycans play an important role in FcγR binding (Kang et al., 2020).

Structural studies of IgG4 are essential to understand how this bind to FcγRs. There are two high-resolution crystal structures for full length IgG4 (PDB IDs: 5D43 and 6GFE) (Scapin et al., 2015; Blech et al., 2019). These structures only provide a single snapshot of the IgG4 structure and do not illustrate the full conformational space that the Fc and Fab regions can occupy in solution. Six glycosylated IgG4 Fc crystal structures show the glycans facing inward within the C\(\text{H}_2\) domains (PDB IDs: 4C54, 4C55, 5LG1, 5W5M and 5W5N) (Davies et al., 2014; Davies et al., 2017; Tam et al., 2017). One high-resolution structure of a deglycosylated IgG4-Fc (PDB ID: 4D2N) revealed that the C\(\text{H}_2\)-C\(\text{H}_2\) domain interactions partially bury the C\(\text{H}_2\) surface that would otherwise been solvent exposed with the presence of a glycan, with a change in DE loop conformation, to which the glycan is attached (Davies et al., 2015). Unfortunately, crystal structures for IgG-Fc in complex with the FcγRs are limited to Fc regions from the IgG1 subclass. Therefore, it is currently difficult to deduce the structural basis for FcγRs binding to IgG4-Fc and the role that glycans might play in this interaction.

The effect of the two Fc glycans on the full IgG4 structure (and not just its Fc region) is not well understood. To address this question, small angle X-ray and neutron scattering (SAXS, SANS) and analytical ultracentrifugation (AUC), together with atomistic modelling, were jointly applied to intact IgG4 as a powerful solution structural approach. SAXS measured in light water buffers provides data sets measured in a high positive solute-solvent contrast, in which the contribution of the hydrophilic surface regions of the glycoprotein are accentuated, while SANS measured in heavy water buffers provides data sets measured in a high negative solute-solvent contrast, in which the contribution of the buried hydrophobic core of the glycoprotein is accentuated (Perkins, 1986, 2001; Svergun et al., 1998). The tightly-bound hydration layer on the protein surface is
detected by SAXS because its electron density is similar to that of the protein and not to bulk water, while the effect of this same hydration layer is much reduced by SANS measured in heavy water, because its nuclear density is almost the same as that of bulk water. The reproducibility of the two data sets will be a test of their outcome because radiation effects in SAXS and protein aggregation in heavy water by SANS may perturb the output of either method. Their utility is much enhanced by the development of atomistic modelling of the scattering data sets molecular dynamics and Monte Carlo methods (Perkins et al., 2016). Previous atomistic scattering modelling with glycosylated IgG4 revealed that the arrangement of the Fab regions restricted access to the Fc region limiting binding to FcγRs and C1q ligands (Rayner et al., 2014). A study of deglycosylated IgG1 demonstrated that deglycosylated IgG1 showed a more conformationally flexible structure than that of glycosylated IgG1 (Chapter 4, Spiteri et al., 2021). Here, a joint approach is applied to show that deglycosylation also results in a more flexible Fc structure within IgG4, in turn affecting the receptor-binding function of IgG4.

5.2 Materials and Methods

5.2.1 Purification and composition of IgG4

IgG4 A33 (146.9 kDa) was generously provided by Dr John O’Hara and Dr Berni Sweeney (UCB). Its enzymatic deglycosylation was achieved using peptide-N-glycosidase F (PNGase F) (35.5 kDa, New England Biolabs, Massachusetts, USA), selected due to its ability to remove glycans completely from glycosylated Asn residues (Plummer et al., 1984). To digest the glycans, 3.7 μl PNGase F (1850 activity units) was used to deglycosylate 150 μl of IgG4 (14.0 mg/ml). Native IgG4 were incubated at 37°C for time points of 1 hour (TP1), 6 hours (TP6) and 10 hours (TP10). Each deglycosylated IgG4 sample was filtered using Amicon Ultra-0.5 ml centrifugal filters (100 kDa cut-off) which simultaneously allowed the PNGase F to pass through the membrane, while concentrating the deglycosylated IgG4 sample. Immediately before SAXS, SANS or AUC measurements, glycosylated and deglycosylated IgG4 were further purified by gel filtration to remove any non-specific aggregates using a Super 6 Increase 10/300 GL column (Cytiva, Amersham, UK), then concentrated using
Amicon Ultra-15 spin concentrators (100 kDa cut-off) and dialyzed at 4°C into 20 mM L-histidine, 138 mM NaCl, and 2.6 mM KCl buffer, pH 6.0. This histidine buffer was found to increase the stability of IgG4. The sequence of IgG4 A33 was aligned against those from Ser^{222} and IgG4 Pro^{222} from Rayner et al (2014) and an IgG4 Fab structure from B72.3 (PDB ID: 1BBJ) (Brady et al., 1992) and an IgG4 Fc structure (PDB ID: 4C55) (Davies et al., 2014) (Figure 5.2). The N-linked oligosaccharide at Asn^{297} on the C_{H2} domains were assumed to be complex-type biantennary oligosaccharide structures with a Man_{3}GlcNAc_{2} core and two NeuNAc.Gal.GlcNAc antennae (Deisenhofer, 1981). From the sequence, the molecular mass of glycosylated IgG4 A33 was calculated to be 148.1 kDa, its unhydrated volume was 190.7 nm^3, its hydrated volume was 251.2 nm^3, its partial specific volume v was 0.730 ml/g and its absorption coefficient was 14.0 cm\(^{-1}\). The molecular mass of deglycosylated IgG4 A33 was calculated to be 143.7 kDa, its unhydrated volume was 186.0 nm^3, its hydrated volume was 244.6 nm^3, its partial specific volume v was 0.732 ml/g and its absorption coefficient was 14.5 cm\(^{-1}\) (Perkins, 1986). The X-ray and neutron scattering densities of glycan residues are similar to those for hydrophilic (polar) amino acid residues, these being slightly higher than those for hydrophobic (non-polar) amino acid residues (Perkins, 1986). The buffer density was measured on an Anton Paar DMA 5000 density meter at 20°C to be 1.00578 g/ml in light water. In heavy water, the density was 1.11106 g/ml. Buffer viscosities were measured on an Anton Paar AMVn Automated microviscometer at 20°C. The viscosity in light and heavy water, pH 6.0 was measured to be 0.010190 and 0.01384 poise respectively.

The completeness of deglycosylation was verified by Superose 6 gel filtration, SDS-PAGE, and mass spectrometry. In the Mass Spectrometry Facility at the Chemistry Department, University College London, the antibodies were analysed on an Agilent 6510 Quadrupole time-of-flight liquid chromatography mass spectrometry system (Agilent, Stockport, UK). Ten µL of each sample was injected onto a PLRP-S, 1000A, 8 µM, 150 mm x 2.1 mm column, which was maintained at 60°C at a flow of 0.3 ml/min. The separation was achieved using mobile phase A (water with 0.1% formic acid) and B (acetonitrile, with 0.1% formic acid) using a gradient elution. The column effluent was continuously electrospayed into the capillary electrospray ionization source of the Agilent 6510 QTOF mass spectrometer and electrospray ionization mass spectra were
acquired in positive electrospray ionisation mode using the m/z range 1,000–3200 in profile mode. The raw data was converted to zero charge mass spectra using the maximum entropy deconvolution algorithm in the MassHunter software version B.07.00 (Agilent, Stockport, UK). The glycan masses were calculated by subtracting the mass of the full glycosylated IgG4 from the partially deglycosylated glycoform giving the mass of a single glycan chain. The single glycan mass was also found by subtracting the mass of fully glycosylated IgG4 from that for deglycosylated IgG4 and halving this mass.

5.2.2 Sedimentation velocity data and analysis of IgG4

Analytical ultracentrifugation data for glycosylated and deglycosylated IgG4 at timepoints TP1, TP6 and TP10 were obtained on two Beckman XL-I instruments equipped with AnTi50 rotors (Beckman Coulter, High Wycombe, UK). Data were collected at 20°C, at a rotor speed of 40,000 rpm in two-sector cells with column heights of 12 mm for approximately 6 hours. Sedimentation analyses were performed using direct boundary Lamm fits of up to 900 scans using SEDFIT (version 15.01b) (Schuck, 1998, 2000). SEDFIT resulted in size-distribution analyses $c(s)$, for which the algorithm assumes that all species have the same frictional ratio $f/f_0$. The final SEDFIT analyses used a fixed resolution of 200 and optimized the $c(s)$ fits by floating $f/f_0$ and the baseline until the overall root mean square deviations and visual appearance of the fits were satisfactory. The percentage of oligomers in the total loading concentration was derived using the $c(s)$ integration function. The buffer viscosity and density were measured experimentally (above). Values were corrected to $s_{20,w}$ by:

$$s_{20,w} = s_{T,B} \left( \frac{\eta_{T,B}}{\eta_{20,w}} \right) \left( \frac{1 - \bar{v}\rho}{1 - \bar{v}\rho_{T,B}} \right)$$

where $s$ is the observed sedimentation coefficient, the subscripts $T,B$ refers to the temperature of the buffer, the subscripts $20,w$ refers to water at 20°C. $\rho$ is the solvent density, $\eta$ is the solvent viscosity and $\bar{v}$ is the protein partial specific volume.
5.2.3 X-ray and neutron scattering data and analyses for IgG4

X-ray scattering data was obtained during one beam session (October 2017) on Instrument B21 (Cowieson et al., 2020) at the Diamond Light Source at the Rutherford Appleton Laboratory (Didcot, UK), operating with a ring energy of 3 GeV, the beamline operational energy is 12.4 keV. A PILATUS 2M detector with a resolution of 1475 × 1679 pixels (pixel size of 172 × 172 µm) was used with a sample-to-detector distance of 4.01 m giving a \( Q \) range from 0.04 nm\(^{-1}\) to 4 nm\(^{-1}\) (where \( Q = 4 \pi \sin \theta / \lambda \); \( 2\theta = \) scattering angle; \( \lambda = \) wavelength). The glycosylated IgG4 (1.33-4.03 mg/ml) and the TP1 (1.63-4.70 mg/ml), TP6 (1.06-3.10 mg/ml) and TP10 (0.85-3.07 mg/ml). Samples in light water were loaded onto a 96 well plate and placed into an EMBL Arinax sample holder (Pernot et al., 2013; Round et al., 2015). This condition showed the antibody molecule as a hydrated structure in a high positive solute-solvent contrast (Perkins, 1986). An automatic sampler injected 30 µl of sample from the plate into a temperature-controlled quartz cell capillary with a diameter of 1.5 mm. Data sets of 30 frames with a frame exposure time of 1 second each were acquired in duplicate as a control of reproducibility. Checks during data acquisition confirmed the absence of radiation damage. ScÅtter (version 3.0) was used for buffer subtraction and data reduction, in which the 30 frames were averaged (Rambo, 2021).

Neutron scattering data on glycosylated IgG4 (1.76-6.05 mg/ml) and the TP1 (0.71-1.93 mg/ml), TP6 (2.38 mg/ml) and TP10 (5.62 mg/ml) samples in heavy water were obtained in two sessions (March and October 2017) on instrument SANS2D at the ISIS pulsed neutron source at the Rutherford Appleton Laboratory (Didcot, UK) (Heenan et al., 2011). This condition showed the antibody structure with a near-invisible hydration shell in a high negative solute-solvent contrast (Perkins, 1986). No conformational differences in the antibody between light and heavy water were detected in this study or previously (Wright et al., 2019). A pulsed neutron beam was derived from proton beam currents of ~40 µA. SANS2D data were recorded with 4 m of collimation, a 4 m sample-to-detector distance, a 12 mm sample aperture, and a wavelength range of 0.175-1.65 nm made available by a time of flight. This gave a \( Q \) range from 0.05 nm\(^{-1}\) to 4 nm\(^{-1}\). The data were acquired using a two-dimensional \(^3\)He detector with 512 × 512 pixels of 7.5 × 75 mm\(^2\) in size. Samples of volume 1 ml were measured in
2 mm path length circular banjo cells for 1-7 h in a thermostated sample rack at 20°C. Data were reduced using MANTID software (Arnold et al., 2014). The MANTID data reduction steps include corrections for the Q resolution, i.e. beam divergence effects and smearing from the shape and size of the slits, as well as the wavelength overlap in each pulse (Arnold et al., 2014). Using SASview software, the Guinier analyses (below) were found to be almost unaffected if the smearing was turned on or off.

Guinier analyses of the scattering data give information of the radius of gyration \( R_G \), cross-sectional radius \( (R_{XS}) \), molecular mass. The scattering curve \( I(Q) \) intensities at low Q are defined by the \( R_G \) value which is the averaged distance of each scattering point from the centre of scattering. In a given solute-solvent contrast, the radius of gyration \( R_G \) is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low Q gave the \( R_G \) value and the forward scattering at zero angle \( I(0) \) (Glatter et al., 1982):

\[
\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}
\]

For antibodies, this expression is valid in a \( Q.R_G \) range up to 1.5, and was used in previous studies (Rayner et al., 2014, 2015; Wright et al., 2019), although the usual upper range reported in the literature is 1.0-1.3. If the structure is elongated, the mean radius of gyration of the cross-sectional structure \( R_{XS} \) and the mean cross-sectional intensity at zero angle \( I(Q) \) is obtained from (Glatter et al., 1982):

\[
\ln[I(Q)Q] = [I(Q)Q]_{Q \rightarrow 0} - \frac{R_{XS}^2 Q^2}{2}
\]

For immunoglobulins, it has been long recognised that the cross-sectional plot exhibits two regions, a steeper innermost one and a flatter outermost one (Pilz et al., 1970), and the two analyses are denoted by \( R_{XS-1} \) and \( R_{XS-2} \) respectively. The \( R_{XS-1} \) parameter represents the averaged overall spatial separation of the Fab and Fc regions, while the \( R_{XS-2} \) parameter represents the averaged spatial cross-section of the two Fab and one Fc region. The \( R_G \) and \( R_{XS} \) analyses were performed using SCT (Table 5.1) (Wright et al., 2015). The \( Q \) ranges for the \( R_G \), \( R_{XS-1} \) and \( R_{XS-2} \) values were 0.10–0.22, 0.29–0.52, and 0.66–
Table 5.1 Experimental data by X-ray and neutron scattering and analytical ultracentrifugation for glycosylated and deglycosylated IgG4.

<table>
<thead>
<tr>
<th></th>
<th>Concentration (mg/ml)</th>
<th>$R_g$ (nm)</th>
<th>$R_{XS-1}$ (nm)</th>
<th>$R_{XS-2}$ (nm)</th>
<th>$L$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X-ray data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG4 Glycosylated</td>
<td>4.03</td>
<td>4.94 ± 0.27</td>
<td>2.51 ± 0.16</td>
<td>1.42 ± 0.12</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2.79</td>
<td>4.89 ± 0.29</td>
<td>2.50 ± 0.20</td>
<td>1.41 ± 0.14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
<td>4.90 ± 0.29</td>
<td>2.50 ± 0.18</td>
<td>1.39 ± 0.13</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.33</td>
<td>4.82 ± 0.35</td>
<td>2.50 ± 0.19</td>
<td>1.40 ± 0.16</td>
<td>15</td>
</tr>
<tr>
<td>IgG4 TP1</td>
<td>4.7</td>
<td>5.00 ± 0.27</td>
<td>2.51 ± 0.16</td>
<td>1.41 ± 0.11</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>4.2</td>
<td>4.98 ± 0.29</td>
<td>2.52 ± 0.16</td>
<td>1.40 ± 0.12</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.87</td>
<td>4.88 ± 0.31</td>
<td>2.51 ± 0.17</td>
<td>1.41 ± 0.14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.63</td>
<td>4.88 ± 0.32</td>
<td>2.50 ± 0.19</td>
<td>1.39 ± 0.15</td>
<td>15</td>
</tr>
<tr>
<td>IgG4 TP6</td>
<td>3.1</td>
<td>4.97 ± 0.29</td>
<td>2.52 ± 0.16</td>
<td>1.40 ± 0.13</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2.24</td>
<td>4.94 ± 0.32</td>
<td>2.51 ± 0.16</td>
<td>1.40 ± 0.14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.79</td>
<td>4.91 ± 0.34</td>
<td>2.51 ± 0.18</td>
<td>1.40 ± 0.14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.06</td>
<td>4.85 ± 0.40</td>
<td>2.49 ± 0.20</td>
<td>1.38 ± 0.17</td>
<td>15</td>
</tr>
<tr>
<td>IgG4 TP10</td>
<td>3.07</td>
<td>4.95 ± 0.29</td>
<td>2.51 ± 0.16</td>
<td>1.38 ± 0.13</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.93</td>
<td>4.91 ± 0.30</td>
<td>2.50 ± 0.19</td>
<td>1.39 ± 0.15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.21</td>
<td>4.86 ± 0.36</td>
<td>2.49 ± 0.20</td>
<td>1.36 ± 0.17</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>0.85</td>
<td>4.72 ± 0.41</td>
<td>2.44 ± 0.23</td>
<td>1.33 ± 0.18</td>
<td>15</td>
</tr>
<tr>
<td><strong>Neutron data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG4 Glycosylated</td>
<td>6.05</td>
<td>5.32 ± 0.99</td>
<td>2.41 ± 0.29</td>
<td>1.15 ± 0.22</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>4.52</td>
<td>5.12 ± 0.77</td>
<td>2.47 ± 0.31</td>
<td>1.20 ± 0.24</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.76</td>
<td>5.19 ± 0.88</td>
<td>2.43 ± 0.30</td>
<td>1.11 ± 0.25</td>
<td>16</td>
</tr>
<tr>
<td>IgG4 TP1</td>
<td>1.93</td>
<td>5.19 ± 0.78</td>
<td>2.43 ± 0.39</td>
<td>1.13 ± 0.24</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>0.71</td>
<td>5.32 ± 0.63</td>
<td>2.48 ± 0.35</td>
<td>1.22 ± 0.25</td>
<td>16</td>
</tr>
<tr>
<td>IgG4 TP6</td>
<td>2.38</td>
<td>5.12 ± 0.75</td>
<td>2.46 ± 0.26</td>
<td>1.21 ± 0.22</td>
<td>16</td>
</tr>
<tr>
<td>IgG4 TP10</td>
<td>5.62</td>
<td>5.35 ± 0.79</td>
<td>2.50 ± 0.22</td>
<td>1.20 ± 0.19</td>
<td>16</td>
</tr>
</tbody>
</table>

**Analytical ultracentrifugation data**

<table>
<thead>
<tr>
<th></th>
<th>$s_{20,w}$ (S)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG4 Glycosylated</td>
<td>3.89</td>
</tr>
<tr>
<td></td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>1.02</td>
</tr>
<tr>
<td>IgG4 TP1</td>
<td>3.25</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>0.81</td>
</tr>
<tr>
<td>IgG4 TP6</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>3.52</td>
</tr>
<tr>
<td></td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>0.88</td>
</tr>
<tr>
<td>IgG4 TP10</td>
<td>2.61</td>
</tr>
<tr>
<td></td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
</tr>
</tbody>
</table>

*Standard error of the mean for $s_{20,w}$ data as follows; ± 0.00 S for glycosylated, ± 0.01 S for TP1, ± 0.02 S for TP6 and ± 0.07 S for TP10.
1.05 nm\(^{-1}\), respectively, as previously (Rayner et al., 2014, 2015; Wright et al., 2019). Indirect transformation of the scattering data \(I(Q)\) in reciprocal space into real space to give the distance distribution function \(P(r)\) was carried out using GNOM (version 4.6) (Semenyuk et al., 1991; Svergun, 1992).

\[
P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q)Qr \sin(Qr) \, dQ
\]

\(P(r)\) corresponds to the distribution of distances \(r\) between the volume elements in the macromolecule. This yields the maximum dimension of the macromolecule \(L\) and its most commonly occurring distance vector \(M\) in real space. For this \(P(r)\) analysis, the X-ray \(I(Q)\) curve utilized up to 755 data points in the \(Q\) range between 0.032 and 1.50 nm\(^{-1}\) for both glycosylated and deglycosylated IgG4. The neutron \(P(r)\) curve utilized up to 155 \(I(Q)\) data points in the \(Q\) range between 0.055 and 1.60 nm\(^{-1}\) for both glycosylated and deglycosylated IgG4.

### 5.2.4 Atomistic modelling of IgG4

Starting structures were created for each of glycosylated and deglycosylated IgG4 A33 based on the A33 sequence provided by UCB. The latter was aligned with the sequences of IgG4 Ser\(^{222}\) and Pro\(^{222}\) (Rayner et al., 2014) (Figure 5.2). This multiple sequence alignment was generated using Clustal Omega software (EMBL-EBI) (Sievers et al., 2011). The Fab structure (Figure 5.1) was based on that found in the chimeric IgG4 b72.3 crystal structure (PDB ID: 1BBJ) (Brady et al., 1992) and the Fc structure was based on that for the serum-derived IgG4 antibody Fc crystal structure (PDB ID: 4C55) (Davies et al., 2014). Modeller (version 9.19) (Webb et al., 2016) was used to generate the human IgG4 structure by adding the A33 Fab sequence to the IgG4 b72,3 Fab structure. The IgG4 hinge was built using a PyMOL script build_seq (PyMOL Script Repository, Queen’s University, Ontario, Canada), based on the sequence of \(216\text{ESKYGPPCCPPEFLGGP}^{235}\). The two N-linked oligosaccharides at Asn\(^{297}\) on the C\(_{\text{H}2}\) domains were approximated as complex-type biantennary oligosaccharide structures with a \(\text{Man}_3\text{GlcNAc}_2\) core and two NeuNAc.Gal.GlcNAc antennae (Deisenhofer, 1981). The glycan template was taken from the GitHub repository (https://github.com/dww100), which was energy
minimized using NAMD (Phillips et al., 2005) for 1 nanosecond to achieve a relaxed structure. This glycan was added to the Fc region by bringing the C1 atom in the first GlcNAc residue to within 0.14 nm to the N sidechain atom of Asn297 in the C12 domain of IgG4, while ensuring no steric clashes with the Fc residues and the glycan chain. The PDB file was then opened on Discovery Studio (Dassault Systèmes BIOVIA, San Diego) where “CONECT” records were created for these glycosidic bonds. The CHARMM force field parameters and protein structure file (PSF), including those for the disulphide bridges and glycans were generated using the CHARMM-Gui GlycanReader tool (Miller et al., 2008; Jo et al., 2011; Lee et al., 2016) in order to be compatible with the CHARMM36 forcefield (MacKerell et al., 1998; Mackerell et al., 2004; Guvench et al., 2009; Raman, Guvench et al., 2010b; Best et al., 2012). To relax this structure, the full IgG4 structure with and without glycans were then energy minimised using the simulation engine NAMD (version 2.9) with the CHARMM36 forcefield.

For the Monte Carlo simulations to generate trial structures, the starting IgG4 structure was renumbered and its naming nomenclature was adjusted to match the required format for the Torsion Angle Monte Carlo (TAMC) module on SASSIE-web (Zhang et al., 2017). For TAMC to work, the IgG4 residue numbering was changed to be continuous for two segments, one segment corresponding to the first Fab region, its hinge and the Fc region, and the other segment to only the second Fab region and the hinge connected to this. A library of physically realistic glycosylated and deglycosylated structural conformations was generated by subjecting the starting structures to the TAMC module in SASSIE-web (Zhang et al., 2017). The flexible regions were assigned within the hinge, namely 216ESK218 and 228APE230 on one side of IgG4, and just 216ESK218 on the other side (green circles, Figure 5.1B, green text, Figure 5.2E). These tripeptides corresponded to surface-accessible structures outside the structurally-defined Fab and Fc regions and the disulphide-linked hinge core. These tripeptides could be structurally varied to create the required IgG4 conformers for testing against the scattering curve. The rest of the IgG4 structure was held rigid. Making 216ESK218 flexible on both sides of IgG4 rendered both Fab regions to be conformationally mobile, and making 231APE230 flexible made the Fc region mobile. For each of these nine linker residues, the backbone phi (φ) and psi (ψ) torsion angles were varied by 15°. In the Monte Carlo simulation,
many attempted moves will be physically unrealistic and were therefore discarded. For the glycosylated IgG4 simulations, 800,000 moves were attempted of which 111,382 were accepted. For the deglycosylated simulations, in which the glycan chains were omitted, 600,000 moves were attempted, of which 117,135 models were accepted.

For each of the 111,382 and 117,135 models, a scattering curve was generated using the SasCalc module in SASSIE-web. SasCalc calculates the scattering curve $I(Q)$ using an all-atom expression for the scattering intensity in which the orientations of the $Q$ vectors are taken from a quasi-uniform spherical grid generated by the golden ratio (Watson et al., 2013). For X-ray modelling, consideration of the hydration shell would require the explicit addition of a monolayer of water molecules to the protein surface before calculating $I(Q)$, and would require much computational effort as well as only affecting the scattering curve at larger $Q$ values (Watson et al., 2013). Thus, the hydration shell was not considered here for X-rays, and was not required for neutrons. These scattering curves were compared to the X-ray and neutron experimental scattering curves extrapolated to zero concentration, using the $R$-factor function in SASSIE-web.

This function calculates the difference between the modelled curve $I_{\text{Model}}(Q_i)$ and the interpolated experimental curves $I_{\text{Expt}}(Q_i)$, this function being analogous to that used in protein crystallography:

$$ R = \frac{\sum \| I_{\text{Expt}}(Q_i) - \eta I_{\text{Model}}(Q_i) \|}{\sum \| I_{\text{Expt}}(Q_i) \|} \times 100 $$

where $Q_i$ is the $Q$ value of the $i$th data point, $I_{\text{Expt}}(Q_i)$ is the experimental scattering intensity and $I_{\text{Model}}(Q_i)$ is the theoretically modelled scattering intensity, and $\eta$ is a scaling factor used to match the theoretical curve to the experimental $I(Q)$ (Wright et al., 2015). Lower $R$-factor values represent better fits. An iterative search to minimize the $R$-factor was used to determine $\eta$ (Wright et al., 2015). In the extrapolated experimental scattering curves, the lowest $Q$ values in the range before the fitted Guinier $R_g$ region were interpolated to zero $Q$ using MATLAB in order to satisfy the input requirement for the SasCalc module in SASSIE-web. Interpolation makes the $Q$ spacing uniform between the data points, and extrapolation extends the full $I(Q)$ curve to zero $Q$. The resulting 680 and 72 $I(Q)$
values in the $Q$ range of 0.0-1.5 nm$^{-1}$ were utilised for the X-ray and neutron curve fits respectively, and defined the $Q$ spacing for the SasCalc and the R-factor values. The use of $\chi^2$ analyses to evaluate the fits was not possible because this requires the experimental data points to have errors associated with them, which were not available. For the neutron curve fits, no correction was required for a flat incoherent background because the IgG4 concentrations were relatively low and the dialyses had sufficiently reduced the proton content in the buffers. The 111,382 glycosylated and 117,135 deglycosylated models gave an $R$-factor vs. $R_G$ distribution that encompassed the experimental extrapolated $R_G$ value. This $R$-factor analysis was repeated for four experimental X-ray scattering curves at different concentrations for each of glycosylated and deglycosylated IgG4 (Table 5.2). The same analysis was repeated for two neutron scattering curves at different concentrations, for each of glycosylated and deglycosylated IgG4 (Table 5.4). For each concentration, the best-fit 100 models with the smallest $R$-factors were accepted.

Principal component analysis (PCA) provided by the Bio3d package in R (Grant et al., 2006) was used to identify the main classes of best-fit IgG4 conformations found in the 800 best-fit glycosylated and deglycosylated models from eight X-ray scattering fits (Table 5.3). A separate analysis of the 400 best-fit models from four neutron scattering fits was performed (Table 5.5). To remove any bias in the PCA clustering of coordinate sets caused by the presence or absence of the glycans, the glycan coordinates were removed from the best-fit glycosylated models prior to generating the PCA. The X-ray and neutron models were assessed through two separate PCA. The mid-point structure for each PCA group was identified using a centroid model computed using R.

In order to model AUC parameters the theoretical $S_{20,w}$ values were generated for the best-fit 800 and 400 glycosylated and deglycosylated IgG4 models using HullRad (Fleming et al., 2018). Hullrad includes glycan residues for glycosylation, however there are inconsistencies in the PDB nomenclature for glycans. The nomenclature in the Hullrad script was thus modified to ensure that the IgG4 glycosylation was correctly incorporated in the $S_{20,w}$ calculation.
Table 5.2 Modelling fits for the SAXS and AUC data in light water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>$R_G$ before minimization</th>
<th>$R_G$ after minimization</th>
<th>$R_{XS-1}$</th>
<th>$R_{XS-2}$</th>
<th>$L$</th>
<th>$R$-factor before minimization</th>
<th>$R$-factor after minimization</th>
<th>$S_{20,w}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(nm)</td>
<td>(nm)</td>
<td>(nm)</td>
<td>(nm)</td>
<td>(nm)</td>
<td>(%)</td>
<td>(%)</td>
<td></td>
</tr>
<tr>
<td>Library of glycosylated models</td>
<td>111382</td>
<td>4.38-5.43</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.65-9.99</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 4.03 mg/ml</td>
<td>100</td>
<td>4.90-5.10</td>
<td>4.90-5.10</td>
<td>2.39-2.58</td>
<td>1.25-1.61</td>
<td>NA</td>
<td>0.89-1.57</td>
<td>0.81-1.55</td>
<td>6.68-6.89</td>
</tr>
<tr>
<td>Best fit at 4.03 mg/ml</td>
<td>1</td>
<td>4.96</td>
<td>4.97</td>
<td>2.54</td>
<td>1.5</td>
<td>15</td>
<td>0.89</td>
<td>0.86</td>
<td>6.73</td>
</tr>
<tr>
<td>Top 100 at 2.79 mg/ml</td>
<td>100</td>
<td>4.83-5.03</td>
<td>4.83-5.03</td>
<td>2.39-2.59</td>
<td>1.18-1.67</td>
<td>NA</td>
<td>0.69-1.53</td>
<td>0.61-1.52</td>
<td>6.69-6.91</td>
</tr>
<tr>
<td>Best fit at 2.79 mg/ml</td>
<td>1</td>
<td>4.93</td>
<td>4.93</td>
<td>2.5</td>
<td>1.36</td>
<td>15</td>
<td>0.69</td>
<td>0.61</td>
<td>6.72</td>
</tr>
<tr>
<td>Top 100 at 2.28 mg/ml</td>
<td>100</td>
<td>4.83-5.03</td>
<td>4.83-5.03</td>
<td>2.39-2.59</td>
<td>1.20-1.68</td>
<td>NA</td>
<td>0.65-1.47</td>
<td>0.59-1.48</td>
<td>6.69-6.91</td>
</tr>
<tr>
<td>Best fit at 2.28 mg/ml</td>
<td>1</td>
<td>4.93</td>
<td>4.93</td>
<td>2.5</td>
<td>1.35</td>
<td>15</td>
<td>0.65</td>
<td>0.59</td>
<td>6.72</td>
</tr>
<tr>
<td>Top 100 at 1.38 mg/ml</td>
<td>100</td>
<td>4.79-4.93</td>
<td>4.79-4.93</td>
<td>2.42-2.61</td>
<td>1.16-1.61</td>
<td>NA</td>
<td>0.75-1.47</td>
<td>0.68-1.46</td>
<td>6.72-6.99</td>
</tr>
<tr>
<td>Best fit at 1.38 mg/ml</td>
<td>1</td>
<td>4.84</td>
<td>4.84</td>
<td>2.53</td>
<td>1.4</td>
<td>15</td>
<td>0.75</td>
<td>0.68</td>
<td>6.85</td>
</tr>
<tr>
<td>Library of deglycosylated models</td>
<td>117135</td>
<td>4.07-5.47</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.53-14.15</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Top 100 at 3.07 mg/ml</td>
<td>100</td>
<td>4.90-5.08</td>
<td>4.90-5.08</td>
<td>2.42-2.61</td>
<td>1.19-1.56</td>
<td>NA</td>
<td>0.53-1.03</td>
<td>0.56-1.07</td>
<td>6.39-6.54</td>
</tr>
<tr>
<td>Best fit at 3.07 mg/ml</td>
<td>1</td>
<td>5.02</td>
<td>5.02</td>
<td>2.48</td>
<td>1.45</td>
<td>15</td>
<td>0.53</td>
<td>0.57</td>
<td>6.47</td>
</tr>
<tr>
<td>Top 100 at 1.93 mg/ml</td>
<td>100</td>
<td>4.88-5.03</td>
<td>4.88-5.03</td>
<td>2.41-2.61</td>
<td>1.17-1.59</td>
<td>NA</td>
<td>0.66-1.11</td>
<td>0.66-1.15</td>
<td>6.39-6.59</td>
</tr>
<tr>
<td>Best fit at 1.93 mg/ml</td>
<td>1</td>
<td>4.98</td>
<td>4.98</td>
<td>2.48</td>
<td>1.45</td>
<td>15</td>
<td>0.66</td>
<td>0.66</td>
<td>6.52</td>
</tr>
<tr>
<td>Top 100 at 1.21 mg/ml</td>
<td>100</td>
<td>4.83-4.96</td>
<td>4.83-4.96</td>
<td>2.43-2.57</td>
<td>1.21-1.52</td>
<td>NA</td>
<td>0.81-1.07</td>
<td>0.79-1.10</td>
<td>6.44-6.66</td>
</tr>
<tr>
<td>Best fit at 1.21 mg/ml</td>
<td>1</td>
<td>4.92</td>
<td>4.91</td>
<td>2.49</td>
<td>1.31</td>
<td>15</td>
<td>0.81</td>
<td>0.79</td>
<td>6.49</td>
</tr>
<tr>
<td>Top 100 at 0.85 mg/ml</td>
<td>100</td>
<td>4.73-4.86</td>
<td>4.73-4.86</td>
<td>2.41-2.56</td>
<td>1.15-1.47</td>
<td>NA</td>
<td>1.42-1.62</td>
<td>1.41-1.67</td>
<td>6.50-6.70</td>
</tr>
<tr>
<td>Best fit at 0.85 mg/ml</td>
<td>1</td>
<td>4.74</td>
<td>4.73</td>
<td>2.51</td>
<td>1.3</td>
<td>15</td>
<td>1.42</td>
<td>1.41</td>
<td>6.68</td>
</tr>
</tbody>
</table>
Table 5.3 Summary of Principal component analysis for fits of X-ray scattering and analytical ultracentrifugation data in light water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>$R_g$ before minimization</th>
<th>$R_g$ after minimization</th>
<th>$R_{XS-1}$</th>
<th>$R_{XS-2}$</th>
<th>$L$</th>
<th>$R$-factor before minimization</th>
<th>$R$-factor after minimization</th>
<th>$s_{20,w}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycosylated</td>
<td>41</td>
<td>4.84-5.10</td>
<td>4.84-5.10</td>
<td>2.40-2.58</td>
<td>1.16-1.52</td>
<td>NA</td>
<td>0.65-1.57</td>
<td>0.59-1.51</td>
<td>6.69-6.84</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>60</td>
<td>4.77-5.08</td>
<td>4.77-5.08</td>
<td>2.42-2.58</td>
<td>1.21-1.56</td>
<td>NA</td>
<td>0.70-1.62</td>
<td>0.71-1.65</td>
<td>6.41-6.69</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>4.91</td>
<td>4.91</td>
<td>2.55</td>
<td>1.4</td>
<td>15</td>
<td>1.1</td>
<td>1.03</td>
<td>6.54</td>
</tr>
<tr>
<td>PCA Group 2</td>
<td>342</td>
<td>4.73-5.08</td>
<td>4.73-5.08</td>
<td>2.37-2.58</td>
<td>1.18-1.68</td>
<td>NA</td>
<td>0.54-1.61</td>
<td>0.56-1.63</td>
<td>6.38-6.99</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>218</td>
<td>4.82-5.08</td>
<td>4.82-5.08</td>
<td>2.37-2.58</td>
<td>1.18-1.68</td>
<td>NA</td>
<td>0.74-1.56</td>
<td>0.69-1.52</td>
<td>6.68-6.99</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>124</td>
<td>4.73-5.07</td>
<td>4.73-5.07</td>
<td>2.41-2.56</td>
<td>1.21-1.56</td>
<td>NA</td>
<td>0.54-1.61</td>
<td>0.56-1.63</td>
<td>6.41-6.69</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>4.8</td>
<td>4.8</td>
<td>2.43</td>
<td>1.38</td>
<td>15</td>
<td>1.6</td>
<td>1.64</td>
<td>6.57</td>
</tr>
<tr>
<td>PCA Group 3</td>
<td>53</td>
<td>4.75-5.04</td>
<td>4.75-5.04</td>
<td>2.41-2.55</td>
<td>1.15-1.56</td>
<td>NA</td>
<td>0.73-1.62</td>
<td>0.71-1.67</td>
<td>6.45-6.77</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>1</td>
<td>4.97</td>
<td>4.97</td>
<td>2.54</td>
<td>1.53</td>
<td>NA</td>
<td>1.24</td>
<td>1.2</td>
<td>6.77</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>52</td>
<td>4.75-5.04</td>
<td>4.75-5.04</td>
<td>2.41-2.55</td>
<td>1.15-1.56</td>
<td>NA</td>
<td>0.73-1.62</td>
<td>0.71-1.67</td>
<td>6.45-6.70</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>4.8</td>
<td>4.8</td>
<td>2.42</td>
<td>1.33</td>
<td>15</td>
<td>1.6</td>
<td>1.63</td>
<td>6.68</td>
</tr>
<tr>
<td>PCA Group 4</td>
<td>304</td>
<td>4.77-5.09</td>
<td>4.77-5.09</td>
<td>2.41-2.61</td>
<td>1.17-1.59</td>
<td>NA</td>
<td>0.53-1.60</td>
<td>0.57-1.60</td>
<td>6.39-6.90</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>140</td>
<td>4.79-5.09</td>
<td>4.79-5.09</td>
<td>2.44-2.61</td>
<td>1.20-1.56</td>
<td>NA</td>
<td>0.88-1.53</td>
<td>0.88-1.42</td>
<td>6.72-6.90</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>164</td>
<td>4.77-5.06</td>
<td>4.77-5.06</td>
<td>2.41-2.61</td>
<td>1.17-1.59</td>
<td>NA</td>
<td>0.53-1.60</td>
<td>0.57-1.60</td>
<td>6.39-6.68</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>4.9</td>
<td>4.9</td>
<td>2.55</td>
<td>1.38</td>
<td>15</td>
<td>1.03</td>
<td>1</td>
<td>6.5</td>
</tr>
</tbody>
</table>
Table 5.4 Modelling fits for the neutron scattering and analytical ultracentrifugation data in heavy water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>$R_G$ before minimization (nm)</th>
<th>$R_G$ after minimization (nm)</th>
<th>$R_{XS-1}$ (nm)</th>
<th>$R_{XS-2}$ (nm)</th>
<th>$L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library of glycosylated models</td>
<td>111382</td>
<td>4.32-5.37</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.13-14.85</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 6.05 mg/ml</td>
<td>100</td>
<td>5.12-5.22</td>
<td>5.12-5.22</td>
<td>2.43-2.58</td>
<td>1.21-1.51</td>
<td>NA</td>
<td>2.30-2.90</td>
<td>2.35-2.90</td>
<td>6.53-6.67</td>
</tr>
<tr>
<td>Best fit at 6.05 mg/ml</td>
<td>1</td>
<td>5.2</td>
<td>5.2</td>
<td>2.51</td>
<td>1.45</td>
<td>15</td>
<td>2.3</td>
<td>2.35</td>
<td>6.63</td>
</tr>
<tr>
<td>Top 100 at 4.52 mg/ml</td>
<td>100</td>
<td>4.99-5.16</td>
<td>4.99-5.16</td>
<td>2.40-2.55</td>
<td>1.11-1.51</td>
<td>NA</td>
<td>1.34-1.85</td>
<td>1.36-1.90</td>
<td>6.60-6.78</td>
</tr>
<tr>
<td>Best fit at 4.52 mg/ml</td>
<td>1</td>
<td>5.07</td>
<td>5.07</td>
<td>2.46</td>
<td>1.36</td>
<td>15</td>
<td>1.34</td>
<td>1.36</td>
<td>6.71</td>
</tr>
<tr>
<td>Top 100 at 1.75 mg/ml</td>
<td>100</td>
<td>5.00-5.18</td>
<td>4.99-5.17</td>
<td>2.37-2.56</td>
<td>1.07-1.47</td>
<td>NA</td>
<td>1.76-2.30</td>
<td>1.80-2.33</td>
<td>6.59-6.76</td>
</tr>
<tr>
<td>Best fit at 1.75 mg/ml</td>
<td>1</td>
<td>5.14</td>
<td>5.13</td>
<td>2.43</td>
<td>1.31</td>
<td>15</td>
<td>1.76</td>
<td>1.8</td>
<td>6.73</td>
</tr>
<tr>
<td>Library of deglycosylated</td>
<td>117135</td>
<td>4.02-5.42</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2.06-19.47</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 100 at 5.62 mg/ml</td>
<td>100</td>
<td>5.13-5.29</td>
<td>5.13-5.28</td>
<td>2.48-2.63</td>
<td>1.10-1.53</td>
<td>NA</td>
<td>2.04-2.77</td>
<td>2.02-2.70</td>
<td>6.23-6.41</td>
</tr>
<tr>
<td>Best fit at 5.62 mg/ml</td>
<td>1</td>
<td>5.18</td>
<td>5.17</td>
<td>2.54</td>
<td>1.41</td>
<td>15</td>
<td>2.04</td>
<td>2.02</td>
<td>6.39</td>
</tr>
</tbody>
</table>
Table 5.5 Summary of Principal component analysis for fits of neutron scattering and analytical ultracentrifugation data in heavy water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>$R_g$ before minimization (nm)</th>
<th>$R_g$ after minimization (nm)</th>
<th>$R_{XS-1}$ (nm)</th>
<th>$R_{XS-2}$ (nm)</th>
<th>$L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$S_{20,w}$ (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycosylated</td>
<td>18</td>
<td>5.02-5.16</td>
<td>5.02-5.16</td>
<td>2.42-2.51</td>
<td>1.32-1.51</td>
<td>NA</td>
<td>1.61-1.83</td>
<td>1.59-1.87</td>
<td>6.60-6.73</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>50</td>
<td>5.16-5.29</td>
<td>5.15-5.28</td>
<td>2.48-2.62</td>
<td>1.29-1.53</td>
<td>NA</td>
<td>2.04-2.72</td>
<td>2.02-2.70</td>
<td>6.28-6.39</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.19</td>
<td>5.18</td>
<td>2.53</td>
<td>1.42</td>
<td>15</td>
<td>2.58</td>
<td>2.56</td>
<td>6.35</td>
</tr>
<tr>
<td>PCA Group 2</td>
<td>50</td>
<td>5.06-5.22</td>
<td>5.06-5.21</td>
<td>2.42-2.63</td>
<td>1.11-1.49</td>
<td>NA</td>
<td>1.34-1.85</td>
<td>1.36-1.90</td>
<td>6.23-6.75</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>32</td>
<td>5.06-5.15</td>
<td>5.06-5.16</td>
<td>2.42-2.52</td>
<td>1.28-1.47</td>
<td>NA</td>
<td>1.34-1.85</td>
<td>1.36-1.90</td>
<td>6.63-6.75</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>18</td>
<td>5.14-5.22</td>
<td>5.14-5.21</td>
<td>2.51-2.63</td>
<td>1.11-1.49</td>
<td>NA</td>
<td>2.35-2.70</td>
<td>2.36-2.70</td>
<td>6.23-6.38</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.09</td>
<td>5.09</td>
<td>2.49</td>
<td>1.39</td>
<td>15</td>
<td>1.77</td>
<td>1.76</td>
<td>6.68</td>
</tr>
<tr>
<td>PCA Group 3</td>
<td>30</td>
<td>4.99-5.09</td>
<td>4.99-5.10</td>
<td>2.40-2.54</td>
<td>1.26-1.46</td>
<td>NA</td>
<td>1.60-1.85</td>
<td>1.52-1.86</td>
<td>6.35-6.78</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>30</td>
<td>4.99-5.09</td>
<td>4.99-5.10</td>
<td>2.40-2.54</td>
<td>1.26-1.46</td>
<td>NA</td>
<td>1.60-1.85</td>
<td>1.52-1.86</td>
<td>6.65-6.78</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.02</td>
<td>5.03</td>
<td>2.54</td>
<td>1.34</td>
<td>15</td>
<td>1.82</td>
<td>1.8</td>
<td>6.72</td>
</tr>
<tr>
<td>PCA Group 4</td>
<td>52</td>
<td>5.03-5.24</td>
<td>5.04-5.24</td>
<td>2.42-2.63</td>
<td>1.10-1.49</td>
<td>NA</td>
<td>1.59-2.68</td>
<td>1.59-2.65</td>
<td>6.28-6.74</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>20</td>
<td>5.03-5.13</td>
<td>5.04-5.14</td>
<td>2.42-2.55</td>
<td>1.11-1.43</td>
<td>NA</td>
<td>1.59-1.82</td>
<td>1.59-1.81</td>
<td>6.66-6.74</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>32</td>
<td>5.13-5.24</td>
<td>5.13-5.24</td>
<td>2.51-2.63</td>
<td>1.10-1.49</td>
<td>NA</td>
<td>2.09-2.68</td>
<td>2.06-2.65</td>
<td>6.28-6.41</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>5.16</td>
<td>5.15</td>
<td>2.59</td>
<td>1.19</td>
<td>15</td>
<td>2.1</td>
<td>2.06</td>
<td>6.65</td>
</tr>
</tbody>
</table>
5.3 Results

5.3.1 Purification and characterisation of glycosylated and deglycosylated IgG4

A protocol for the deglycosylation of the monoclonal human IgG4 A33 antibody was set up using peptide:N-glycosidase F (PNGase F) digests according to the manufacturer's protocol. The completeness of deglycosylation was verified by a three-fold combination of routine gel filtration, SDS-PAGE and mass spectrometry, and also by analytical ultracentrifugation (see below):

(i) At the timepoints of one hour, six hours and ten hours after the start of the digests (TP1, TP6 and TP10), the IgG4-digested products eluted from a gel filtration column slightly prior to that for glycosylated IgG4 (Figure 5.3A). The proteins eluted as a main symmetrical peak at 17.88 ml, 17.84 ml, 17.82 ml and 17.86 ml for glycosylated IgG4, and deglycosylated IgG4 at timepoints TP1, TP6 and TP10, respectively (Figure 5.3A). This procedure ensured that the IgG4 sample was monodisperse with no aggregates present immediately before analytical ultracentrifugation or scattering experiments.

(ii) When the IgG4 samples were submitted to non-reducing and reducing SDS-PAGE analyses at equimolar concentrations, purified glycosylated and deglycosylated IgG4 showed a single band between 200 and 116 kDa on 4-12% Bis Tris NuPage gel under non-reducing conditions, which is consistent with the expected masses of ~147 kDa for IgG4 (Figure 5.3B). Under reducing conditions two bands were present corresponding to the heavy chain (with an apparent mass of ~55 kDa) and the light chain (with an apparent molecular mass of between 31 and 21 kDa) (Figure 5.3B). These apparent molecular masses were as expected from the known sequence.

(iii) Liquid chromatography mass spectrometry measurements showed multiple peaks for glycosylated IgG4 (G) that were assigned to the presence of at least five glycoforms, separated by masses of 157-168 Da that corresponded to single sugar residues (Figure 5.3C) (Hui et al., 2019). The most intense IgG4 glycosylated population had an observed deconvoluted mass of 146,848 Da.

After an hour of digest, a partially deglycosylated IgG4 (P) was observed in which the number of glycoforms was diminished, and an additional peak was
Figure 5.3 Purification, SDS-PAGE and mass spectrometry of human glycosylated and deglycosylated IgG4.

A Elution peaks from a Superose 6 Increase 10/300 gel filtration column for four IgG4 samples, these being glycosylated (black) and from the TP1 (blue), TP6 (red) and TP10 (magenta) timepoints. The dashed vertical lines indicate the peak positions. B Lane 1 and 6, molecular mass markers are denoted in kDa. Lanes 2-5, non-reduced SDS-PAGE of glycosylated IgG4, TP1, TP6 and TP10 after gel filtration. 7-10, reduced SDS-PAGE of glycosylated IgG4, TP1, TP6 and TP10 after gel filtration. C Mass spectra of glycosylated and deglycosylated IgG4, using the same colour scheme as in A. Peaks labelled G represents glycosylated species, P represents partially glycosylated species, and D represents fully deglycosylated species.
observed at 145,240 Da. After six or ten hours, only the single dominant deglycosylated peak (D) was seen at 143,636 Da.

The peak at 145,240 kDa was attributed to IgG4 in which one of the two glycans at Asn297 was not present. The mass of each glycan chain was calculated by subtracting the glycosylated and deglycosylated masses and halving the outcome to give 1606 Da. This glycan mass was also calculated by subtracting the glycosylated and half-deglycosylated masses to give 1608 Da. These mass values agreed well with an assumed glycan composition of Gal$_3$Man$_3$GlcNAc$_4$ that gives a mass of 1622 Da.

### 5.3.2 Analytical ultracentrifugation of glycosylated and deglycosylated IgG4

Sedimentation velocity experiments investigated the mass and solution of glycosylated and deglycosylated IgG4 at the TP1, TP6 and TP10 timepoints. The SEDFIT analyses involved fits of as many as 900 scans, and the good agreement between the experimental boundary scans and fitted lines is clear (left, Figure 5.4A). In the resulting size distribution analyses $c(s)$, a monomer peak that monitored the overall IgG4 solution structure was observed at average $s_{20,w}^{0}$ values of 6.52 S for glycosylated IgG4, 6.30 S for IgG4 TP1, 6.29 S for IgG4 TP6 and 6.31 S for IgG4 TP10 (right panel, Figure 5.4A). For the glycosylated forms, these values were consistent with those of 6.44 S, 6.80 S and 6.60 S for IgG4 (Ser) (Abe et al., 2010), IgG4 (Ser$^{222}$) and IgG4 (Pro$^{222}$) respectively (Rayner et al., 2014) and previously reported (Perkins, 1986; Brady et al., 1992; Davies et al., 2014). From the $c(s)$ analyses, the molecular masses of the IgG4 monomer peak were 146 kDa (glycosylated), 146 kDa (TP1), 148 kDa (TP6) and 149 kDa (TP10). These values agree well with the composition-calculated masses of 148.1 kDa and 143.1 kDa for the glycosylated and deglycosylated monomers respectively. These also agree well with the values from mass spectrometry of 146,848 Da and 143,636Da for glycosylated and deglycosylated IgG4 (Figure 5.3C). The 0.2 S reduction (3%) in the $s_{20,w}$ values on deglycosylation is attributable to the 4 kDa reduction (3%) in the IgG4 mass, according to the Svedberg equation in which $s_{20,w}$ is proportional to the mass.
A The experimentally observed sedimentation boundaries for a concentration series of glycosylated IgG4 and likewise for deglycosylated IgG4 at the TP1 (blue), TP6 (red) and TP10 (magenta) timepoints in histidine buffer. Scans were recorded at 30,000 rpm and 20°C, from which 31-66 boundaries (black outlines) are shown from totals of up to 900 scans. The SEDFIT fits are shown as white lines. The peaks in the corresponding size distribution analyses c(s) revealed a monomer peak (M) at $s^{0}_{20,w}$ values of 6.03-6.55 S for glycosylated IgG4 and the three deglycosylation timepoints TP1, TP6 and TP10. B The $s^{0}_{20,w}$ values for the monomer peaks are shown as a function of concentration for glycosylated IgG4 (●), and the TP1 (●), TP6 (●) and TP10 (●) timepoints.

Figure 5.4 Sedimentation velocity analyses of glycosylated and deglycosylated IgG4.
divided by the frictional coefficient. This calculation assumes that the IgG4 conformation (i.e. the frictional coefficient) is unchanged after deglycosylation. In the 2014 study, minor peaks for dimers were visible at about 9 S for IgG4 (Ser^{222}) and (Pro^{222}) \cite{Rayner2014}. In this work, no dimer peaks were visible for IgG4 A33 (right panel, Figure 5.4A). This difference may result from the use of histidine buffer in this current study, in distinction to the phosphate buffer saline used before. If IgG4 A33 did form dimers, the histidine buffer may have increased the exchange rate between monomer and dimer such that separate monomer and dimer peaks would no longer be seen.

### 5.3.3 X-ray and neutron scattering of glycosylated and deglycosylated IgG4

The solution structures of glycosylated and deglycosylated IgG4 at the TP1, TP6 and TP10 time points were analysed by both X-ray and neutron scattering. The two methods provided different perspectives of the same solution structure. X-rays in light water buffers detect the hydration shell surrounding the protein structure, whereas the effect of the hydration shell is reduced by neutrons in heavy water buffers for reasons of the different solute-solvent contrast in use \cite{Perkins1986, Perkins2001, Svergun1998}. The IgG4 X-ray data collection at concentrations between 0.85-4.03 mg/ml used time frame analyses to ensure the absence of radiation damage effects. The resulting $R_G$ and $R_{XS-1}/R_{XS-2}$ values monitor of the elongation of the overall IgG4 structure and its approximate cross-sectional structures respectively.

Guinier analyses resulted in high quality linear plots for all samples and revealed three regions of the $I(Q)$ curves, as expected for antibodies \cite{Pilz1970, Rayner2014, Rayner2015}. From these, the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values from the individual scattering curves were obtained within satisfactory $Q.R_G$ and $Q.R_{XS}$ limits of 0.5-1.4, 0.7-1.3, 0.9-1.5 (Figure 5.5A). A slight concentration dependence was observed from the X-ray $I(0)/c$ values, values that suggested a small amount of oligomer formation in the concentration series (Figure 5.6A; Table 5.1). After extrapolation to zero concentration, the X-ray $R_G$ values that monitor the overall structure for glycosylated IgG4, and the deglycosylated TP1, TP6 and TP10 samples were $4.92 \pm 0.19$ nm, $4.91 \pm 0.27$ nm, $4.91 \pm 0.17$ nm
and 4.91 ± 0.13 nm, respectively. The present X-ray $R_G$ values for glycosylated IgG4 A33 agreed well with previous $R_G$ values of 5.07 – 4.82 nm for two other human monoclonal IgG4 Ser$^{222}$ and IgG4 Pro$^{222}$ antibodies. The $RXS-1$ values from the individual curves (Figure 5.6A) are an approximate monitor of the cross-sectional structure for glycosylated and deglycosylated IgG4. These were extrapolated to zero concentration to show that these were essentially unchanged from 2.50 ± 0.01 nm to 2.49 ± 0.10 nm, respectively, indicating minimal rearrangement of the Fab and Fc. The $RXS-2$ values from the individual curves for IgG4 glycosylated and deglycosylated TP1, TP6 and TP10 timepoints were extrapolated to zero concentration to show that these were slightly decreased from 1.41 ± 0.04 nm to 1.38 ± 0.15 nm, 1.39 ± 0.05 nm and 1.37 ± 0.10 nm respectively, suggesting that the Fc region became more compact upon removal of the glycan chains.

The corresponding neutron scattering data sets for glycosylated and deglycosylated IgG4 (TP1, TP6, TP10) in 100% $^2$H$_2$O buffer were analysed between 0.71-9.92 mg/ml, this concentration range being similar to that used above for SAXS. The Guinier analyses revealed high quality linear fits for the same three $R_G$, $RXS-1$ and $RXS-2$ parameters as for X-rays (Figure 5.5B). The mean neutron $R_G$ values for glycosylated IgG4 and deglycosylated TP1, TP6 and TP10 were 5.21 ± 0.09 nm, 5.26 ± 0.07 nm, 5.10 ± 0.02 nm and 5.35 nm (only one measurement was acquired for this time point) respectively. A concentration dependence was observed in the neutron $R_G$ data for IgG4, however this was not reproducible in the other parameters $I(0)/c$, $RXS-1$ and $RXS-2$. It is possible that the higher concentrations used for glycosylated IgG4 (average concentration 4.60 mg/ml), compared to an average of 1.32 mg/ml for TP1, 1.78 mg/ml for TP6 and 5.62 mg/ml for TP10, gave this outcome, but again the reduced number of data points limited the precision of the data sets. Fewer concentrations were used in SANS, compared to SAXS, due to the nature of a neutron scattering experiment where data collection is longer (minimum of one hour compared to one minute for X-rays) and the volumes required were far greater (800 µl compared to 25 µl for X-rays). The mean neutron $RXS-1$ values for IgG4 glycosylated and deglycosylated TP1, TP6 and TP10 were 2.43 ± 0.02 nm, 2.46 ± 0.02 nm, 2.48 ± 0.02 nm and 2.48 nm, respectively. The mean neutron $RXS-2$ values for IgG4 glycosylated and deglycosylated TP1, TP6 and TP10 were 1.15 ± 0.03 nm, 1.18
± 0.04 nm, 1.23 ± 0.02 nm and 1.20 nm, respectively. The neutron values confirmed the X-ray analyses.

The distance distribution function $P(r)$ was derived from Fourier transformation of the scattering curve $I(Q)$, and provided structural information in real space for glycosylated and deglycosylated IgG4. The X-ray and neutron $P(r)$ analyses gave $R_g$ values that were similar to those from the X-ray Guinier analyses, showing that the two analyses were self-consistent (open symbols, Figure 5.6). The maximum lengths of glycosylated and deglycosylated IgG4 were determined from the value of $r$ when the X-ray $P(r)$ curve intersected zero on the $r$ axis and was 15 nm for all four IgG4 samples (Figure 5.7A). The maximum lengths of glycosylated and deglycosylated IgG4, were similar but slightly increased to 16 nm for all four neutron samples (Figure 5.7B).

The maxima in the $P(r)$ curves corresponded to the most frequently occurring atomic distances between scattering elements within the structures, these being a monitor of the IgG4 structure. For all IgG4 samples, two peaks, $M1$ and $M2$, were visible that were characteristic of antibody-shaped proteins. $M1$ corresponded predominantly to the shorter distances within each Fab and Fc region, and was expected to be almost invariant because these regions should not change shape. $M2$ corresponded primarily to the longer distances between pairs of Fab and Fc regions, and monitored changes in the relative separation of the Fab and Fc regions (Figures 5.1 and 5.7). No concentration dependence was observed in the positions of the $M1$ and $M2$ peaks, which were measured directly from their maximum values (Figure 5.7C, D). No differences in the $M1$ and $M2$ positions were observed between glycosylated and deglycosylated IgG4. For the X-ray data for IgG4, the $M1$ peaks were observed at 4.54 ± 0.02 nm for glycosylated IgG4 and 4.53 ± 0.01 nm for deglycosylated IgG4. The X-ray $M2$ peaks were observed at 7.29 ± 0.07 nm for glycosylated IgG4 and 7.27 ± 0.05 nm for deglycosylated IgG4. The neutron $P(r)$ analyses showed that the $M1$ peak was at 4.38 ± 0.08 nm for glycosylated IgG4 and 4.29 ± 0.03 nm for deglycosylated IgG4, and the $M2$ peak was at 7.23 ± 0.20 nm for glycosylated IgG4 and 7.17 ± 0.12 nm for deglycosylated IgG4. Both the X-ray and neutron analyses were consistent with each other.
Figure 5.5 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG4 (legend overleaf).
Figure 5.5 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG4 (continued).

A The SAXS curves for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG4 at concentrations of 0.85-4.70 mg/ml. The filled circles between the arrows represent the $Q.R_G$ and $Q.R_{XS}$ fit ranges used to determine the $R_G$ and $R_{XS}$ values. The $Q$ range used for the $R_G$ values was 0.10-0.27 nm$^{-1}$; those for the $R_{XS-1}$ and $R_{XS-2}$ values were 0.29-0.52 nm$^{-1}$ and 0.66-1.05 nm$^{-1}$ respectively.

B The SANS curves for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG4 at concentrations of 0.71-6.05 mg/ml. The $Q$ range used for the $R_G$ values was 0.07-0.27 nm$^{-1}$ and those for the $R_{XS-1}$ and $R_{XS-2}$ values were 0.28-0.52 nm$^{-1}$ and 0.66-1.05 nm$^{-1}$ respectively.
The filled symbols show the values determined from the Guinier analyses and the open symbols in the $R_G$ panels indicate those determined from the $P(r)$ analyses. The colours denote the glycosylated IgG4 (black), TP1 (blue), TP6 (red) and TP10 (magenta) timepoints.

A The SAXS $R_G$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for glycosylated (●, ○) and deglycosylated TP1 (●, ○), TP6 (●, ○) and TP10 (●, ○). The solid lines corresponded to linear regression fits of glycosylated IgG4, and the dashed lines to the fits for deglycosylated IgG4.

B The SANS $R_G$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG4, each corresponding to a single measurement in histidine buffer in $^2$H$_2$O. The solid and dashed lines correspond to the mean values for glycosylated and deglycosylated IgG4.

Figure 5.6 Concentration dependence of the SAXS and SANS Guinier analyses.
5.3.4 Atomistic modelling of glycosylated and deglycosylated IgG4

The scattering curves of IgG4 before and after glycan removal were examined for conformational differences using atomistic modelling simulations of the two structures were performed. This was initiated from high-resolution crystal structures for the human Fab and Fc regions. The Fab amino acid sequence in the structure was converted into that for IgG4 A33 using Modeller (Figure 5.2A-D). The Fab and Fc regions were joined by a peptide 216ESKYGPPCPAPCPAPEFLGGP235 that included the 12-residue hinge sequence (Figure 5.2E), which was built using PyMOL. The glycosylated IgG4 models were created by adding complex-type biantennary glycans to the two Asn297 sidechains in the Fc region (Figure 5.1B). This starting structure was subjected to energy minimisation.

Physically-realistic IgG4 models without steric clashes were created for comparison with the experimental X-ray curves. By varying the torsion angles at three flexible regions at the start and end of the two IgG4 hinges (green circles, Figure 5.1), trial IgG4 structures were created that involved movements of the two Fab and one Fc regions relative to each other. For glycosylated IgG4, 800,000 models were generated in eight Monte Carlo simulations, from which 111,382 models were accepted because these showed no steric clashes between separate residues in the model. For deglycosylated IgG4, 600,000 models were generated in six Monte Carlo simulations, of which 117,135 models were likewise acceptable. To ensure that no systematic trends were overlooked in the modelling outcome, four X-ray and up to three neutron scattering curves for up to four concentrations were fitted for each of the four samples in question (Figure 5.8). For both glycosylated and deglycosylated IgG4, comparison of the four experimental X-ray scattering curves at 0.85-4.03 mg/ml with the 111,382 and 117,135 theoretical curves gave a goodness-of-fit $R$-factor vs $R_G$ distribution with clear minima in all eight cases (Figure 5.8A, B). The minima agreed with the experimental $R_G$ values (Figure 5.6A). The minima showed that enough trial X-ray models had been generated to result in good fits in each case. Filtering of the models to select these with the lowest $R$-factors gave the 100 best-fit models for each concentration (red, Figure 5.8). The range of the 100 $R$-factors for each of
Figure 5.7 SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG4 (legend overleaf).
Figure 5.7 SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG4 (continued).

The colours denote the glycosylated IgG4 (black), TP1 (blue), TP6 (red) and TP10 (magenta) timepoints. A The peak maxima at $M1$ and $M2$ and the maximum length $L$ are indicated by arrows. The SAXS and SANS $P(r)$ curves for glycosylated and deglycosylated (TP1, TP6 and TP10) IgG4 are shown at concentrations of 0.85-4.70 mg/ml. B The corresponding $P(r)$ curves for the SANS curves for IgG4 0.71-6.05 mg/ml. C, D The concentration dependence of the peak maxima $M1$ and $M2$ for glycosylated (●) and deglycosylated TP1 (●), TP6 (●) and TP10 (●) are shown. The fitted lines are the mean values for glycosylated IgG4 (solid line), and for TP1, TP6 and TP10 averaged together (dashed lines).
the four concentrations was low at between 0.65-1.57% for the best-fit glycosylated models and 0.53-1.62% for the best-fit deglycosylated models (Table 5.2). This indicated good quality X-ray curve fits between the experimental and modelled curves.

The eight sets of 100 best-fit models (Figure 5.8) were examined in order to identify the resulting best-fit IgG4 conformations from the curve fits. For this, PCA were performed (David et al., 2014). The PCA determines the correlated motions of protein residues as linearly uncorrelated variables, each being termed a principal component (David et al., 2014). These “essential motions” were extracted from a covariance matrix of the atomic coordinates of the frames in the selected structure set. The eigenvectors of this matrix each have an associated eigenvalue that characterises the clustering of the models based on structural coordinates (or variance). In order to eliminate bias in the PCA, the glycan chains were removed from the glycosylated IgG4 models before comparison with the deglycosylated models. Interestingly, despite the lack of clear change in the Guinier or P(r) analyses, PCA indicated differences between the glycosylated and deglycosylated best-fit X-ray IgG4 models (black and magenta respectively, Figure 5.9A-D; Table 5.3). Thus, the distributions of the best-fit 400 glycosylated and 400 deglycosylated X-ray models were each clustered into four distinct Groups 1-4, with some overlap between glycosylated and deglycosylated groups. The glycosylated models mostly occurred in the PCA Groups 2, while the deglycosylated models mostly occurred in the PCA Groups 1 and 3, with significant overlap of glycosylated and deglycosylated models in Group 4. The visually-excellent X-ray curve fits confirmed the validity of the modelling fits (Figures 5.10A, B). Of particular note was the agreement of the experimental and theoretical double peaks in the P(r) curves shown as insets.

The same outcome was found with the neutron modelling, thus confirming the reproducibility of the curve fits, although the precision of the neutron scattering curves was reduced compared to that of the X-ray fits. The same 111,382 and 117,135 theoretical curves were compared with the neutron scattering curves at 1.75-6.05 mg/ml to show again that 100 best-fit structures could be identified at clear minima in each of the R-factor vs \( R_G \) neutron distributions (Figures 5.8C, D). The neutron PCA also indicated clear differences between the glycosylated and deglycosylated neutron IgG4 models (black and
The 111,382 models that were accepted for glycosylated IgG4 and the 117,135 models that were accepted for deglycosylated IgG4 are shown as circles. A, B Four experimental X-ray scattering curves for four concentrations of glycosylated and deglycosylated IgG4 are shown. C, D two experimental neutron scattering curves for two concentrations of glycosylated and deglycosylated IgG4 are shown. These experimental curves were fitted to the 111,382 and 117,135 modelled curves. Those models with $R_G$ values closest to the experimental $R_G$ values showed the lowest goodness-of-fit $R$-factor as expected. The top 100 best-fit models (red circles) showed the lowest goodness-of-fit $R$-factors. The experimental $R_G$ is represented by a solid vertical line and the dashed vertical lines represent the ± 2% upper and lower boundaries of these $R_G$ values.

Figure 5.8 Atomistic modelling analyses for glycosylated and deglycosylated IgG4.
Figure 5.9 Principal component analysis (PCA) of the best-fit glycosylated and deglycosylated IgG4 models (legend overleaf).
Figure 5.9. Principal component analysis (PCA) of the best-fit glycosylated and deglycosylated IgG4 models (continued). Glycosylated models are represented in black and deglycosylated models are represented in magenta. In this, groups 1, 2, 3 and 4 are represented by ○, Δ, +, and × in that order, and the centroid model for each group is represented by large numbers (blue) and a ★.

A-D The eight sets of 100 best-fit models from the experimental X-ray scattering curves for glycosylated and deglycosylated IgG4 were grouped by PCA into four groups as shown in three panels A, B and C of PC2 vs PC1, PC3 vs PC2 and PC3 vs PC1. D, The first three eigenvalue rankings (PC1, PC2 and PC3) captured 50.1% of the variance in the 800 models. (E-H) The two sets of 100 best-fit models from the experimental neutron scattering curves for glycosylated and deglycosylated IgG4 were grouped by PCA into four groups as shown in E, F and G of PC2 vs PC1, PC3 vs PC2 and PC3 vs PC1. H, The first three eigenvalue rankings (PC1, PC2 and PC3) captured 55.8% of the variance in the 200 models.
Figure 5.10 Scattering curve fits to the experimental data for the best-fit model for each of the glycosylated and deglycosylated IgG4 samples (legend overleaf).
FIGURE 5.10. Scattering curve fits to the experimental data for the best-fit model for each of the glycosylated and deglycosylated IgG4 samples (continued).

The experimental curve is denoted by black circles and the best-fit modelled curves are denoted by solid red lines. The distance distribution curves \( P(r) \) are shown in the top right of each panel. **A** glycosylated and **B** deglycosylated IgG4 X-ray scattering curve fits for four concentrations each. For the four X-ray fits in **A**, the glycosylated IgG4 models were taken from PCA group 1 (2.79 mg/ml, 2.28 mg/ml) and group 2 (4.03 mg/ml and 1.33 mg/ml) in that order (Table 5.3). In **B**, the deglycosylated IgG4 models used for the fits as shown corresponded to PCA group 2 (0.85 mg/ml) and group 4 (3.07 mg/ml, 1.93 mg/ml and 1.21 mg/ml in that order). **C** glycosylated and **D** deglycosylated IgG4 neutron curve fits shown for three and one concentrations respectively. In the neutron fits, the glycosylated IgG4 models corresponded to PCA group 2 (4.25 mg/ml) (Table 5.5). Likewise, the deglycosylated IgG4 models corresponded to PCA group 1 (5.62 mg/ml).
Figure 5.11 Normalised Kratky plots for the experimental and best fit glycosylated and deglycosylated IgG4 scattering curves.

A X-ray experimental data (solid lines) and model fits (dashed lines) were shown in black for glycosylated IgG4 at 4.03 mg/ml and in magenta for deglycosylated IgG4 at 3.07 mg/ml. B Neutron experimental data (solid lines) and model fits (dashed lines) were shown in black for glycosylated IgG4 at 4.52 mg/ml and in magenta for deglycosylated IgG4 at 5.62 mg/ml.
Figure 5.12 Views of representative best-fit X-ray and neutron structures.

The blue cartoon denotes the protein backbone of the starting glycosylated and deglycosylated IgG4 structures. The Fab regions of the 100 best-fit models were superimposed onto these starting structures, in order to focus on movements in the Fc region. The blue and magenta wireframe envelopes denote the overall space occupied by the glycosylated and deglycosylated Fc regions, respectively, in the 100 best-fit structures for each of the A four X-ray and B one neutron analyses. In all of the representations, the magenta wireframes (deglycosylated IgG4) occupy a greater region of space compared to the blue wireframes (glycosylated IgG4). C The cartoon representation based on part (c) of A showed the larger range of Fc conformations as arrowed (magenta) after deglycosylation compared to that for glycosylated IgG4 (blue).
magenta respectively, Figure 5.9E-H; Table 5.5). The distributions of the best-fit 100 glycosylated and 100 deglycosylated neutron models (4.52 mg/ml in Figure 5.8C, D) were each again clustered into four PCA groups, with some overlap between glycosylated and deglycosylated groups. The glycosylated models were mostly in PCA Groups 2 and 3, while the deglycosylated models were mostly in PCA Groups 1 and 4. Visually-excellent neutron curve fits were obtained (Figure 5.10C, D).

Given the lack of change in the Guinier or $P(r)$ analyses following deglycosylation, further insights into the experimental data and their modelling were obtained from Kratky analyses of $(Q.R_G)^2.I(Q)/I(0)$ vs $Q.R_G$ for the experimental scattering curves at the highest concentrations in use and the scattering curves from the modelled best fit structures. These plots indicate whether the macromolecule in question is globular in its structure or possesses intrinsically disordered regions (Receiveur-Brechot et al., 2012). The Kratky plots demonstrated two clear peaks in both the experimental and theoretical modelled curves. For the X-ray Kratky curves (Figure 5.11A), the $Q.R_G$ values for the experimental glycosylated peaks were 2.02 and 3.95, in good accord with the modelled values of 1.89 and 3.92. The $Q.R_G$ values for the experimental deglycosylated peaks of 2.01 and 4.06 were also in good accord with the modelled deglycosylated peaks of 2.01 and 4.18. It was interesting to note that the second peak showed higher intensities for deglycosylated IgG4 (magenta) than for glycosylated IgG4 (black), suggesting that there was a small increase in antibody disorder after deglycosylation. For the neutron Kratky curves (Figure 5.11B), the $Q.R_G$ values for the experimental peaks for glycosylated IgG4 were 2.06 and 4.22, which were similar to the modelled peaks at 1.95 and 4.11. The $Q.R_G$ values for the experimental peaks for deglycosylated IgG4 were 2.15 and 4.52, but showed less agreement for the modelled peaks at 2.03 and 4.18. Given that the second neutron peak showed higher intensities for deglycosylated IgG4 (magenta) when compared with glycosylated IgG4 (black), the experimental data suggested that greater disorder was present in IgG4 after deglycosylation.

As another test of the scattering modelling, the $s^2_{20,w}$ values for the eight sets of best-fit 100 glycosylated and deglycosylated models from each X-ray concentration (Figures 5.8, 5.10) were calculated using HullRad (Fleming et al., 2018). This gave an $s^2_{20,w}$ range of 6.68-6.99 S for the four X-ray concentrations.
for glycosylated IgG4 and 6.39-6.70 S for deglycosylated IgG4 (Table 5.2). These values agreed well with the experimental $s_{20,w}^{d}$ values of 6.54-6.55 S for glycosylated IgG4 and 6.03-6.33 S for deglycosylated IgG4 (Table 5.1). These agreements corroborated the outcome of the atomistic scattering modelling, given that the mean difference between the modelled and experimental values should typically be ± 0.21 S for related macromolecules (Perkins et al., 2009). This $s_{20,w}^{d}$ modelling was however unable to distinguish changes before and after deglycosylation.

5.4 Discussion

The present scattering and atomistic modelling study investigated the structural importance of the C$_{H2}$ glycans in IgG4, thus providing functional insight on their role. Compared to several earlier crystallographic studies on just the IgG4 Fc region alone, the joint SAXS-SANS-AUC-MC approach provided an informative outcome on full-sized IgG4 of the structural changes accompanying the removal of the two conserved Asn$^{297}$ glycans in the Fc region (Figure 5.1B). Deglycosylation was directly confirmed using a combination of gel filtration, mass spectrometry, and AUC. The AUC sedimentation data showed that IgG4 samples were monomeric and that the $s_{20,w}^{d}$ values decreased to reflect the decreased molecular mass of IgG4 following glycan removal. Subsequently, extensive SAXS and SANS data collection on glycosylated and deglycosylated IgG4 established their Guinier $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values, and their distance distribution curves $P(r)$. While a slight concentration dependence was observed in the $R_G$ parameters, interestingly no differences in these were initially observed between glycosylated and deglycosylated IgG4. This is in contrast to IgG1 studies where clear differences in the $R_{XS-1}$ and $M2$ parameters were observed corresponding to an increase in the averaged Fab-Fc separations after deglycosylation (Spiteri et al., 2021b). The experimental Kratky plots suggested a greater disorder in IgG4 following its deglycosylation (Figure 5.11), and this represented the main experimental difference seen between the two forms of IgG4.

The use of atomistic scattering modelling using Monte Carlo simulations of 111,382-117,135 stereochemically-correct IgG4 conformations clarified the
structural importance of the C_{H2} glycans in IgG4. The interpretation of the output was much facilitated by PCA analyses of the 100 best-fit structures. While some overlap between the glycosylated and deglycosylated models was seen, the two sets of PCA structures clustered into distinct groups (Figure 5.9). The clearest view of the final result was presented in wireframe representations of the 100 best fit structures (Figure 5.12). Views of the 100 best-fit X-ray models at four concentrations showed that the glycosylated Fc region occupied a smaller volume (blue wireframe, Figure 5.12A), than a larger one for deglycosylated Fc (magenta wireframe, Figure 5.12A). The cartoon view (Figure 5.12C) showed that the deglycosylated Fc region of IgG4 occupied more conformational space than the glycosylated one. In the deglycosylation study of IgG1, the conformational flexibility of deglycosylated IgG1 (Spiteri et al., 2021b) was also significantly greater than of deglycosylated IgG4. This outcome implies that, by the X-ray modelling, the shorter length of the IgG4 hinge results in a reduced level of conformational flexibility compared to IgG1, thus the differences observed between glycosylated and deglycosylated IgG4 were smaller than those for IgG1. The 100 best-fit IgG4 neutron models were limited by the reduced precision of the scattering data but are consistent with this interpretation (Figure 5.12B). Overall, the molecular simulations were able to explain the structural differences seen before and after deglycosylation.

In relation to function, the advantage of this combined SAXS-SANS-AUC-MC approach is the ability to address the full IgG4 structure in solution, this being the functional native structure. Previous studies of the IgG1-Fc and IgG4-Fc regions indicate that the Fc glycans makes no or very little contact with the FcγR receptors that mediate the effector function of IgG antibodies (Watson et al., 2013; Davies et al., 2014; Wright et al., 2015, 2019). Functional studies of these IgG-FcγR complexes are limited by their crystal structures only being available for IgG1-Fc complexes, there being no available structures of IgG4-Fc in complex with the FcγRs. Functional studies of deglycosylated IgG4 demonstrated abrogated binding to FcγRIIIa, which implicated the glycans to assist in the FcγR interaction (Kang et al., 2020). The lower binding affinity of IgG4 is explained by the main finding in the present study that the deglycosylated IgG4-Fc region is more conformationally labile than the seemingly more restricted glycosylated IgG4-Fc region (Spiteri et al., 2021b). In support of this outcome, crystallography
studies of the deglycosylated Fc region showed that its FG loop, which is vital in FcγR and C1q binding, can adopt two distinct conformations that would reduce their binding interactions (Davies et al., 2014). As a different perspective altogether, one previous structural study of the full-length IgG4 antibody suggested that the IgG4 glycans may not reside in the internal cavity of the Fc region, as shown in Figure 5.1B, but are solvent exposed (Scapin et al., 2015). That study attributed this outcome to the shorter IgG4 hinge that forced the C12 domains into an unorthodox conformation that differed from the available IgG4-Fc crystal structures (Davies et al., 2014; Davies et al., 2017; Tam et al., 2017). Such a C12 domain rearrangement may itself reduce IgG4 function because the contact residues for FcγR and C1q binding have been displaced, which appears unlikely.

The atomistic scattering modelling approach resulted in a molecular explanation of the changes induced in IgG4 following glycan removal. Following is a discussion of molecular modelling approaches to this. Traditionally, solution scattering is a low-resolution method with a structural resolution of around 2 nm, while protein crystallography routinely achieves resolutions that are over ten times better. Fitting atomistic structures to the scattering curves thus improves the utility of solution scattering. The first IgG4 modelling analysis in 2014 used “constrained” modelling based on the SCT/SCTPL modelling approach (available in SASSIE-web (Wright et al., 2015) with the Fab and Fc crystal structures that were connected by a hinge region that was conformationally randomised using molecular dynamics to give 20,000 trial IgG4 models (Rayner et al., 2014). Ten different fits from X-ray and neutron data out to Q values of 1.00-2.00 nm⁻¹ were reported for two variants of human IgG4, namely IgG4 Ser²₂⁸ and IgG4 Pro²₂⁸, where the latter stabilised the hinge to prevent Fab arm exchange. Different NaCl buffers were studied. The resulting good curve fits revealed that the IgG4-Fab regions restricted access to the Fc region. This Fc restriction was offered as a possible explanation for the inability of IgG4 to activate complement. These studies revealed overall asymmetric solution structures, which agree with published crystal structures of the full length human IgG4 (Scapin et al., 2015; Blech et al., 2019). This new structure for glycosylated IgG4 from the Monte Carlo modelling approach in SASSIE agreed with this (Figure 5.12). The best-fit R-factors were 2.5-4.1%. In a previous modelling study of glycosylated IgG4, further
X-ray and neutron data collections out to $Q$ values of 1.5 nm$^{-1}$ were performed, and accompanied by scattering curve fits to 190,437 trial models produced from molecular dynamics and SASSIE Monte Carlo simulations (Wright et al., 2019). The best-fit R-factors were 3%. Both symmetric and asymmetric IgG4 solution structures were determined. Nonetheless, docking simulations of the IgG4-FcγRII interaction showed greater steric clashes, when compared to IgG1; this explained the observed lower binding affinity of the IgG4-FcγRII interaction compared to that of IgG1. Of particular interest in the latter study is that the earlier use of the SCT/SCTPL approach explicitly incorporated hydration shells in a coarse-grained approach (Wright et al., 2015). The SASSIE atomistic modelling study (Watson et al., 2013) did not include atomistic representation of hydration shells because this is computationally expensive. Nonetheless the outcomes from both IgG4 modelling studies were similar and resulted in asymmetric IgG4 solution structures. In this present third modelling study, the conformational libraries of IgG4 A33 were based on X-ray and neutron data that extend out to $Q$ values of 1.5 nm$^{-1}$ and resulted in fits with low R-factor of 1.50% or less (Table 5.2 and Table 5.4). The R-factor improvement is attributed to the improved signal-noise ratio in the experimental scattering curves obtained from the B21 Instrument. The ability to combine atomistic modelling approaches with improved high quality SAXS curves with little noise at larger $Q$ values enabled us to study smaller perturbations in the structure of IgG4 following its deglycosylation. In conclusion the joint use of improved simulations and data collection extend the utility of molecular modelling of scattering curves for large biological macromolecules.
Chapter 6 Thermostability studies of glycosylated and deglycosylated IgG1 and IgG4 demonstrate the stabilising role of the Fc glycan
6.1 Introduction

The human immunoglobulin G (IgG) class is the most abundant antibody in serum and comprises four subclasses IgG1, IgG2, IgG3 and IgG4, named in order of their concentration in serum (Hamilton, 1987). Each IgG is arranged into the characteristic Y shape with two Fab regions connected to a Fc region through a hinge, which varies in length between the subclasses. IgG4 has the shorter hinge with 12 residues, held together by two disulphide bonds at Cys\(^{226}\) and Cys\(^{229}\) (EU numbering) (Vidarsson et al., 2014) and IgG1 has a slightly longer hinge with 15 amino acids, also held together by two disulphide bonds at Cys\(^{226}\) and Cys\(^{229}\) (Vidarsson et al., 2014) (Figures 6.1 and 6.2). IgG1 and IgG4 are widely used as biotherapeutics due to their high specificity to antigens through Fab binding, and their ability to incite effector functions through the Fc binding Fc\(\gamma\) receptors (Fc\(\gamma\)Rs) on immune cells. IgG1 and IgG4 are functionally different, with IgG4 being considered an anti-inflammatory antibody. IgG4 does not activate the complement cascade through the binding of C1q and does not bind the proinflammatory Fc\(\gamma\)Rs, namely, Fc\(\gamma\)RIIIa and Fc\(\gamma\)RIIIb, and only weakly associates with the other Fc\(\gamma\)Rs (Chapter 1). This is in contrast to IgG1 which can bind all the Fc\(\gamma\)Rs and activate complement (Vidarsson et al., 2014).

The conserved N-linked glycosylation of the Fc region plays a key functional role in all four IgG subclasses (Figure 6.1B). A complex-type biantennary glycan with a Man\(_3\)GlcNAc\(_2\) core and two NeuNAC.Gal.GlcNAc antennae is attached at Asn\(^{297}\) on each CH2 domain (Deisenhofer, 1981) (Figure 6.2). However, the glycan structure is chemically heterogenous (Wuhrer et al., 2007). The Fc glycans modulate several IgG-Fc effector functions (Wada et al., 2019). The role that the Fc glycan plays in the structure of IgG1 and IgG4 is still not fully understood. Recent studies that applied a combined analytical ultracentrifugation, small angle scattering and atomistic modelling of IgG1 (Chapter 4; Spiteri et al., 2021b) and IgG4 (Chapter 5; Spiteri et al., 2021a) have found that the Fc glycan plays a role in stabilising the conformation of the Fc region.

To follow these solution structural studies, here a multidisciplinary structural approach has been employed to clarify the thermal stability of
Figure 6.1 The human IgG1 and IgG4 domains and their glycosylation.

A and B The heavy chains are comprised of $V_H$, $C_{H1}$, $C_{H2}$ and $C_{H3}$ domains, and the light chains are comprised of $V_L$ and $C_L$ domains. For both subclasses the heavy chains are connected by two Cys-Cys disulphide bridges at Cys$^{226}$ and Cys$^{229}$ and an N-linked oligosaccharide at Asn$^{297}$ is present on each of the $C_{H2}$ domains. A The hinge region in IgG1 connecting the Fab and Fc regions was constructed from 23 residues $^{216}EPKSCDKTHTCPPCPAPELGGP^{238}$. B The hinge region in IgG4 connecting the Fab and Fc regions was constructed from 20 residues $^{216}ESKYGGPPCPAPELGGP^{235}$.

C At the left, the glycosylation of IgG-Fc at two Asn$^{297}$ residues in the IgG1 Fc region is shown as stick models. The central schematic shows the approximated glycosylation pattern used in this study ($N$-acetyl glucosamine, GlcNAc; mannose, Man; galactose, Gal; $N$-acetyl neuraminic acid, NeuNAc. At the right, the detailed view of a single $C_{H2}$ domain with its glycan chain is shown, with the glycan colours coordinated with those in the schematic.
Figure 6.2 Sequence alignment of human IgG1 and IgG4 (legend overleaf).
Figure 6.2 Sequence alignment of human IgG1 and IgG4 (continued).

A-G The IgG1 A33 and IgG4 A33 sequences were kindly provided by UCB. A, B, the V\textsubscript{L} and C\textsubscript{L} domains; C-E, the V\textsubscript{H} and C\textsubscript{H1} domains and the hinge; F, G, the C\textsubscript{H2} and C\textsubscript{H3} domains, with Asn\textsuperscript{297} in dark blue. EU sequence numbering was used, and the complementary determining regions (CDR) sequences were identified in red. The IgG4 S225P mutation was identified in cyan and red. Beneath the alignments, consensus symbols indicated the degree of conservation, where (*) indicates full conservation, (:) indicates conservation between groups of strongly similar properties based on the Gonnet PAM 250 matrix, (.) indicates conservation between groups of weakly similar properties, and a space indicates no conservation.
glycosylated and deglycosylated IgG. Static light scattering (SLS) and intrinsic fluorescence was used to follow the thermal melting of glycosylated and deglycosylated IgG1 and IgG4. Previous thermostability studies have already demonstrated that the differences in the glycosylation pattern of the Fc glycans influences the melting temperature and aggregation propensity of IgG1 (Bowden et al., 2012; Wada et al., 2019). The greater the number of glycan moieties included in the Fc glycan chain, the higher the melting temperature, i.e. greater stability (Wada et al., 2019). When IgG1 was completely deglycosylated the melting temperature was significantly lower (~5.2 ± 1.0°C) than the fully glycosylated form (Bowden et al., 2012). Small angle X-ray and neutron scattering (SAXS and SANS) was then used to clarify the conformational state of both IgG subclasses during their thermal melting. This combination of melting temperatures derived from intrinsic fluorescence with SAXS and SANS data collected across different temperatures clarified the molecular basis for the stabilising role of the Fc glycan.

6.2. Materials and Methods

6.2.1 Purification and composition of IgG1 and IgG4

IgG1 A33 (148 kDa) and IgG4 A33 (146.9 kDa) was generously supplied by Dr John O’Hara and Dr Berni Sweeney (Union Chimique Belge). Its enzymatic deglycosylation was performed using peptide:N-glycosidase F (PNGase F) (35.5 kDa; New England Biolabs, Massachusetts, USA) for reason of its ability to remove glycans completely from glycosylated Asn residues (Plummer et al., 1984). To digest the glycans, 3.7 μl PNGase F (1850 activity units) was used to deglycosylate 150 μl IgG1 A33 and IgG4 (16.3 mg/ml). Native IgG1 and native IgG4 were incubated at 37°C for 10 hours. Each deglycosylated sample was filtered, through three successive dilutions using Amicon Ultra-0.5 ml centrifugal filters (100 kDa cut-off), which simultaneously allowed the PNGase F to pass through the membrane, while concentrating the deglycosylated sample. Immediately before the UNit, SAXS and SANS measurements, glycosylated and deglycosylated IgG1 and IgG4 were purified by gel filtration to remove any non-specific aggregates using a Superose 6 Increase 10/300 GL column (Cytiva,
Amersham, UK), then concentrated using Amicon Ultra-15 ml spin concentrators (100 kDa cut-off) and dialyzed at 4°C into 20 mM L-histidine, 138 mM NaCl, and 2.6 mM KCl buffer, pH 6.0. This histidine buffer was found to increase the stability of IgG1 and IgG4. The sequences of IgG1 A33 and IgG4 A33 were aligned against one another (Figure 2). The N-linked glycans at Asn\textsuperscript{297} on the C\textsubscript{H}2 domains were approximated as complex-type biantennary structures with a Man\textsubscript{3}GlcNAc\textsubscript{2} core and two NeuNAc.Gal.GlcNAc antennae (Deisenhofer, 1981). From these sequence, the molecular masses of glycosylated IgG1 A33 and IgG4 A33 were respectively calculated to be 148.4 kDa and 148.1 kDa, their unhydrated volume were 191.4 nm\textsuperscript{3} and 190.7 nm\textsuperscript{3}, their hydrated volume were 252.0 nm\textsuperscript{3} and 251.2 nm\textsuperscript{3} (based on 0.3 g of water/g of glycoprotein and an electrostricted volume of 0.0245 nm\textsuperscript{3} per bound water molecule), their partial specific volume \( \bar{\rho} \) were 0.731 ml/g and 0.730 ml/g and their absorption coefficients were 14.0 (1%, 1-cm pathlength, 280 nm) (Perkins, 1986). The molecular masses of deglycosylated IgG1 A33 and IgG4 A33 were, respectively, 144.0 kDa and 143.7 kDa, their unhydrated volume was 186.7 nm\textsuperscript{3} and 186.0 nm\textsuperscript{3}, their hydrated volume was 245.4 nm\textsuperscript{3} and 244.6 nm\textsuperscript{3}, their partial specific volume \( \bar{\rho} \) was 0.733 ml/g and 0.732 ml/g and their absorption coefficient was 14.4 cm\textsuperscript{-1} and 14.5 cm\textsuperscript{-1} (Perkins, 1986). The X-ray and neutron scattering densities of glycan residues are similar to those for hydrophilic (polar) amino acid residues, these being slightly higher than those for hydrophobic (non-polar) amino acid residues (Perkins, 1986).

The completeness of deglycosylation was verified by Superose 6 gel filtration, SDS-PAGE, and mass spectrometry. In the Mass Spectrometry Facility at the Chemistry Department, University College London, the antibodies were analysed on an Agilent 6510 Quadrupole time-of-flight liquid chromatography mass spectrometry system (Agilent, UK). Ten \( \mu \)L of each sample was injected onto a PLRP-S, 1000A, 8 \( \mu \)M, 150 mm \( \times \) 2.1 mm column, which was maintained at 60 °C at a flow of 0.3 ml/min. The separation was achieved using mobile phase A (water with 0.1% formic acid) and B (acetonitrile, with 0.1% formic acid) using a gradient elution. The column effluent was continuously electrosprayed into the capillary electrospray ionization source of the Agilent 6510 QTOF mass spectrometer and electrospray ionization mass spectra were acquired in positive electrospary ionisation mode using the \( m/z \) range 1,000–3200 in profile.
mode. The raw data was converted to zero charge mass spectra using the maximum entropy deconvolution algorithm in the MassHunter software version B.07.00. The single glycan mass was also found by subtracting the mass of fully glycosylated IgG1/IgG4 from that for deglycosylated IgG1/IgG4 and halving this mass.

6.2.2 Thermal melting of glycosylated and deglycosylated IgG1 and IgG4

The thermal stability of glycosylated and deglycosylated IgG1 and IgG4 were each measured using a UNit system (Unchained Labs, Pleasanton, California, USA). Each sample was measured in a range of concentrations (1.0 - 4.0 mg/ml) in 20 mM L-histidine, 138 mM NaCl, and 2.6 mM KCl buffer, pH 6.0. The sample was step-heated from 20 to 95°C at 1°C/min and with 30 s equilibration at each temperature. The intrinsic fluorescence spectrum of the folded protein at each temperature was recorded in triplicate. The Barycentric Mean (BCM) of each spectrum was calculated by the instrument software and plotted against temperature. In all cases, it was assumed that rapid equilibrium unfolding is convoluted with irreversible aggregation from the unfolded protein during thermal ramping, and hence the melting temperature $T_m$ values were presented as apparent, $T_{m,app}$. The measured $T_{m,app}$ values were reproducible for a given thermal ramp rate between replicates runs, and gave a measure of the conformational stability that allows comparison of the glycosylated and deglycosylated IgG samples.

Adequate curve fits were achieved by dividing each curve into two segments, where the first and second segments encompassed the first transition ($T_{m,app 1}$) that consisted of data points from approximately 320-346 K (46.9-72.9 °C), and the second transition ($T_{m, app 2}$) consisted of data points from approximately 335-367 K (61.9-93.9 °C), with variation of a few degrees between curves. The first curve segment was fitted to a two-state unfolding model, to extract the first apparent midpoint of unfolding transitions ($T_{m,app 1}$). The second curve segment was fitted to a three-state unfolding model to extract the second apparent midpoint of unfolding transitions ($T_{m,app 2}$):

The two-state unfolding model is given by (Nashine et al., 2013; Chakroun et al., 2016; Robinson et al., 2018; Zhang et al., 2020)
with an equilibrium constant for the transition between native and denatured state given by:

\[ K = \exp \left[ \frac{\Delta H_{\text{vh}}}{R} \left( \frac{1}{T_{m,\text{app}} - 1} \right) \right] \]

where \( T \) is the experimental temperature, \( T_{m,\text{app}} \) is the temperature at which half the protein is denatured. \( I_T, I_N \) and \( I_D \) are spectroscopic signals of the protein at each temperature, for the native, and fully denatured state, respectively. \( a \) and \( b \) are the baseline slopes of the native and denatured region of the curve. \( \Delta H_{\text{vh}} \) is the van’t Hoff enthalpy and \( R \) is the molar gas constant. All temperatures in these equations are in Kelvin.

The three-state unfolding model is given by (Zhang et al., 2020):

\[ I_T = \frac{(I_N + aT) + (I_D + bT) K_1 + (I_D + cT) K_1 K_2}{1 + K_1 + K_1 K_2} \]

where \( K_1 \) is the equilibrium constant for the transition between the native and intermediate state and \( K_2 \) is the equilibrium constant between the intermediate and denatured states given by:

\[ K_1 = \exp \left[ \frac{\Delta H_{\text{vm1}}}{R} \left( \frac{1}{T_{m,\text{app1}} - 1} \right) \right] \text{ and } K_2 = \exp \left[ \frac{\Delta H_{\text{vm2}}}{R} \left( \frac{1}{T_{m,\text{app2}} - 1} \right) \right] \]

where \( T_{m,\text{app1}} \) and \( T_{m,\text{app2}} \) are the temperatures at which the protein is half denatured for the transition of native to intermediate and intermediate to denatured states, respectively. \( a, b \) and \( c \) are the baseline slopes of the native, intermediate and denatured region of the curve. \( \Delta H_{\text{vm1}} \) and \( \Delta H_{\text{vm2}} \) are the van’t Hoff enthalpy for both transitions and \( R \) is the molar gas constant. All temperatures in this equation are in Kelvin. \( T_{m,\text{app1}} \) determined from fitting of the three-state unfolding model to the second curve segment was discarded, as it overlapped with \( T_{m,\text{app1}} \) derived from the two-state model fitting of the first curve segment. \( T_{m,\text{app1}} \) was not determined from the three-state unfolding model because its a signal was weaker as fewer points were fitted for this transition.

The apparent mole fraction of the unfolded species was given by:

\[ f_{\text{app}} = \frac{\exp \left[ \frac{\Delta H_{\text{vh}}}{R} \left( \frac{1}{T_{m} - 1} \right) \right]}{1 + \exp \left[ \frac{\Delta H_{\text{vh}}}{R} \left( \frac{1}{T_{m} - 1} \right) \right]} \]

where \( f_{\text{app}} \) is the apparent mole fraction of the unfolded species at temperature \( T \) (Zhang et al., 2020).
Static light scattering (SLS) data was collected concurrently to the fluorescent BCM data at two wavelengths, 266 nm and 473 nm. This monitored the aggregation state of the IgG samples. To analyse the SLS data, temperature versus SLS intensity was plotted and fitted to a Boltzmann equation on OriginPro (Version 2020b, OriginLab Corporation, Northampton, MA, USA) in order to extract the aggregation temperature, $T_{agg}$. This entailed the use of the Boltzmann equation:

$$y = A_2 + \frac{(A_1 - A_2)}{1 + e^{(x-x_0)/d_x}}$$

where $A_1$ is the $y$-axis minimum, $A_2$ is the $y$-axis maximum, $x_0$ is the 50% threshold defined by $\frac{A_1 - A_2}{2}$, and $d_x$ is a temperature constant. The $T_{agg}$ value is derived from the point closest to the onset of aggregation at a signal change of 10%, therefore this point was calculated from the fitted Boltzmann curve when:

$$T_{agg} = dx \left[ \ln \left( \frac{A1}{10\%} - \frac{A2 - 1}{A2} \right) \right] + x_0$$

### 6.2.3 X-ray and neutron scattering data and analyses for IgG1 and IgG4

X-ray scattering data were obtained during one beam session (February 2019) on Instrument B21 (Cowieson et al., 2020) at the Diamond Light Source at the Rutherford Appleton Laboratory (Didcot, UK), operating with a ring energy of 3 GeV, and an operational energy of 12.4 keV. A PILATUS 2M detector with a resolution of $1475 \times 1679$ pixels (pixel size of $172 \times 172$ $\mu$m) was used with a sample-to-detector distance of 4.01 m giving a $Q$ range from 0.04 nm$^{-1}$ to 4 nm$^{-1}$ (where $Q = 4 \pi \sin \theta / \lambda$; $2\theta$ = scattering angle; $\lambda$ = wavelength). The glycosylated and deglycosylated IgG1 and IgG4 at concentrations of 1.0 and 4.00 mg/ml in light water buffer were loaded onto a 96 well plate and placed into an EMBL Arinax sample holder (Pernot et al., 2013; Round et al., 2015). This condition showed the antibody molecule as a hydrated structure in a high positive solute-solvent contrast (Perkins, 1986). An automatic sampler injected 30 $\mu$l of sample from the well plate into a temperature-controlled quartz cell capillary with a diameter of 1.5 mm. The temperature in the capillary was varied in a range between 20-60$^\circ$C, with increments of 5$^\circ$C, and each sample was held in the capillary and equilibrated to the temperature for 3 minutes before being exposed
to the beam. Data sets of 30 frames with a frame exposure time of 1 second each were acquired in duplicate as a control of reproducibility. Checks during data acquisition confirmed the absence of radiation damage. Buffer subtraction and data reduction, in which the 30 frames were averaged was done automatically by an inhouse pipeline (on B21) (Basham et al., 2015).

Neutron scattering data on glycosylated and deglycosylated IgG1 and IgG4 at concentrations of 1.0 mg/ml and 4.0 mg/ml in heavy water buffer were obtained in one beam session (March 2019) on instrument SANS2D at the ISIS pulsed neutron source at the Rutherford Appleton Laboratory (Didcot, UK) (Heenan et al., 2011). This condition showed the antibody structure in a high negative solute-solvent contrast (Perkins, 1986). No conformational differences in the antibody between light and heavy water were detected in this study or previously (Wright et al., 2019). A pulsed neutron beam was derived from proton beam currents of ~40 μA. SANS2D data were recorded with 4 m of collimation, a 4 m sample-to-detector distance, a 12 mm sample aperture, and a wavelength range of 0.175-1.65 nm made available by time of flight. This gave a $Q$ range from 0.05 nm$^{-1}$ to 4 nm$^{-1}$. The data were acquired using a two-dimensional 3He detector with 512 × 512 pixels of 7.5 × 7.5 mm$^2$ in size. Samples of volume 1 ml were measured in 2 mm path length circular banjo cells for 1 hour in a thermostatted sample rack, for which the temperature was adjusted to 20°C, 35°C, 50°C, 65°C, and 80°C. The sample was equilibrated to the temperature for 5 minutes prior to beam exposure and data collection. Data were reduced using MANTID software (Arnold et al., 2014). The MANTID data reduction steps include corrections for the $Q$ resolution, i.e. beam divergence effects and smearing from the shape and size of the slits, as well as the wavelength overlap in each pulse (Arnold et al., 2014). Using SASview software, the Guinier analyses (below) were found to be almost unaffected if the smearing was turned on or off.

Guinier analyses of the scattering data gave information of the radius of gyration $R_G$, cross-sectional radius $R_{XS}$ and molecular mass. The scattering curve $I(Q)$ intensities at low $Q$ are defined by the $R_G$ value which is the averaged distance of each scattering point from the centre of scattering. In a given solute-solvent contrast, the radius of gyration $R_G$ is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect.
Guinier analyses at low $Q$ gave the $R_G$ value and the forward scattering at zero angle $I(0)$ (Glatter et al., 1982):

$$\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}$$

For antibodies, this expression is valid in a $Q,R_G$ range up to 1.5, and was used in previous studies (Rayner et al., 2015; Wright et al., 2019), although the usual upper range reported in the literature is 1.0-1.3. If the structure is elongated, the mean radius of gyration of the cross-sectional structure $R_{XS}$ and the mean cross-sectional intensity at zero angle $[I(Q)Q]_{Q\rightarrow 0}$ is obtained from (Pilz et al., 1970):

$$\ln[I(Q)Q] = [I(Q)Q]_{Q\rightarrow 0} - \frac{R_{XS}^2 Q^2}{2}$$

For immunoglobulins, it has been long recognised that the cross-sectional plot exhibits two regions, a steeper innermost one and a flatter outermost one (Pilz et al., 1970) and the two analyses are denoted by $R_{XS-1}$ and $R_{XS-2}$ respectively. The $R_{XS-1}$ parameter represents the averaged overall spatial separation of the Fab and Fc regions, while the $R_{XS-2}$ parameter represents the averaged spatial cross-section of the two Fab and one Fc region. The $R_G$ and $R_{XS}$ analyses were performed using SCT (Table 6.2 and 6.4) (Wright et al., 2015). The $Q$ ranges for the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values were 0.10–0.22, 0.29–0.52, and 0.66–1.05 nm$^{-1}$, respectively, as previously (Rayner et al., 2015; Wright et al., 2019; Spiteri et al., 2021b, 2021a). Indirect transformation of the scattering data $I(Q)$ in reciprocal space into real space to give the distance distribution function $P(r)$ was carried out using GNOM (version 4.6) (Semenyuk et al., 1991; Svergun, 1992).

$$P(r) = \frac{1}{2\pi^2} \int_{0}^{\infty} I(Q)Qr \sin(Qr) dQ$$

$P(r)$ corresponds to the distribution of distances $r$ between the volume elements in the macromolecule. This yields the maximum dimension of the macromolecule $L$ and its most commonly occurring distance vector $M$ in real space. For this $P(r)$ analysis, the X-ray $I(Q)$ curve utilized up to 223 data points in the $Q$ range between 0.0032 and 1.50 nm$^{-1}$ for both glycosylated and deglycosylated IgG1 and IgG4. The neutron $P(r)$ curve utilized up to 66 $I(Q)$ data points in the $Q$ range between 0.010 and 1.50 nm$^{-1}$ for both glycosylated and deglycosylated IgG1 and IgG4.
6.3 Results

6.3.1 Purification and characterisation of glycosylated and deglycosylated IgG1 and IgG4

Human IgG1 A33 and IgG4 A33 possess identical $V_H$ and $V_L$ domains but differ in their heavy chain sequences. Hence any differences seen between them reflects their hinge regions and Fc regions and not their Fab regions. A protocol for the deglycosylation of the monoclonal human IgG1 A33 and IgG4 A33 antibody was set up using peptide-$N$-glycosidase F (PNGase F) digests according to the manufacturer's protocol (section 6.2.1). The completeness of deglycosylation was verified by a three-fold combination of routine gel filtration, SDS-PAGE and mass spectrometry (see below):

(i) At ten hours after the start of the digests, the IgG1 and IgG4-digested products eluted from Superose 6 Increase gel filtration column slightly prior to that for glycosylated IgG1 and IgG4 (Figure 6.3A and 6.3B). A 10 hour digest was sufficient for glycan removal (Spiteri et al., 2021b, 2021a). Glycosylated and deglycosylated IgG1 eluted as a main symmetrical peak at 17.88 ml and 17.86 ml, respectively (Figure 6.3A). Glycosylated and deglycosylated IgG4 eluted as a symmetrical peak at 17.54 ml and 17.53 ml, respectively, (Figure 6.3B). This procedure ensured that each sample was monodisperse with no aggregates present immediately before UNit or scattering experiments.

(ii) When the IgG1 and IgG4 samples were submitted to non-reducing and reducing SDS-PAGE analyses at equimolar concentrations, purified glycosylated and deglycosylated IgG1 showed a single band between 200 and 116 kDa on 4-12% Bis Tris NuPage gel under non-reducing conditions, which is consistent with the expected masses of ~147 kDa for IgG1 (Figure 6.3C). Under reducing conditions two bands were present corresponding to the heavy chain (with an apparent mass of ~50 kDa) and the light chain (with an apparent molecular mass of between ~25 kDa) (Figure 6.3C). These apparent molecular masses were as expected from the known sequence. Purified glycosylated and deglycosylated IgG4 showed a single band between 200 and 116 kDa on 4-12% Bis Tris NuPage gel under non-reducing conditions, which is consistent with the expected masses of ~147 kDa for IgG4 (Figure 6.3D). Under reducing conditions two bands were present corresponding to the heavy chain (with an apparent mass of ~55 kDa)
Figure 6.3 Purification, SDS-PAGE and mass spectrometry of human glycosylated and deglycosylated IgG1 and IgG4 (legend overleaf).
Figure 6.3 Purification, SDS-PAGE and mass spectrometry of human glycosylated and deglycosylated IgG1 and IgG4 (continued).

A and B Elution peaks from a Superose 6 Increase 10/300 gel filtration column for glycosylated (black) and deglycosylated (magenta) IgG1 and IgG4 samples. The dashed vertical lines indicate the peak positions. C and D Lane 1 and 4, molecular mass markers are denoted in kDa. Lanes 2 and 3, non-reduced SDS-PAGE of glycosylated and deglycosylated IgG1/IgG4. 5 and 6, reduced SDS-PAGE of glycosylated and deglycosylated IgG1/IgG4. E and F Mass spectra of glycosylated and deglycosylated IgG1 and IgG4, using the same colour scheme as in A and B. Peaks labelled G represents glycosylated species, and D represents fully deglycosylated species.
and the light chain (with an apparent molecular mass of between 31 and 21 kDa) (Figure 6.3D). These apparent molecular masses were as expected from the known sequence.

(iii) Liquid chromatography mass spectrometry measurements showed multiple peaks for glycosylated IgG1 (G) that were assigned to the presence of at least four glycoforms, separated by masses of 160-230 Da that corresponded to single sugar residues (Figure 6.3E) (Hui et al., 2019). Glycosylated IgG4 (G) showed multiple peaks that were assigned to the presence of at least five glycoforms, separated by masses of 157-168 Da that also corresponded to single sugar residues (Figure 6.3F). The most intense IgG1 glycosylated population had an observed deconvoluted mass of 147,010 Da and for IgG4 the most intense glycosylated population had an observed deconvoluted mass of 146,848 Da. After ten-hour digest with PNGase F, only single dominant deglycosylated peak (D) was seen at 143,958 Da for IgG1 and 143,636 Da for IgG4. The mass of each glycan chain was calculated by subtracting the glycosylated and deglycosylated masses and halving the outcome to give 1,526 Da for the IgG1 Fc glycan and 1,606 Da for IgG4 Fc glycan.

6.3.2 Melting and aggregation analysis of glycosylated and deglycosylated IgG1 and IgG4

Thermal denaturation curves were obtained for glycosylated and deglycosylated IgG1 and IgG4, using intrinsic fluorescence as a structural probe. The impact of Fc glycan on the stability of IgG1 and IgG4 was determined from the apparent melting temperature ($T_{m,\text{app}}$). Analysis of the intrinsic fluorescence data, converted to the barycentric mean (BCM) by the instrument, revealed that glycosylated and deglycosylated IgG1 at pH 6 underwent two transitions (Figure 6.4A). The curves were fitted in two segments (Section 6.2.2) to a two-state and three-state unfolding model. The resulting mean apparent melting temperatures ($T_{m,\text{app}}$) for glycosylated IgG1 were $64.8 \pm 2.5^\circ C$ and $83.6 \pm 0.4^\circ C$, respectively for each transition (Figure 6.5A, Table 6.1). A previous study of IgG1 A33 at 1.0 mg/ml in a pH 4.5, 100 mM ionic strength buffer found that $T_{m,\text{app}}$ was
Figure 6.4 Thermal unfolding curves for glycosylated and deglycosylated IgG1 and IgG4 (legend overleaf).
Figure 6.4 Thermal unfolding curves for glycosylated and deglycosylated IgG1 and IgG4 (continued).

Experimentally derived curves for glycosylated (●) and deglycosylated (●) IgG1 (left, A, C and E) and IgG4 (right, B, D and F). The least-squared fitted curves are shown in solid lines, in red for glycosylated and in black for deglycosylated. Temperature displayed in Kelvin for ease of fit. For each sample, glycosylated and deglycosylated IgG1 and IgG4 one curve is shown corresponding to 4 mg/ml. The curves shown were generated in the same repeat. (A and B) The experimental barycentric mean curve (BCM), the fit was achieved by fitting a two-state or three-state unfolding model in to two curve segements, the first segment spanning 320-346 K and the second spanning 335-367 K. Fitting of the BCM gave an apparent melting temperature for each of the two transitions, $T_{m, app 1}$ and $T_{m, app 2}$ (C and D) Experimental data for static light scattering (SLS) at 266 nm where the fitted curve was achieved by the Boltzmann equation. (E and F) Experimental data for SLS at 472 nm, fit was achieved by the Boltzmann equation. Fitting of SLS gave the aggregation temperature $T_{agg}$. 
Figure 6.5 Thermostability parameters for IgG1 and IgG4.

Thermostability parameters; $T_{m,\text{app} \ 1}$, $T_{m,\text{app} \ 2}$, $T_{\text{agg}}$ derived from fitting of experimental data for glycosylated (●) and deglycosylated (●) IgG1 (left panels A, C and E) and IgG4 (right panels B, D and F). Linear regression shown as solid lines, in black for glycosylated and in magenta for deglycosylated. Data collected for all samples across a concentration range of 1.00 mg/ml-4.00 mg/ml. Temperature has been converted from Kelvin to Centigrade.

(A and B) The $T_{m,\text{app} \ 1}$ and $T_{m,\text{app} \ 2}$ obtained from the experimental data (C and D) $T_{\text{agg}}$ data derived from SLS at 266 nm. (E and F) $T_{\text{agg}}$ data derived from static light scattering at 473 nm.
Table 6.1 Experimental thermal unfolding data for glycosylated and deglycosylated IgG1.

<table>
<thead>
<tr>
<th>Unit Data</th>
<th>Concentration (mg/ml)</th>
<th>$T_{m,\text{app}}$ 1 (°C)</th>
<th>$T_{m,\text{app}}$ 2 (°C)</th>
<th>$T_{\text{agg}}$ 266 nm (°C)</th>
<th>$T_{\text{agg}}$ 473 nm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG1 Glycosylated</td>
<td>4</td>
<td>64.3 ± 0.1</td>
<td>83.3</td>
<td>78.2 ± 0.5</td>
<td>80.6 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>63.8</td>
<td>83.4</td>
<td>78.2 ± 0.4</td>
<td>80.6 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>68.4 ± 0.2</td>
<td>84.1 ± 0.2</td>
<td>78.3 ± 0.7</td>
<td>80.7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>62.5 ± 0.6</td>
<td>75.4 ± 2.1</td>
<td>79.9 ± 0.8</td>
<td>80.8 ± 0.2</td>
</tr>
<tr>
<td>IgG1 Deglycosylated</td>
<td>4</td>
<td>56.4 ± 0.1</td>
<td>79.8</td>
<td>77.3 ± 0.7</td>
<td>79.4 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>56.2</td>
<td>80.3 ± 0.5</td>
<td>77.3 ± 0.6</td>
<td>79.4 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>56.1 ± 0.2</td>
<td>79.4 ± 0.2</td>
<td>77.3 ± 1.1</td>
<td>79.5 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>56.1 ± 0.1</td>
<td>79.7 ± 0.5</td>
<td>78.2 ± 1.3</td>
<td>79.8 ± 1.2</td>
</tr>
</tbody>
</table>

Table 6.2 Experimental thermal unfolding data for glycosylated and deglycosylated IgG4.

<table>
<thead>
<tr>
<th>Unit Data</th>
<th>Concentration (mg/ml)</th>
<th>$T_{m,\text{app}}$ 1 (°C)</th>
<th>$T_{m,\text{app}}$ 2 (°C)</th>
<th>$T_{\text{agg}}$ 266 nm (°C)</th>
<th>$T_{\text{agg}}$ 473 nm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG4 Glycosylated</td>
<td>4</td>
<td>58.0 ± 0.7</td>
<td>75.3 ± 0.1</td>
<td>69.5 ± 1.0</td>
<td>71.9 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>57.4 ± 0.6</td>
<td>75.7 ± 0.3</td>
<td>70.2 ± 0.8</td>
<td>72.6 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>57.7 ± 0.5</td>
<td>75.6 ± 0.3</td>
<td>70.9 ± 0.9</td>
<td>73.4 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>57.5 ± 0.3</td>
<td>75.3 ± 0.7</td>
<td>71.3 ± 1.2</td>
<td>73.1 ± 1.1</td>
</tr>
<tr>
<td>IgG4 Deglycosylated</td>
<td>4</td>
<td>51.4 ± 0.1</td>
<td>74.4 ± 0.1</td>
<td>68.6 ± 0.6</td>
<td>70.5 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>51.1 ± 0.1</td>
<td>74.6 ± 0.1</td>
<td>68.4 ± 1.1</td>
<td>70.8 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>51.1 ± 0.1</td>
<td>75.4 ± 0.7</td>
<td>68.0 ± 0.5</td>
<td>70.8 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>51.3 ± 0.2</td>
<td>75.5 ± 0.6</td>
<td>69.0 ± 0.6</td>
<td>72.0 ± 0.1</td>
</tr>
</tbody>
</table>
56 ± 3°C and $T_{\text{m,app}}$ was 75.7 ± 0.4°C (Zhang et al., 2020). The difference between the two experiments may result from the different buffer. Here an L-histidine buffer at pH 6.0 was used (Section 6.2.1) which stabilised the IgG1 structure (Spiteri et al., 2021b). The mean apparent melting temperatures for deglycosylated IgG1 were 56.2 ± 0.1°C ($T_{\text{m,app}}$1) and 79.8 ± 0.4°C ($T_{\text{m,app}}$2) (Table 6.1). The difference in $T_{\text{m,app}}$1 between glycosylated and deglycosylated IgG1, was approximately 8.6°C and for $T_{\text{m,app}}$2 this difference was approximately 3.8°C. This indicates that $T_{\text{m,app}}$1 is possibly correlated to the unfolding of the Fc region, and that the presence of the Fc glycan has a significant impact on the overall stability of the IgG1. $T_{\text{m,app}}$2 is a measure of the Fab region unfolding (Zhang et al., 2020) and is therefore less affected by the removal of the glycan.

The aggregation temperature ($T_{\text{agg}}$) was derived from the fitting of the SLS curves measured by the UNit at two wavelengths, 266 nm and 473 nm (Figure 6.4C-F). Two wavelengths are used for reason of their being technical controls. The mean $T_{\text{agg}}$ for glycosylated IgG1 was 78.6 ± 0.8°C (from 266 nm) and 80.6 ± 0.1°C (from 473 nm). The mean $T_{\text{agg}}$ for deglycosylated IgG1 was 77.5 ± 0.5°C (from 266 nm) and 79.5 ± 0.2°C (473 nm). The difference in the mean $T_{\text{agg}}$ between glycosylated and deglycosylated IgG1 was approximately 1.1°C (266 nm) and 1.2°C (473 nm). This demonstrated that the removal of the Fc glycan does not significantly expedite the aggregation process despite the unfolding of the deglycosylated IgG1 at lower temperatures. $T_{\text{m,app}}$2 has been previous been attributed to the unfolding of the Fab region, as the isolated Fab region was found to have a high $T_{\text{m,app}}$ of 78.2 ± 0.2°C (Zhang et al., 2020). Here, $T_{\text{agg}}$ is more closely correlated to $T_{\text{m,app}}$2, which indicates that IgG aggregation is mediated by the Fab region unfolding. The $T_{\text{agg}}$ values derived from the fitting of the SLS curves from two wavelengths, were comparable. For both glycosylated and deglycosylated IgG1, a slight concentration dependence for $T_{\text{agg}}$ was observed whereby samples with a higher concentration aggregated at a lower temperature.

Glycosylated IgG4 underwent three apparent transitions, the final transition being due to sample aggregation (Figure 6.4B). The resulting mean apparent melting temperatures ($T_{\text{m,app}}$1 and $T_{\text{m,app}}$2) for glycosylated IgG4 were 57.7 ± 0.3°C and 75.5 ± 0.2°C, respectively (Figure 6.5B, Table 6.2). For deglycosylated IgG4, the mean $T_{\text{m,app}}$1 was 51.2 ± 0.1°C and $T_{\text{m,app}}$2 was 75.0 ± 0.6°C. The difference between glycosylated and deglycosylated $T_{\text{m,app}}$1 was
approximately 6.4°C and approximately 0.5°C lower for $T_{m,app}$ 2. These values were consistent with those from the analysis of IgG1 (above), further demonstrating that $T_{m,app}$ 1 corresponds to unfolding of the Fc region, which is stabilised by the Fc glycan. $T_{m,app}$ 2 is linked to Fab unfolding hence explanation why it is less sensitive to the presence or absence of the Fc glycan. The mean $T_{agg}$ for glycosylated IgG4 was 70.4 ± 0.8°C (from 266 nm) and 72.7 ± 0.7°C (from 473 nm). The mean $T_{agg}$ deglycosylated IgG4 was 68.5 ± 0.4°C (from 266 nm) and 71.0 ± 0.7°C (473 nm) (Figure 6.5D and F). The $T_{agg}$ values derived from the two wavelengths were comparable, as for IgG1. There was a difference of 2.0°C in the mean $T_{agg}$ derived at 266 nm for glycosylated and deglycosylated IgG4, and a 1.7°C difference in the mean $T_{agg}$ derived at 473 nm. As for IgG1, there was a slight concentration dependence for $T_{agg}$, whereby samples with a higher concentration aggregated at a lower temperature.

### 6.3.3 X-ray and neutron scattering of glycosylated and deglycosylated IgG1 and IgG4

The overall solution structures of glycosylated and deglycosylated IgG1 and IgG4 samples were analysed by X-ray and neutron scattering. The two methods provided different perspectives of the same solution structure. X-rays in light water buffers detect the hydration shell surrounding the protein structure, whereas the effect of the hydration shell is much reduced by neutrons in heavy water buffers for reason of the different solute-solvent contrast in use (Perkins, 1986, 2001; Svergun et al., 1998). Both IgG1 and IgG4 X-ray data collection at concentrations of 1.0 mg/ml and 4.0 mg/ml used time-frame analyses to ensure the absence of radiation damage effects. The resulting $R_G$ and $R_{XS-1}/R_{XS-2}$ values monitored the elongation of the overall IgG1 and IgG4 structure and its approximate cross-sectional structures respectively.

The corresponding neutron scattering data sets for glycosylated and deglycosylated IgG1 and IgG4 in 100% $^2$H$_2$O buffer were analysed at concentrations of 1.0 mg/ml and 4.0 mg/ml, as used above for SAXS. The temperature range used in neutron scattering was 20 to 80°C with 15°C intervals between temperatures. Samples could be probed at higher temperature on SANS2d compared to B21, because the thin quartz sample capillary in the BioSAXS robot on B21 (Diamond) was prone to shatter at temperatures higher
than 60°C, whereas the thicker quartz Banjo cells used in neutron scattering are more robust. Again, the Guinier analyses revealed high quality linear fits for the \( R_g \), \( R_{XS-1} \) and \( R_{XS-2} \) parameters (Figure 6.6A-C, bottom two panels). A temperature dependence was observed for IgG1 and IgG4, this being seen from the \( I(0)/c \) values which increases (Figure 6.7B and D).

Guinier analyses resulted in high quality linear plots and revealed three distinctive regions of the \( I(Q) \) curves, as expected for antibodies (Rayner et al., 2014, 2015; Hui et al., 2015). From these, the \( R_g \), \( R_{XS-1} \) and \( R_{XS-2} \) values that monitor the elongation of the overall IgG structure and its approximate cross-sectional structures respectively derived from the individual scattering curves were obtained within satisfactory \( Q.R_g \) and \( Q.R_{XS} \) limits of 0.5-1.1, 0.7-1.3 and 0.9-1.5 respectively (Figure 6.6A, B and C). For both IgG1 and IgG4 a temperature dependence was observed in the X-ray \( I(0)/c \) values that suggested at higher temperatures there is increased oligomer formation (Figure 6.6A and C). This transition was mirrored in the fluorescence unfolding and SLS data used to determined \( T_{m,app} \) and \( T_{agg} \), respectively, where the temperatures of the transition agreed with Figure 6.4. The distance distribution function \( P(r) \) is derived from Fourier transformation of the scattering curve \( I(Q) \), and provides structural information in real space on glycosylated and deglycosylated IgG. The maxima in the \( P(r) \) curves corresponded to the most frequently occurring distances between scattering elements within the structures. For both IgG1 and IgG4, two peaks, \( M1 \) and \( M2 \), were visible that are characteristic of antibody-shaped proteins. \( M1 \) corresponds primarily to the shorter distances within each Fab and Fc region. \( M2 \) corresponds primarily to the longer distances between pairs of Fab and Fc regions and monitors changes in the separation of the Fab and Fc regions (Figures 6.1 and 6.8). A temperature dependence was observed in the positions of the \( M1 \) and \( M2 \) peaks, and which were measured directly from their maximum values (Figure 6.9). At higher temperatures of greater than 60°C, where the structure of the IgG has been denatured, and the sample is aggregating, the \( M1 \) and \( M2 \) peaks are no longer visible.
Figure 6.6 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG1 and IgG4 at different temperatures (legend overleaf).
Figure 6.6 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG1 and IgG4 at different temperatures (continued).

The SAXS and SANS curves for glycosylated and deglycosylated IgG1/IgG4 at concentrations of 1.00 and 4.00 mg/ml. The filled circles between the arrows represent the $Q.R_G$ and $Q.R_{XS}$ fit ranges used to determine the $R_G$ and $R_{XS}$ values. (A) The $Q$ range used for the $R_G$ values was 0.10-0.22 nm$^{-1}$. The $Q$ ranges (B) for the $R_{XS-1}$ were 0.29-0.52 nm$^{-1}$ and (C) $R_{XS-2}$ values were 0.66-1.05 nm$^{-1}$, respectively.
Figure 6.7 Temperature dependence of the SAXS and SANS Guinier analyses (legend overleaf).
Figure 6.7 Temperature dependence of the SAXS and SANS Guinier analyses (continued).

The colours denote the glycosylated IgG1/IgG4 (black) and deglycosylated IgG1/IgG4 (magenta) timepoints. The $R_g$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for the glycosylated samples at 4.00 mg/ml and 1.00 mg/ml (●, ○), respectively and deglycosylated IgG1/IgG4 samples at 4.00 mg/ml and 1.00 mg/ml (●, ○), respectively. The solid lines corresponded to linear regression fits of glycosylated and deglycosylated samples at 4.00 mg/ml, and the dashed lines to the fits for glycosylated and deglycosylated samples at 1.00 mg/ml.

(A) The SAXS $R_g$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for IgG1, and (B) the SANS $R_g$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for IgG1, each corresponding to a single measurement in L-histidine buffer in $^2$H$_2$O. (C) The SAXS $R_g$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for IgG4, and (D) the SANS $R_g$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for IgG4, each corresponding to a single measurement in L-histidine buffer in $^2$H$_2$O.
6.3.3.1 X-ray and neutron scattering of glycosylated and deglycosylated IgG1

For IgG1 samples at 4.0 mg/ml the $R_G$ was unchanged from 5.12 ± 0.51 nm at 20°C to 5.11 ± 0.51 nm at 60 °C for glycosylated IgG1. In comparison the $R_G$ for deglycosylated IgG1 at 4.0 mg/ml slightly increased from 5.17 ± 0.52 nm at 20°C to 5.69 ± 0.13 nm at 60°C (Figure 6.7A, Table 6.3). The $R_{XS-1}$ values from the individual curves (Figure 6.5B) are an approximate monitor of the cross-sectional structure for glycosylated and deglycosylated IgG1. The $R_{XS-1}$ slightly decreased from 2.51 ± 0.73 nm at 20°C to 2.44 ± 0.71 nm at 60°C for glycosylated IgG1 and slightly increased from 2.53 ± 0.73 nm at 20°C to 2.67 ± 0.77 nm at 60 °C for deglycosylated IgG1. The $R_{XS-2}$ decreased from 1.40 ± 0.93 nm at 20°C to 1.27 ± 0.84 nm at 60°C for glycosylated IgG1 and decreased from 1.39 ± 0.92 nm at 20°C to 1.10 ± 0.73 nm at 60°C for deglycosylated IgG1. As $R_{XS-2}$ is a measure of internal distances of the Fab and Fc regions, this may indicate that the glycan protects the Fc region from compacting or unfolding as the temperature is increased. However, the errors on the $R_{XS}$ might as suggest that these changes are not significant. This is consistent with the trend observed in $T_{m,app}$.

The neutron $R_G$ values increased from 5.38 ± 0.54 nm at 20°C to 11.25 ± 1.13 nm at 80°C for glycosylated IgG1, and from 5.36 ± 0.06 nm at 20 °C to 11.41 ± 1.14 nm at 80°C for deglycosylated IgG1 at 4.0 mg/ml (Figure 6.7B, Table 6.3). This large increase in $R_G$ for all samples, indicated oligomer formation as the temperature is increased. The rate of oligomer formation is faster for both the deglycosylated IgG1 and IgG4 samples. This is also seen in the upward trend in the $I(0)/c$ plot (Figure 6.7B and D). The neutron $R_{XS-1}$ values increased from 2.40 ± 0.70 nm at 20°C to 3.35 ± 0.97 nm at 80°C for glycosylated IgG1 and from 2.40 ± 0.69 nm at 20°C to 3.78 ± 1.10 nm at 80°C, for deglycosylated IgG1. This increase in $R_{XS-1}$ can be attributed to unfolding followed by oligomerisation. The neutron $R_{XS-2}$ was slightly increased from 1.28 ± 0.85 nm at 20°C to 1.36 ± 0.90 nm at 80°C for glycosylated IgG1 and was essentially unchanged from 1.21 ± 0.80 nm at 20°C to 1.22 ± 0.80 nm at 80°C for deglycosylated IgG1. The neutron $R_{XS-2}$ was less informative than that from X-rays. This could be due to lower signal-to-noise at higher $Q$ values, and fewer data points. However, it would appear that that each Fab and Fc region remained intact.
The maximum lengths of glycosylated and deglycosylated IgG1 were determined from the value of $r$ when the $P(r)$ curve intersects zero on the $r$ axis and this was 16 nm for all glycosylated and deglycosylated IgG1 samples at temperatures between 20-60°C (Figure 6.8A and C). This indicates that the IgG1 structure was stable, even with glycan removal. From the UNit thermostability data the global melting temperature was above 60°C, which was outside the temperature range for the SAXS experiments. The neutron maximum lengths of both glycosylated and deglycosylated IgG1, increased from 16 nm at 20°C to 42 nm at 80°C. However, it should be noted that the deglycosylated IgG1 become elongated earlier at 65°C, where the maximum length was 25 nm. At the same temperature, glycosylated IgG1 has an unchanged maximum length of 16 nm.

The X-ray $M1$ peak did not shift from 4.09 nm at 20°C and 60°C for glycosylated IgG1 at 4.0 mg/ml. However, for deglycosylated IgG1 the peak shifted from 4.09 nm at 20°C to 4.46 nm at 60°C (filled circles, Figure 6.9A). For both glycosylated and deglycosylated IgG1 at a concentration of 1.0 mg/ml the $M1$ peak was more prone to shifting as temperature increased. The $M1$ peak for deglycosylated IgG1 at 1.0 mg/ml increased from 4.31 nm at 20°C to 5.31 nm at 60°C. For glycosylated IgG1 at 1.0 mg/ml the peak shift was smaller from 4.16 nm at 20°C to 4.66 nm at 60°C. This could be due to a concentration dependence, where a higher concentration protects IgG1 from denaturation, possibly due to crowding effects or due to lower signal-to-noise statistics at lower concentrations. This concentration dependence is apparent for the X-ray $M2$ peak which shifts downward for glycosylated and deglycosylated samples at 4.00 mg/ml from 7.53 nm at 20°C to 7.39 nm at 60°C and from 7.87 nm at 20°C to 7.39 nm at 60°C, respectively. At 1.0 mg/ml, the $M2$ peak for glycosylated IgG1 observes the same downward trend as at 4.0 mg/ml moving from 7.59 nm to 7.46 nm, but the $M2$ peak for deglycosylated IgG1 shifts upward 7.98 nm to 8.38 nm. This indicates that the molecule elongated faster at lower concentrations and that the presence of the Fc glycan helped to stabilise the structure of IgG1. For neutron analyses the $M1$ peak for glycosylated IgG1 was consistent with X-ray analysis and did not change from 4.12 nm until the sample was heated to 80°C when the double peak maxima disappear leaving a single peak at 13.35 nm at 80°C, which is indicative of oligomeric species (Figure 6.8C). For deglycosylated IgG1 the $M1$ peak was stable at 4.12 nm until about 65°C when only one peak was apparent at 8.81 nm.
at 65°C which increases to 12.70 nm at 80°C. This indicates that the oligomer is getting larger as the temperature is increased. At the same temperature glycosylated IgG1 is stable, with the double peaks being visible again illustrating the stabilising effect of the glycan.

The dimensionless Kratky plots of \((Q.R_G)^2.I(Q)/I(0)\) vs \(Q.R_G\), indicate whether the macromolecule in question is globular in its structure or possesses intrinsically disordered regions (Receveur-Brechet et al., 2012). The Kratky plots demonstrated two clear peaks at \(Q.R_G\) values of 1.90 nm and 3.97 nm for glycosylated and 1.91 nm and 4.11 nm for deglycosylated IgG1 (Figure 6.10A and C). Deglycosylated IgG1 is slightly more disordered than glycosylated IgG1, with the tail of the Kratky plot plateauing above that of glycosylated IgG1. The Kratky plots derived from neutron data are similar to those from X-ray data. At 65°C the second peak for deglycosylated IgG1 disappears, however, at the same temperature glycosylated IgG1 appears to be intact with two peaks being apparent. This further demonstrated that the Fc glycan mediates conformational stability of the IgG1. At 80°C the Kratky plot is uninformative as the sample has aggregated (Figure 6.10C).

### 6.3.3.2 X-ray and neutron scattering of glycosylated and deglycosylated IgG4

For IgG4 X-ray analyses, the \(R_G\) increased from 4.83 ± 0.48 nm at 20°C to 5.93 ± 0.49 nm at 60°C for glycosylated IgG4, and from 5.01 ± 0.50 nm at 20°C to 6.29 ± 0.63 nm at 60°C for deglycosylated IgG4 at 4.0 mg/ml (Figure 6.7C, Table 6.4). The \(R_{XS-1}\) decreased slightly from 2.53 ± 0.73 nm at 20°C to 2.47 ± 0.72 nm at 60°C for glycosylated IgG4 and increased from 2.53 ± 0.73 nm at 20°C to 3.09 ± 0.90 nm at 60°C for deglycosylated IgG4. This increase in \(R_{XS-1}\), which corresponds to Fab-Fab and Fab-Fc cross-sectional separations, indicated the molecule become more elongated as it unfolded. The \(R_{XS-2}\) decreased from 1.40 ± 0.93 nm at 20°C to 1.27 ± 0.84 nm at 60°C for glycosylated IgG4, and decreased from 1.39 ± 0.92 nm at 20°C to 0.83 ± 0.55 nm at 60°C for deglycosylated IgG4. The \(R_{XS-2}\), is a measure of the internal distances of the Fab and Fc regions, which therefore, indicated that they were change shape while unfolding.
For IgG4 neutron analyses, the $R_G$ increased from 5.1 ± 0.5 nm at 20°C to 12.0 ± 1.2 nm at 80°C for glycosylated IgG4, and from 5.1 ± 0.5 nm at 20°C to 12.2 ± 1.2 nm at 80°C for deglycosylated IgG4 at 4.0 mg/ml (Figure 6.7D, Table 6.4). This increase in $R_G$ corresponded to the oligomerisation of IgG4 at higher temperatures. For IgG4, the $R_{XS-1}$ values increased from 2.4 ± 0.7 nm at 20°C to 3.7 ± 1.1 nm at 80°C, and from 2.4 ± 0.7 nm at 20°C to 3.4 ± 1.1 nm at 80°C for deglycosylated IgG4. The $R_{XS-2}$ increased from 1.19 ± 0.78 nm at 20°C to 1.5 ± 1.0 nm at 80°C for glycosylated IgG4, and from 1.2 ± 0.8 nm at 20°C to 1.4 ± 0.9 nm at 80°C for deglycosylated IgG4. This increase in $R_{XS-2}$ corresponded to unfolding of the Fc and Fab regions.

The distance distribution curves for IgG4 gave maximum lengths ($L$) of the molecule ranging from 15.66 nm at 20°C to 16 nm at 60°C for glycosylated IgG4, and 16 nm at 20°C to 22 nm at 60°C for deglycosylated IgG4 (Figure 6.8B and 6.8D). This demonstrated that the Fc glycan stabilises the structure. The neutron maximum length increased from 16 nm at 20°C to 52 nm at 80°C for glycosylated IgG4, and from 15 nm at 20°C to 51 nm at 80°C for deglycosylated IgG4. At 50°C, deglycosylated IgG elongated to a maximum length of 18 nm, compared to 16 nm for glycosylated IgG4 at the same temperature. This was consistent with the X-ray analyses. For IgG4 the neutron maximum lengths were 1 nm smaller when compared to the X-ray value of 16 nm, this being attributed to the reduced contribution visibility of the hydration shell seen by neutron scattering (Figure 6.8D, Table 6.4). These reductions in the neutron $R_G$ and $L$ values have been previously seen in earlier joint SAXS and SANS studies of antibodies (Rayner et al., 2015).

The glycosylated and deglycosylated IgG4 structure being maintained up to 50°C, above this temperature the sample begins to oligomerise (Figure 6.8D). This is consistent with the X-ray analyses (Figure 6.8B and D). The X-ray $M1$ peak shifted from 4.16 nm at 20°C to 4.45 nm at 60°C for glycosylated IgG4, and from 4.23 nm at 20°C to 4.49 nm at 50°C for deglycosylated IgG4, both at a concentration of 4.00 mg/ml. Above 50°C the double peak maxima are no longer visible, with a single peak at 6.67 nm at 55 °C and 7.59 nm at 60°C. At a concentration of 1.00 mg/ml, the peak position shifts from 4.31 nm at 20°C to 4.88 nm at 60°C for glycosylated IgG4 and 4.38 nm at 20°C to 4.60 nm at 50°C for deglycosylated IgG4. At 55°C the double peaks are no longer visible, as the
The colours denote the glycosylated IgG1/IgG4 (black) and deglycosylated IgG1/IgG4 (magenta). The peak maxima at $M1$ and $M2$ and the maximum length $L$ are indicated. (A and B) The SAXS $P(r)$ curves for glycosylated and deglycosylated IgG1/IgG4 are shown across different temperatures 20$^\circ$C-60$^\circ$C. (C and D) The corresponding $P(r)$ curves for the SANS curves for glycosylated and deglycosylated IgG1 and IgG4 across different temperatures 20$^\circ$C-80$^\circ$C.
Figure 6.9 Temperature of the peak maxima from the distance distribution analyses $P(r)$ (legend overleaf).
Figure 6.9 Temperature dependence of the peak maxima from the distance distribution analyses $P(r)$ (continued).

The temperature dependence of the peak maxima $M1$ and $M2$ for glycosylated samples at 4.00 mg/ml and 1.00 mg/ml (●, ○), respectively, and deglycosylated samples at 4.00 mg/ml and 1.00 mg/ml (●, ○) respectively, are shown for data collected at different temperatures. The linear regression for glycosylated (solid line), and for deglycosylated (dashed lines). (A) Peak maxima for IgG1 X-ray data, and (C) Peak maxima for IgG4 X-ray data, both for data collected for different temperatures 20-60°C. For X-ray data, $M1$ and $M2$ peak are not reported for deglycosylated IgG4 measured at 65°C, where the IgG structure has been compromised and the $M1$ and $M2$ peaks are not defined. (B) Values corresponding to IgG1 neutron data and (D) IgG4 neutron data, for different temperatures 20-80°C. For neutron data, $M1$ and $M2$ peak not reported for higher temperatures (65°C-80°C) where IgG1 and IgGs structure has been completely compromised and the $M1$ and $M2$ peaks are not defined.
Figure 6.10 Normalised Kratky plots for the experimental and best fit glycosylated and deglycosylated IgG1 and IgG4 scattering curves. (A and B) X-ray experimental data for glycosylated in black and deglycosylated IgG1 and IgG4 in magenta, shown for data collected at different temperatures 20-60°C. (C and D) Neutron experimental data for glycosylated in black and deglycosylated IgG1 and IgG4 in magenta, shown for data collected at different temperatures 20-80°C.
Table 6.3 Experimental X-ray and neutron scattering data for glycosylated and deglycosylated IgG1.

<table>
<thead>
<tr>
<th>X-ray data</th>
<th>Temperature (°C)</th>
<th>Concentration (mg/ml)</th>
<th>(R_0) (nm)</th>
<th>(R_{XS-1}) (nm)</th>
<th>(R_{XS-2}) (nm)</th>
<th>(L) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IgG1 Glycosylated</strong></td>
<td>20</td>
<td>4.00</td>
<td>5.12 ± 0.51</td>
<td>2.51 ± 0.73</td>
<td>1.42 ± 0.94</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.03 ± 0.50</td>
<td>2.50 ± 0.73</td>
<td>1.38 ± 0.91</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>4.00</td>
<td>5.21 ± 0.52</td>
<td>2.49 ± 0.72</td>
<td>1.41 ± 0.93</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>4.99 ± 0.50</td>
<td>2.50 ± 0.72</td>
<td>1.34 ± 0.88</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.00</td>
<td>5.07 ± 0.51</td>
<td>2.49 ± 0.72</td>
<td>1.38 ± 0.91</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>4.95 ± 0.49</td>
<td>2.37 ± 0.69</td>
<td>1.08 ± 0.71</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>4.00</td>
<td>5.05 ± 0.50</td>
<td>2.45 ± 0.71</td>
<td>1.31 ± 0.87</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.01 ± 0.50</td>
<td>2.36 ± 0.68</td>
<td>1.07 ± 0.71</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>4.00</td>
<td>5.11 ± 0.51</td>
<td>2.44 ± 0.71</td>
<td>1.30 ± 0.86</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.04 ± 0.50</td>
<td>2.36 ± 0.69</td>
<td>1.07 ± 0.71</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td><strong>IgG1 Deglycosylated</strong></td>
<td>20</td>
<td>4.00</td>
<td>5.17 ± 0.52</td>
<td>2.53 ± 0.73</td>
<td>1.39 ± 0.92</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.07 ± 0.51</td>
<td>2.52 ± 0.73</td>
<td>1.37 ± 0.91</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>4.00</td>
<td>5.14 ± 0.51</td>
<td>2.52 ± 0.73</td>
<td>1.39 ± 0.92</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.09 ± 0.51</td>
<td>2.51 ± 0.73</td>
<td>1.33 ± 0.88</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.00</td>
<td>4.78 ± 0.48</td>
<td>2.39 ± 0.69</td>
<td>1.06 ± 0.70</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.15 ± 0.51</td>
<td>2.48 ± 0.72</td>
<td>1.28 ± 0.84</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>4.00</td>
<td>5.69 ± 0.57</td>
<td>2.67 ± 0.77</td>
<td>1.23 ± 0.81</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.28 ± 0.53</td>
<td>2.45 ± 0.71</td>
<td>1.00 ± 0.66</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neutron data</th>
<th>IgG1 Glycosylated</th>
<th>Temperature (°C)</th>
<th>Concentration (mg/ml)</th>
<th>(R_0) (nm)</th>
<th>(R_{XS-1}) (nm)</th>
<th>(R_{XS-2}) (nm)</th>
<th>(L) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>4.00</td>
<td>5.38 ± 0.54</td>
<td>2.40 ± 0.70</td>
<td>1.28 ± 0.85</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.25 ± 0.53</td>
<td>2.26 ± 0.65</td>
<td>1.42 ± 0.94</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>4.00</td>
<td>5.24 ± 0.52</td>
<td>2.39 ± 0.69</td>
<td>1.27 ± 0.84</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.09 ± 0.51</td>
<td>2.52 ± 0.73</td>
<td>1.33 ± 0.88</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.00</td>
<td>4.91 ± 0.49</td>
<td>2.44 ± 0.71</td>
<td>1.24 ± 0.82</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>4.85 ± 0.49</td>
<td>2.38 ± 0.69</td>
<td>1.24 ± 0.82</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>4.00</td>
<td>5.28 ± 0.53</td>
<td>2.46 ± 0.71</td>
<td>1.26 ± 0.83</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>5.53 ± 0.55</td>
<td>2.47 ± 0.72</td>
<td>1.32 ± 0.87</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>4.00</td>
<td>1.13 ± 0.12</td>
<td>3.35 ± 0.97</td>
<td>1.36 ± 0.90</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>3.31 ± 0.96</td>
<td>1.54 ± 1.02</td>
<td>42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IgG1 Deglycosylated</th>
<th>Temperature (°C)</th>
<th>Concentration (mg/ml)</th>
<th>(R_0) (nm)</th>
<th>(R_{XS-1}) (nm)</th>
<th>(R_{XS-2}) (nm)</th>
<th>(L) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>4.00</td>
<td>5.36 ± 0.56</td>
<td>2.40 ± 0.69</td>
<td>1.21 ± 0.80</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>5.61 ± 0.56</td>
<td>2.43 ± 0.71</td>
<td>1.49 ± 0.98</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>4.00</td>
<td>5.17 ± 0.52</td>
<td>2.37 ± 0.69</td>
<td>1.22 ± 0.81</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>5.85 ± 0.58</td>
<td>2.35 ± 0.68</td>
<td>1.29 ± 0.85</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.00</td>
<td>5.10 ± 0.51</td>
<td>2.42 ± 0.70</td>
<td>1.19 ± 0.78</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>4.30 ± 0.43</td>
<td>2.51 ± 0.73</td>
<td>1.31 ± 0.86</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>4.00</td>
<td>7.42 ± 0.74</td>
<td>3.37 ± 0.98</td>
<td>1.03 ± 0.68</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>6.74 ± 0.67</td>
<td>3.00 ± 0.87</td>
<td>1.22 ± 0.81</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>4.00</td>
<td>1.14 ± 0.12</td>
<td>3.78 ± 1.10</td>
<td>1.21 ± 0.80</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.16 ± 0.12</td>
<td>3.51 ± 1.02</td>
<td>1.44 ± 0.95</td>
<td>42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.4 Experimental X-ray and neutron scattering data for glycosylated and deglycosylated IgG4.

<table>
<thead>
<tr>
<th>SAXS data</th>
<th>Temperature (°C)</th>
<th>Concentration (mg/ml)</th>
<th>$R_g$ (nm)</th>
<th>$R_{X\beta}$-1 (nm)</th>
<th>$R_{X\beta}$-2 (nm)</th>
<th>$L$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG4 glycosylated</td>
<td>20</td>
<td>4.00</td>
<td>4.83 ± 0.48</td>
<td>2.53 ± 0.73</td>
<td>1.40 ± 0.93</td>
<td>15.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.78 ± 0.48</td>
<td>2.48 ± 0.72</td>
<td>1.34 ± 0.89</td>
<td>15.66</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>4.00</td>
<td>4.91 ± 0.49</td>
<td>2.51 ± 0.73</td>
<td>1.39 ± 0.92</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.78 ± 0.48</td>
<td>2.50 ± 0.73</td>
<td>1.35 ± 0.89</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.00</td>
<td>4.90 ± 0.49</td>
<td>2.45 ± 0.71</td>
<td>1.29 ± 0.85</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.82 ± 0.48</td>
<td>2.37 ± 0.69</td>
<td>1.08 ± 0.71</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>4.00</td>
<td>4.93 ± 0.49</td>
<td>2.72 ± 0.92</td>
<td>1.40 ± 0.93</td>
<td>15.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.87 ± 0.49</td>
<td>2.39 ± 0.69</td>
<td>1.27 ± 0.84</td>
<td>16</td>
</tr>
<tr>
<td>IgG4 Deglycosylated</td>
<td>20</td>
<td>4.00</td>
<td>5.01 ± 0.50</td>
<td>2.53 ± 0.73</td>
<td>1.39 ± 0.92</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.68 ± 0.47</td>
<td>2.52 ± 0.73</td>
<td>1.33 ± 0.88</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>4.00</td>
<td>4.92 ± 0.49</td>
<td>2.52 ± 0.73</td>
<td>1.37 ± 0.91</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.68 ± 0.47</td>
<td>2.51 ± 0.73</td>
<td>1.32 ± 0.87</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.00</td>
<td>4.91 ± 0.49</td>
<td>2.47 ± 0.72</td>
<td>1.26 ± 0.83</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.65 ± 0.46</td>
<td>2.35 ± 0.68</td>
<td>0.97 ± 0.64</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>4.00</td>
<td>6.29 ± 0.63</td>
<td>3.09 ± 0.90</td>
<td>1.10 ± 0.73</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>5.45 ± 0.55</td>
<td>2.66 ± 0.77</td>
<td>0.83 ± 0.55</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SANS data</th>
<th>Temperature (°C)</th>
<th>Concentration (mg/ml)</th>
<th>$R_g$ (nm)</th>
<th>$R_{X\beta}$-1 (nm)</th>
<th>$R_{X\beta}$-2 (nm)</th>
<th>$L$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG4 Glycosylated</td>
<td>20</td>
<td>4.00</td>
<td>5.11 ± 0.51</td>
<td>2.39 ± 0.69</td>
<td>1.19 ± 0.78</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>5.24 ± 0.52</td>
<td>2.34 ± 0.68</td>
<td>1.52 ± 1.00</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>4.00</td>
<td>4.90 ± 0.49</td>
<td>2.36 ± 0.68</td>
<td>1.21 ± 0.80</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>5.73 ± 0.57</td>
<td>2.28 ± 0.66</td>
<td>1.30 ± 0.85</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.00</td>
<td>4.92 ± 0.49</td>
<td>2.40 ± 0.70</td>
<td>1.22 ± 0.80</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.16 ± 0.42</td>
<td>2.52 ± 0.73</td>
<td>1.09 ± 0.72</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>4.00</td>
<td>7.36 ± 0.74</td>
<td>3.18 ± 0.92</td>
<td>1.05 ± 0.69</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>6.51 ± 0.65</td>
<td>2.89 ± 0.84</td>
<td>1.30 ± 0.86</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>4.00</td>
<td>6.28 ± 0.63</td>
<td>3.09 ± 0.90</td>
<td>1.10 ± 0.73</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>5.45 ± 0.55</td>
<td>2.66 ± 0.77</td>
<td>0.83 ± 0.55</td>
<td>19</td>
</tr>
<tr>
<td>IgG4 Deglycosylated</td>
<td>20</td>
<td>4.00</td>
<td>5.10 ± 0.51</td>
<td>2.40 ± 0.70</td>
<td>1.22 ± 0.80</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>5.52 ± 0.55</td>
<td>2.38 ± 0.69</td>
<td>1.24 ± 0.10</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>4.00</td>
<td>5.22 ± 0.52</td>
<td>2.36 ± 0.68</td>
<td>1.17 ± 0.77</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>5.70 ± 0.57</td>
<td>0.66</td>
<td>1.18 ± 0.20</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.00</td>
<td>4.97 ± 0.50</td>
<td>2.48 ± 0.72</td>
<td>1.12 ± 0.74</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>4.86 ± 0.59</td>
<td>2.42 ± 0.70</td>
<td>1.10 ± 0.74</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>4.00</td>
<td>8.65 ± 0.87</td>
<td>4.04 ± 1.17</td>
<td>0.88 ± 0.58</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>6.16 ± 0.36</td>
<td>2.74 ± 0.79</td>
<td>1.05 ± 0.69</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>4.00</td>
<td>1.22 ± 1.05</td>
<td>3.83 ± 1.11</td>
<td>1.43 ± 0.94</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.00</td>
<td>1.22 ± 1.05</td>
<td>3.83 ± 1.11</td>
<td>1.43 ± 0.94</td>
<td>51</td>
</tr>
</tbody>
</table>
sample has oligomerised (Figure 6.8B). For glycosylated IgG4 at concentrations of 4.00 mg/ml and 1.00 mg/ml, the M2 peak shifted upward from 7.26 nm at 20°C to 7.48 nm at 60°C, and from 7.73 nm at 20°C to 8.50 nm at 60°C, respectively. Again, a concentration dependence is present that seems to stabilise the IgG4 structure (Figure 6.9C). For deglycosylated IgG4 at concentrations of 4.00 mg/ml and 1.00 mg/ml, the M2 peak shifted downward from 7.60 nm at 60°C to 7.31 nm at 50°C for 4.0 mg/ml and from 8.00 nm at 20°C to 7.91 nm at 50°C for 4.00 mg/ml, this indicates that the structure is compacting slightly before oligomerising at 55°C. plots derived from neutron data are similar to those from X-ray data. At 65°C there is only one clear peak for both glycosylated and deglycosylated IgG4 at 1.90 nm, indicating that the protein is unfolding and becoming more disordered. At 80°C the protein is uninformative as the sample has aggregated (Figure 6.10D).

The IgG4 Kratky plots demonstrated two clear peaks at Q.RG values of 1.89 nm and 4.01 nm for glycosylated and 1.90 nm and 4.15 nm deglycosylated (Figure 6.10B and D). These peaks are intact up to approximately 50°C. However, at 60°C the second Kratky peak for deglycosylated IgG4 shifts to 5.79 nm but the glycosylated Kratky plot is intact at this temperature (Figure 6.10B).

### 6.4 Discussion

In comparison to traditional thermostability studies which focussed on melting curves derived from intrinsic fluorescence (Bowden et al., 2012; Wada et al., 2019; Zhang et al., 2020), the combination of thermostability data with small angle scattering data at different temperatures provides a novel insight into the thermodynamics and structural events in antibody unfolding. This study used a combined structural approach to follow the thermal melting using intrinsic fluorescence (BCM) to derive melting temperatures, SLS to derive the aggregation temperatures, and SAXS and SANS to monitor several structural parameters as the temperature was increased. The key parameters from these experiments were the Guinier plots, the distance distribution curves P(r) curves and Kratky plots which provide a measure of structural flexibility. Together, these have clarified the role that the Fc glycan plays on the Fc region and the global structure of IgG1 and IgG4.
The complete deglycosylation of IgG1 and IgG4 was validated by a combination of gel filtration and routine mass spectrometry. Analyses of thermal melting experiments on the UNit identified $T_{m, \text{app} 1}$ as the first major unfolding transition, postulated to be indicative of Fc unfolding (Zhang et al., 2020), and revealed a -8.6°C decrease in $T_{m, \text{app} 1}$ upon deglycosylation of IgG1, and a -6.4°C decrease upon deglycosylation of IgG4. This significant difference for both IgG1 and IgG4 further demonstrates that $T_{m, \text{app} 1}$ is linked to Fc unfolding. Zhang et al. (2020) suggested that $T_{m, \text{app} 2}$ could be linked to the unfolding of the Fab region. They found that the $T_{m, \text{app}}$ of the isolated Fab region was 78.2 ± 0.2 °C, and were therefore were able to attribute the $T_{m, \text{app} 2}$ value observed in the full-length IgG1 to the Fab region. Despite an abundance of SAXS and SANS data it was not possible to deconvolute the Fab and Fc regions in the analysis. In this study a difference in $T_{m, \text{app} 2}$ of 3.8°C was found between glycosylated and deglycosylated IgG1, and a difference in $T_{m, \text{app} 2}$ of 0.5°C between glycosylated and deglycosylated IgG4. In contrast to IgG1, there is a smaller difference between glycosylated and deglycosylated IgG4 melting temperatures, is consistent with previous finding that the IgG4 Fc glycan has a more limited role in maintaining the spatial conformation of the Fc (Chapter 5, Spiteri et al., 2021a). In a similar previous study for IgG1 the Fc glycan was found to play a key role in restricting the spatial conformation of Fc, and this provided an explanation for the increased binding capacity of native IgG1 to FcγRs and C1q (Chapter 4, Spiteri et al., 2021b). Given the central role that the Fc glycan plays in the IgG1 structure it follows that the global structure would be more easily destabilised upon glycan removed.

Despite changes in $T_{m, \text{app} 1}$ and $T_{m, \text{app} 2}$ values, the mean $T_{\text{agg}}$ remained more consistent for both glycosylated and deglycosylated IgG1 and IgG4. However, difference in aggregation potential could be seen between the glycosylated and deglycosylated forms by SAXS and SANS, which may indicate that SLS is not as sensitive to early catching aggregation. The aggregation potential of IgG1 and IgG4 could also be followed by the distance distribution $P(r)$ curves derived from SAXS and SANS, by analysing at the maximum length ($L$) of the structure. $L$ is effected by the concentration of oligomeric species, thus an increase $L$ due to unfolding is difficult to deconvolve from simultaneous oligomerisation, indeed this is a technical hurdle in thermal melting studies.
However, the maximum length parameter can still be informative when considering it as a relative parameter to make comparisons between glycosylated and deglycosylated samples. As the temperature was increased deglycosylated IgG1 lost its quaternary structural arrangement and began to elongate at a lower temperature (65°C) than its glycosylated counterpart, which was still intact (Figure 6.8C). In IgG4, the deglycosylated structure also starts to elongate at a lower temperature, at approximately 60°C, than glycosylated IgG4, which elongates at 65°C. Another informative parameter that can be extracted from the $P(r)$ is the position of the $M1$ and $M2$ peaks that monitors the mean separation of the Fab and Fc regions. For deglycosylated IgG1 these peaks shifted position more readily (Figure 6.9A and B). Demonstrating that internal Fc structure and the global IgG1 structure was being significantly affected by the removal of the glycans. This was less apparent in IgG4, and the $M2$ peak appeared to shift downward, indicating that the Fab-Fab/Fab-Fc domains move closer to one another possibly due to unfolding events (Figure 6.9C and D). This was also seen in the cross-sectional distances, $R_{XS-1}$ and $R_{XS-2}$, derived from Guinier analysis (Figure 6.7C).

Parameters derived from Guinier analysis, primarily the $R_G$, revealed that deglycosylated IgG1 and IgG4 oligomerised at a lower temperature, at approximately 60°C for IgG1 and 55°C for IgG4. The temperatures at which $R_G$ and $L$ increase corresponded more closely to $T_{m,app 1}$, indicating that elongation and oligomerisation of the sample is predominately mediated by the first transition, i.e. changes in the Fc region. It should be noted that due to limitations in beam time, larger temperatures increments were used, 5°C in SAXS and 15°C in SANS, therefore it was difficult to achieve the same resolution as in the UNit experiments which had intervals of 1°C. Furthermore, to validate the reproducibility of the data and to better understand the effect of concentration on thermostability of IgG1 and IgG4, data at more concentrations should be collected. This study only presents two concentrations 4 mg/ml and 1 mg/ml, due to limitations in sample availability and beamtime limitations.

This study has demonstrated that despite IgG1 and IgG4 having high sequence and structure homology, IgG1 is significantly more thermostable than IgG4. With a difference in $T_{m,app 1}$ and $T_{m,app 2}$ of 7.1°C and 8.1°C, respectively, between glycosylated IgG1 and IgG4, these are significant changes despite very...
high sequence similarity for IgG1 and IgG4 A33, being 96.05% similar (Figure 6.2). In previous structural studies it was found that the IgG4 Fc (Spiteri et al., 2021a, Chapter 5) is less conformationally restricted than that for IgG1 (Spiteri et al., 2021b, Chapter 4) and IgG3 (Chapter 7), this is possibly due to the IgG4 hinge being shorter allowing for movements in the Fab regions to force concurrent movements in the IgG4-Fc region and this provided an explanation for its reduced receptor binding. The combined structural approach applied in this study was able to confirm findings in previous studies and demonstrate the Fc glycan critical for maintaining the stability of the IgG1 and IgG4 structure. For this study we show that a shorter hinge is less stable than a longer hinge, while conformational and thermal stability are correlated.
Chapter 7 Solution structures of human myeloma IgG3 immunoglobulin reveal extended Fab and Fc regions relative to the other IgG subclasses: implications for function
7.1 Introduction

Immunoglobulin IgG3 is one of the four subclasses IgG1-IgG4 that make up the human IgG antibodies. It is the third most abundant IgG in serum and comprises 5-8% of serum IgG. IgG3 plays a role in protection against intracellular bacteria, parasites and viruses. IgG3 is unique in the four subclasses, as it is the only subclass which has not as yet been exploited as a biotherapeutic. It is of particular interest for reason of its elongated hinge with 62 amino acids and 11 disulphide bonds that connects the Fab and Fc regions (Figure 7.1A), which potentially gives the Fab and Fc regions in IgG3 a greater rotational freedom than in IgG1, IgG2 and IgG4 which have shorter hinges of lengths 15, 12 and 12 residues respectively (Damelang et al., 2019). IgG3 mediates a broad range of effector functions, including being the most effective activator of complement through the binding of C1q and able to initiate antibody effector functions such as antibody dependent cell-mediated cytotoxicity. The superior ability of IgG3 to bind to FcγR receptors and C1q was thought to be due to its elongated hinge (Richardson et al., 2019), however this can also be attributed to variation in the IgG3 sequence in the C\(_{H2}\) region (de Taeye et al., 2020). Of all the IgG subclasses, IgG3 has the most known allotypes, with about 15 in total that correspond to polymorphisms in its constant regions that vary the hinge length or the Fc sequence (de Taeye et al., 2020). Some IgG3 allotypes have a reduced half-life of about seven days, compared to the longer 21-day half-life for IgG1, IgG2 and IgG4 (Damelang et al., 2019). This reduced half-life is attributed to a His\(^{435}\)Arg substitution in the Fc region (EU numbering) (Stapleton et al., 2011). This His\(^{435}\)Arg substitution is also critical as it blocks IgG3 from binding to a protein A column, thus allowing for IgG3 to be purified from IgG1, IgG2 and IgG4 in human serum (van Loghem et al., 1982).

The molecular structure of full-length human IgG3 is poorly understood, given that its elongated hinge has made structural studies difficult using methods such as X-ray crystallography. Nonetheless a high resolution structure for IgG3-Fc (PDB ID: 5W38) (Shah et al., 2017) is available. Several solution studies of full-length IgG3 using small angle X-ray scattering (SAXS) and analytical ultracentrifugation (AUC) have been reported (Gregory et al., 1987; Lu et al., 2007; Liu et al., 2019). These studies showed that IgG3 has an elongated
structure compared with IgG1, IgG2 and IgG4. By SAXS, IgG3 was found to have a more flexible hinge than those of IgG1 and IgG2 (Liu et al., 2019). Structural predictions of the hinge have suggested that its length should be around 9 nm (Pumphrey, 1986). IgG3 also showed partial O-glycosylation at the IgG3 hinge region (Plomp et al., 2015). All four human IgG subclasses have a conserved N-linked glycosylation site in the Fc region which plays a key functional role (Figure 1B). A complex-type biantennary glycan with a Man3GlcNAc2 core and two NeuNAc.Gal.GlcNAc antennae is attached at Asn297 (EU numbering) on each C\(\text{H}2\) domain (Deisenhofer, 1981) (Figure 7.2). This glycan is chemically heterogenous (Wuhrer et al., 2007) and influences the interaction with Fc\(\gamma\)R receptors and complement C1q binding (Jefferis, 2007). The structural role of the two Fc glycans in IgG3 is not well understood, but deglycosylated IgG3 was shown to have impaired binding to Fc\(\gamma\)R receptors, leading to the conclusion that Fc\(\gamma\)R binding depended on the conformation of the Fc region in IgG3 (Walker et al., 1989).

The issues of the unknown IgG3 molecular structure and its glycans were here addressed by a joint application of AUC, SAXS, small angle neutron scattering (SANS) and atomistic modelling to myeloma human IgG3. This powerful solution structural strategy has recently provided molecular structures for glycosylated and deglycosylated monoclonal IgG1 and IgG4, and also for human myeloma IgG2 (Hui et al., 2019; Spiteri et al., 2021b, 2021a). AUC provides overall solution structures for IgG3 as well as confirming its monodispersity in solution. Scattering provides low resolution structural data on IgG3. Here, SAXS provides data sets measured in high positive solute-solvent contrast, in which the contribution of the hydrophilic surface regions of the glycoprotein are accentuated, while SANS measured with heavy water buffers provides data sets measured in high negative solute-solvent contrast, in which the contribution of the buried hydrophobic core of the glycoprotein is accentuated (Perkins, 1986, 2001; Svergun et al., 1998). The tightly-bound hydration layer is detected by SAXS because its electron density is similar to that of the protein and not to bulk water, while this same hydration layer is almost invisible by SANS measured in heavy water, because its nuclear density is almost the same as that of bulk water. The reproducibility of the SAXS and SANS data sets confirms the outcome of both methods, because experimental artefacts from radiation effects...
in SAXS and aggregation in heavy water by SANS may perturb the outputs of either method. SAXS and SANS data sets can now be represented as molecular structures by the relatively recent development of atomistic modelling using molecular dynamics and Monte Carlo methods to fit the scattering curves (Perkins et al., 2016). Here, this joint SAXS-SANS-AUC approach was applied to determine experimentally the first full-length IgG3 molecular structure for myeloma IgG3 in order to explain better its immune function. The effect of the removal of the Fc glycans on the molecular structure of full-length IgG3 was also discussed.

7.2 Materials and Methods

7.2.1 Purification and composition of IgG3

IgG3 κ HP3 (164.8 kDa) was prepared by Dr Margaret Goodall (University of Birmingham) from human myeloma serum stored at -20°C. The serum was thawed at 37°C, and spun at 3000 rpm for 15 minutes. 12.5 ml of clear serum was dialysed into 1 litre of 0.01M phosphate buffer at pH 7.5, repeated three times. IgG was isolated from other serum components using a DE 52 Whatman (DEAE) column (12 cm x 12 cm), in 0.01 M phosphate buffer pH 7.7, varying the salt concentration sequentially (0.2, 0.3, 0.4, 0.05 and 0.06 M). The most IgG3 was found from the salt concentration eluate of 0.01 M. IgG3 was further purified using a 1 ml pre-packed Staphylococcus Protein A coupled to a Sepharose column (Cytiva, Amersham, UK). Staphylococcus Protein A binds IgG1, IgG2 and IgG4 but not IgG3 which flowed through the column. The purity of IgG3 following purification from serum was confirmed using SDS-PAGE.

To remove the N-glycans from IgG3, enzymatic deglycosylation was achieved using peptide:N-glycosidase F (PNGase F) (35.5 kDa, New England Biolabs, Massachusetts, USA), selected due to its ability to remove glycans completely from the glycosylated Asn residues (Plummer et al., 1984). Thus 20 μl PNGase F (1850 activity units) was used in excess to deglycosylate 1000 μl of IgG3 (8.8 mg/ml) by incubation at 37°C for 10 hours. Each deglycosylated IgG3 sample was concentrated using Amicon Ultra-0.5 ml centrifugal filters (50 kDa cut-off) which simultaneously allowed the PNGase F to pass through
Figure 7.1 The human IgG3 domains and its glycosylation (legend overleaf).
**Figure 7.1. The human IgG3 domains and its glycosylation (continued).**

A The heavy chains are comprised of V_H, C_H1, C_H2 and C_H3 domains, and the light chains are comprised of V_L and C_L domains. The heavy chains are connected in the hinge region by eleven Cys-Cys disulphide bridges at Cys^{228}, Cys^{231}, Cys^{237}, Cys^{243}, Cys^{246}, Cys^{252}, Cys^{258}, Cys^{261}, Cys^{267}, Cys^{273} and Cys^{276} as shown. An N-linked oligosaccharide at Asn^{297} (EU numbering) or Asn^{344} (continuous numbering) is present on each of the C_H2 domains. The hinge region connecting the Fab and Fc regions was constructed from 70 residues \(^{216}ELKTPLGDTTHTCPRCP(EPKSCDTPPCPRCPRCP)\_3APELLGGP^{285}.

B On the left, the IgG3 Fc glycans at both Asn^{297} (Asn^{344}) residues in the Fc region are shown as stick models. The three hinge peptides that were conformationally varied in the TAMC searches for glycosylated models are shown in orange. Those for the deglycosylated models are shown in yellow. The central schematic shows the glycosylation pattern used in this study (\(\text{N-acetyl glucosamine, GlcNAc;}\) \(\text{mannose, Man;}\) \(\text{galactose, Gal;}\) \(\text{N-acetyl neuraminic acid, NeuNAc.}\) At the right, the detailed view of a single C_H2 domain with its glycan chain is shown, with the glycan colours coordinated with those in the schematic.
the membrane. Immediately before AUC, SAXS and SANS data collection, measurements, glycosylated and deglycosylated IgG3 were further purified by gel filtration to remove any non-specific aggregates using a Super 6 Increase 10/300 GL column (Cytiva, Amersham, UK), then concentrated using Amicon Ultra-15 spin concentrators (50 kDa cut-off) and dialyzed at 4°C into 20 mM L-histidine, 138 mM NaCl, and 2.6 mM KCl buffer, pH 6.0. This histidine buffer improved the stability of IgG3 in solution.

As no sequences were available for myeloma IgG3, a hypothetical sequence was created based on the sequence of the IgG4 A33 Fab regions (Spiteri et al., 2021a) and the Fc sequence from the IgG3 Fc crystal structure (PDB ID: 5W38) (Shah et al., 2017). The hinge sequence was assumed to be ELKTPLGDTTHTCPRPCP(EPKSCDTPPPCPRPCP)_3APELLGGP. The assumed IgG3 Fab sequence was aligned with the IgG4 b72.3 crystal structure (PDB ID: 1BBJ) (Brady et al., 1992) in order to build an IgG3 A33 Fab structure (Figure 7.2). This multiple sequence alignment was generated using Clustal Omega software (Goujon et al., 2010). The N-linked glycan at Asn297 on the C_H2 domains were approximated to be complex-type biantennary oligosaccharides with a Man3GlcNAc2 core and two NeuNAc.Gal.GlcNAc antennae (Deisenhofer, 1981). From this sequence, the molecular mass of glycosylated IgG3 was calculated to be 158.1 kDa, its unhydrated volume was 203.7 nm³, its hydrated volume was 245.9 nm³, its partial specific volume $\bar{\rho}$ was 0.776 ml/g and its absorption coefficient was 12.4 cm⁻¹. The molecular mass of deglycosylated IgG3 was calculated to be 154.0 kDa, its unhydrated volume was 199.4 nm³, its hydrated volume was 241.5 nm³, its $\bar{\rho}$ was 0.780 ml/g and its absorption coefficient was 12.8 cm⁻¹ (Perkins, 1986). The X-ray and neutron scattering densities of glycan residues are similar to those for hydrophilic (polar) amino acid residues, these being slightly higher than those for hydrophobic (non-polar) amino acid residues (Perkins, 1986). The buffer density was measured on an Anton Paar DMA 5000 density meter at 20°C to be 1.00578 g/ml in light water and 1.11106 g/ml in heavy water. Buffer viscosities were measured on an Anton Paar AMVN Automated microviscometer at 20°C to be 0.010190 and 0.01384 poise respectively for light and heavy water, pH 6.0.
Figure 7.2 Sequence alignment of IgG3 (legend overleaf).
Figure 7.2. Sequence alignment of IgG3 (continued).

(A-G) Given that no IgG3 sequence was available, the IgG3 Fab sequence was taken to be that from IgG4 A33 Fab. Its atomistic modelling was based on the IgG4 b72.3 crystal structure (PDB ID: 1BBJ) whose sequence is included. The IgG3 hinge and Fc sequences were taken from its crystal structure (PDB ID: 5W38). Beneath the alignments, (*) indicates full conservation, (:) indicates conservation between groups of strongly similar properties based on the Gonnet PAM 250 matrix, (.) indicates conservation between groups of weakly similar properties, and a space indicates no conservation. A, B, the V\textsubscript{L} and C\textsubscript{L} domains; C-E, the V\textsubscript{H} and C\textsubscript{H1} domains and the hinge, with the TAMC-varied peptides for glycosylated IgG3 are shown in orange and for deglycosylated IgG3 shown in yellow. In A, C, the three complementarity-determining regions (CDRs) sequences are in red. F, G, the C\textsubscript{H2} and C\textsubscript{H3} domains, with Asn\textsuperscript{297} (Asn\textsuperscript{344}) in magenta. In E-G, the standard EU residue numbering that matches IgG1, IgG2 and IgG4 is shown in grey, and the continuous numbering is shown in black.
The completeness of deglycosylation was verified by Superose 6 gel filtration, SDS-PAGE, and mass spectrometry. In the Mass Spectrometry Facility at the Chemistry Department University College London, the antibodies were analysed on an Agilent 6510 Quadrupole time-of-flight liquid chromatography mass spectrometry system (Agilent, UK). Ten µL of each sample was injected onto a PLRP-S, 1000A, 8 µM, 150 mm × 2.1 mm column, which was maintained at 60°C at a flow of 0.3 ml/min. The separation was achieved using mobile phases A (water with 0.1% formic acid) and B (acetonitrile, with 0.1% formic acid) using a gradient elution. The column effluent was continuously electrospayed into the capillary electrospray ionization source of the Agilent 6510 QTOF mass spectrometer and electrospray ionization mass spectra were acquired in positive electrospray ionisation mode using the m/z range 1,000–3200 in profile mode. The raw data was converted to zero charge mass spectra using the maximum entropy deconvolution algorithm in the MassHunter software version B.07.00 (Agilent, Stockport, UK). The single glycan mass was found by subtracting the mass of fully glycosylated IgG3 from that for deglycosylated IgG3 and halving this mass.

7.2.2 Sedimentation velocity data and analysis of IgG3

Analytical ultracentrifugation data for glycosylated and deglycosylated IgG3 were obtained on two Beckman XL-I instruments equipped with AnTi50 rotors. Data were collected at 20°C, at a rotor speed of 30,000 rpm in two-sector cells with column heights of 12 mm for approximately 6 hours. Sedimentation analyses were performed using direct boundary Lamm fits of up to 600 scans using SEDFIT (version 15.01b) (Schuck, 1998, 2000). SEDFIT resulted in size-distribution analyses c(s), for which the algorithm assumes that all species have the same frictional ratio f/f0. The final SEDFIT analyses used a fixed resolution of 200 and optimized the c(s) fits by floating f/f0 and the baseline until the overall root mean square deviations and visual appearance of the fits were satisfactory. The percentage of oligomers in the total loading concentration was derived using the c(s) integration function. The buffer viscosity and density were the experimental values. Values were corrected to s20,w by:

\[ s_{20,w} = s_{T,B} \left( \frac{\eta_{T,B}}{\eta_{20,w}} \right) \left( 1 - \bar{v} \rho \right)_{20,w} \left( 1 - \bar{v} \rho \right)_{T,B} \]
where $s$ is the sedimentation coefficient, the subscripts $T,B$ refers to the temperature of the buffer and $^{20,w}$ refers to water at 20°C, $\rho$ is the solvent density, $\eta$ is the solvent viscosity and $\bar{v}$ is as above.

### 7.2.3 X-ray and neutron scattering data and analyses for IgG3

X-ray scattering data was obtained during one beam session (July 2019) on Instrument B21 (Cowieson et al., 2020) at the Diamond Light Source at the Rutherford Appleton Laboratory (Didcot, UK), operating with a ring energy of 3 Gev, and a beamline operational energy of 12.4 keV. A PILATUS 2M detector with a resolution of $1475 \times 1679$ pixels (pixel size of $172 \times 172 \mu m$) was used with a sample-to-detector distance of 4.01 m giving a Q range from $0.04 \text{ nm}^{-1}$ to $4 \text{ nm}^{-1}$ (where $Q = 4 \pi \sin \theta / \lambda$; $2\theta$ = scattering angle; $\lambda$ = wavelength). The glycosylated (0.50-2.46 mg/ml) and deglycosylated (0.50-1.00 mg/ml) IgG3 samples in light water buffer were loaded onto a 96 well plate and placed into an EMBL Arinax sample holder (Pernot et al., 2013; Round et al., 2015). This measurement condition showed the antibody molecule as a hydrated structure in a high positive solute-solvent contrast (Perkins, 1986). An automatic sampler injected 30 µl of sample from the plate into a temperature-controlled quartz cell capillary with a diameter of 1.5 mm. Data sets of 30 frames with a frame exposure time of 1 second each were acquired in duplicate as a control of reproducibility. Checks during data acquisition confirmed the absence of radiation damage. An inhouse B21 pipeline termed DAWN (Basham et al., 2015) was used for buffer subtraction and data reduction, in which the 30 frames were averaged.

Neutron scattering data on glycosylated (0.70-1.85 mg/ml) and deglycosylated (0.65-1.21 mg/ml) IgG3 samples in heavy water buffer were obtained in one session (October 2020) on instrument SANS2D at the ISIS pulsed neutron source at the Rutherford Appleton Laboratory (Didcot, UK) (Heenan et al., 2011). This condition showed the antibody structure with a near-invisible hydration shell in a high negative solute-solvent contrast (Perkins, 1986). No conformational differences in the antibody between light and heavy water were detected in this study or previously (Wright et al., 2019). A pulsed neutron beam was derived from proton beam currents of $\sim40 \mu A$. SANS2D data were recorded with 4 m of collimation, a 4 m sample-to-detector distance, a 12 mm sample aperture, and a wavelength range of 0.175-1.65 nm made available by a
time of flight. This gave a $Q$ range from 0.05 nm$^{-1}$ to 4 nm$^{-1}$. The data were acquired using a two-dimensional $^3$He detector with 512 x 512 pixels of 7.5 x 7.5 mm$^2$ in size. Samples of volume 1 ml were measured in 2 mm path length circular banjo cells for 1-7 h in a thermostatted sample rack at 20$^\circ$C. Data were reduced using MANTID software (Arnold et al., 2014). The MANTID data reduction steps include corrections for the $Q$ resolution, i.e. beam divergence effects and smearing from the shape and size of the slits, as well as the wavelength overlap in each pulse (Arnold et al., 2014).

Guinier analyses of the scattering data give information of the radius of gyration $R_G$, the cross-sectional radius of gyration $R_{XS}$, and the molecular mass. The scattering curve $I(Q)$ intensities at low $Q$ are defined by the $R_G$ value which is the averaged distance of each scattering point from the centre of scattering. In a given solute-solvent contrast, the radius of gyration $R_G$ is a measure of structural elongation if the internal inhomogeneity of scattering densities within the protein has no effect. Guinier analyses at low $Q$ gave the $R_G$ value and the forward scattering at zero angle $I(0)$ (Glatter et al., 1982):

$$\ln I(Q) = \ln I(0) - \frac{R_G^2 Q^2}{3}$$

For antibodies, this expression is valid in a $Q, R_G$ range up to 1.5, and was used in previous studies (Rayner et al., 2014, 2015; Wright et al., 2019), although the usual upper range reported in the literature is 1.0-1.3. If the structure is elongated, the mean radius of gyration of the cross-sectional structure $R_{XS}$ and the mean cross-sectional intensity at zero angle $[I(Q)Q\rightarrow 0$ is obtained from (Glatter et al., 1982):

$$\ln[I(Q)Q] = [I(Q)Q]_{Q\rightarrow 0} - \frac{R_{XS}^2 Q^2}{2}$$

For the immunoglobulins IgG1, IgG2 and IgG4, it has been long recognised that the cross-sectional plot exhibits two regions, a steeper innermost one and a flatter outermost one (Pilz et al., 1970), and the two analyses are denoted by $R_{XS,1}$ and $R_{XS,2}$ respectively. The $R_{XS,1}$ parameter represents the averaged overall spatial separation of the Fab and Fc regions, this being one shorter Fab-Fab separation and two longer Fab-Fc separations (Figure 7.1A), while the $R_{XS,2}$ parameter represents the averaged spatial cross-section of each of the two Fab and one Fc regions. The $R_G$ and $R_{XS}$ analyses were performed using SCT (Table 7.1)
Table 7.1 Experimental data by X-ray and neutron scattering and analytical ultracentrifugation for glycosylated and deglycosylated IgG3.

<table>
<thead>
<tr>
<th></th>
<th>Concentration (mg/ml)</th>
<th>$R_G$ (nm)</th>
<th>$R_{XS-1}$ (nm)</th>
<th>$R_{XS-2}$ (nm)</th>
<th>$L$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X-ray data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG3 glycosylated</td>
<td>2.46</td>
<td>6.98 ± 0.62</td>
<td>1.41 ± 0.17</td>
<td>1.60 ± 0.10</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>6.99 ± 0.63</td>
<td>1.41 ± 0.18</td>
<td>1.61 ± 0.10</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>6.96 ± 0.63</td>
<td>1.41 ± 0.15</td>
<td>1.61 ± 0.22</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>6.99 ± 0.82</td>
<td>1.41 ± 0.19</td>
<td>1.60 ± 0.27</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>6.84 ± 0.90</td>
<td>1.32 ± 0.27</td>
<td>1.61 ± 0.31</td>
<td>28</td>
</tr>
<tr>
<td>IgG3 deglycosylated</td>
<td>1.00</td>
<td>7.05 ± 0.71</td>
<td>1.41 ± 0.19</td>
<td>1.59 ± 0.15</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>6.98 ± 1.06</td>
<td>1.41 ± 0.29</td>
<td>1.58 ± 0.25</td>
<td>28</td>
</tr>
<tr>
<td><strong>Neutron data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG3 glycosylated</td>
<td>1.85</td>
<td>6.26 ± 1.07</td>
<td>0.89 ± 0.30</td>
<td>1.15 ± 0.21</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>6.20 ± 1.33</td>
<td>0.71 ± 0.31</td>
<td>0.97 ± 0.24</td>
<td>25</td>
</tr>
<tr>
<td>IgG3 deglycosylated</td>
<td>1.21</td>
<td>6.24 ± 1.58</td>
<td>1.10 ± 0.35</td>
<td>1.22 ± 0.27</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>6.33 ± 1.64</td>
<td>1.00 ± 0.30</td>
<td>1.02 ± 0.43</td>
<td>27</td>
</tr>
<tr>
<td><strong>AUC data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG3 glycosylated in light water</td>
<td>0.39</td>
<td>5.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>5.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>5.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG3 deglycosylated in light water</td>
<td>0.46</td>
<td>6.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>6.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>6.33</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Standard error of the mean for $s_{20,w}$ data as follows; ± 0.04 S for glycosylated and ± 0.03 S for deglycosylated.
(Wright et al., 2015). The $Q$ ranges for the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values were 0.10–0.22, 0.22–0.28, and 0.50–1.10 nm$^{-1}$ respectively, these being distinct from IgG1, IgG2 and IgG4 (Rayner et al., 2014, 2015; Hui et al., 2019; Wright et al., 2019; Spiteri et al., 2021b, 2021a). Indirect transformation of the scattering data $I(Q)$ in reciprocal space into real space to give the distance distribution function $P(r)$ was carried out using GNOM (version 5.0) (Semenyuk et al., 1991; Svergun, 1992).

$$P(r) = \frac{1}{2\pi^2} \int_0^\infty I(Q)Qr \sin(Qr) \, dQ$$

$P(r)$ corresponds to the distribution of distances $r$ between the volume elements in the macromolecule. This yields the maximum dimension of the macromolecule $L$ and its most commonly occurring distance vector $M$ in real space. For this $P(r)$ analysis, the X-ray $I(Q)$ curve utilized up to 788 data points in the $Q$ range between 0.09 and 1.50 nm$^{-1}$ for both glycosylated and deglycosylated IgG3. The neutron $P(r)$ curve utilized up to 134 $I(Q)$ data points in the $Q$ range between 0.165 and 1.60 nm$^{-1}$ for both glycosylated and deglycosylated IgG3.

### 7.2.4 Atomistic modelling of IgG3

Starting structures were created for each of glycosylated and deglycosylated IgG3 based on the sequence of Figure 7.2. The Fab structure (Figure 7.1) was based on the chimeric IgG4 b72.3 crystal structure (PDB ID: 1BBJ) (Brady et al., 1992) and the Fc structure was based on the serum-derived IgG3 antibody Fc crystal structure (PDB ID: 5W38) (Davies et al., 2014). Modeller (version 9.19) (Webb et al., 2016) was used to generate the full-length IgG3 structure by substituting the A33 Fab sequence into the chimeric IgG4 b72.3 structure. The IgG3 hinge region was built using a PyMOL script build_seq (PyMOL Script Repository, Queen’s University, Ontario, Canada), based on the 216ELKTPLGDTTHTCPRCP(EPKSCDTTPPCPRCP)$\beta$APELLGGP285 sequence.

This 62-residue hinge structure was subjected to energy minimisation and molecular dynamics (MD) independently of the Fab and the Fc regions in order to identify hinge conformations. Hinge structures were prepared for all-atom molecular dynamics simulations using Glycan Reader and Modeler (Jo et al., 2008, 2017; Park et al., 2019) at the CHARMM-GUI website (http://www.charmmgui.org/). Protonatable residues were identified using the PDB2PQR website (http://apbs-rest-test.westus2.cloudapp.azure.com/pdb2pqr) and their
protonated states adjusted for pH 6 in CHARMM-GUI. The TIP3P model was used to simulate explicit water molecules. The octahedral solvation box was used as this allowed for at least 1.0 nm from the protein in each axis, and 0.138 M NaCl was added. The CHARMM36 force field was used (Mackerell et al., 2004; Huang et al., 2014), and all calculations were performed at 293.1 K. The particle mesh Ewald algorithm was applied to calculate electrostatic forces, and the van der Waals interactions were smoothly switched off at 1.0 nm by a force-switching function (Steinbach et al., 1994). A time step of 2 fs was used in all simulations. Initially, the system was shortly equilibrated in constant particle number, volume, and temperature (NVT) condition using CHARMM36 (Brooks et al., 2009). To assure gradual equilibration of the system, positional restraints for backbone and side-chain heavy atoms were applied and the restraint forces were gradually reduced during the equilibration. The system was simulated for 100 ns and repeated three times using the CHARMM36 force field on the Kathleen high-performance cluster at University College London using NAMD (Phillips et al., 2020). For the production, particle number, pressure and temperature (NPT) simulation, the Langevin coupling coefficient was set to 1 ps$^{-1}$ and a Nosé-Hoover Langevin-piston (Martyna et al., 1994; Feller et al., 1995) was used to maintain constant pressure (1 bar) with a piston period of 50 fs and a piston decay of 25 fs. The time step was 2 fs and trajectories were saved every 100 ps. The electrostatic interactions were updated every 20 fs. The short-range nonbonded and electrostatic interactions were calculated with a cut-off of 1.2 nm. SHAKE was used to constrain all bonds involving hydrogen atoms. Convergence of all repeats was checked through the comparison of the average RMSD using VMD (Humphrey et al., 1996).

A principal component analysis (PCA) of the frames from all three repeat simulations was used to identify commonly occurring hinge conformations. Five such hinge conformations were used to build five full-length IgG3 starting structures by attaching the above structures for the Fab and Fc regions. This approach of creating several start structures using theoretical hinge conformations minimised a potential bias if a single hinge conformation had been used to create full-length IgG3 structures. To facilitate the attachment of the Fc region to the hinge, MD simulations of the hinge included a small segment SVFLFPPK$^{246}$ (EU numbering) in the IgG3-Fc crystal structure that was kept...
unchanged in the MD simulations while allowing the hinge to move. This \textsuperscript{239}SVFLFPK\textsuperscript{246} peptide in the Fc region served as a key anchor site for the Fc region. The two N-linked Asn\textsuperscript{297} glycans on the Fc C\textsubscript{H}2 domains were taken to be complex-type biantennary oligosaccharides with a Man\textsubscript{3}GlcNAc\textsubscript{2} core and two NeuNAc.Gal.GlcNAc antennae (Deisenhofer, 1981). The glycan template for this was taken from the GitHub repository (https://github.com/dww100), which was energy minimized using NAMD (Phillips \textit{et al.}, 2005) for 1 nanosecond to achieve a relaxed structure. This glycan was added to the Fc region by bringing its C1 atom in its first GlcNAc residue to within 0.14 nm of the Asn\textsuperscript{297} sidechain N atom in the C\textsubscript{H}2 domain, while ensuring no steric clashes with the Fc residues and the glycan chain. The PDB file was then opened on Discovery Studio (Dassault Systèmes BIOVIA, San Diego) in order to create “CONECT” records for these glycosidic bonds. The CHARMM force field parameters and protein structure file, including those for the disulphide bridges and glycans were generated using the CHARMM-Gui GlycanReader tool (Miller \textit{et al.}, 2008; Jo \textit{et al.}, 2011; Lee \textit{et al.}, 2016) in order to be compatible with the CHARMM36 forcefield (MacKerell \textit{et al.}, 1998; Mackerell \textit{et al.}, 2004; Guvench \textit{et al.}, 2009; Raman \textit{et al.}, 2010b; Best \textit{et al.}, 2012). This procedure was followed for all five starting structures. To relax all ten starting structures (five with and five without glycans), these were energy minimised using the simulation engine NAMD (version 2.9) with the CHARMM36 forcefield.

For the Monte Carlo simulations to generate trial structures, the residues in the five starting IgG3 structures were renumbered and their naming nomenclature was adjusted to match the required format for the Torsion Angle Monte Carlo (TAMC) module on SASSIE-web for this to work (Zhang \textit{et al.}, 2017). The IgG3 residue numbering was thus changed into one continuous segment that encompassed both the Fab and the Fc regions. A library of physically realistic glycosylated and deglycosylated structural conformations was generated by subjecting the five starting hinge structures to the TAMC module in SASSIE-web (Zhang \textit{et al.}, 2017). The peptides that were conformationally varied were assigned within the hinge. For glycosylated IgG3, these peptides were \textsuperscript{217}LKT\textsuperscript{220} and \textsuperscript{285}PS\textsuperscript{286} on one hinge of IgG3, and \textsuperscript{217}LKTPLG\textsuperscript{222} on the other hinge (Figure 7.1B, 7.2E). For deglycosylated IgG3 these peptides were \textsuperscript{217}LKT\textsuperscript{219} and \textsuperscript{283}GG\textsuperscript{284} on one hinge, and \textsuperscript{217}LKT\textsuperscript{220} on the other hinge of IgG3. Slightly different
peptides were varied as initial tests were run in order to find the peptides that resulted in the least disulphide bond stretching and yield in highest acceptance rate. These six peptides corresponded to surface-accessible structures outside the structurally-defined Fab and Fc regions and the disulphide-linked hinge core. These peptides were structurally varied in TARC to create the required IgG3 conformers for testing against the scattering curve, while the rest of the IgG3 model was held rigid. Allowing $^{217}$LKTP$^{220}$, $^{217}$LKTPGL$^{222}$ or $^{217}$LKT$^{219}$ to be variable on both IgG3 hinges enabled both Fab regions to be conformationally mobile. Allowing $^{285}$PS$^{286}$ or $^{283}$GG$^{284}$ to be variable enabled the Fc region to be mobile too. For each of these linker residues, the backbone phi ($\phi$) and psi ($\psi$) torsion angles were varied in $15^\circ$ steps. During the Monte Carlo simulations, many attempted steps were physically unrealistic and were therefore discarded by TARC. Overall, for glycosylated IgG3, 50,000 moves were attempted for each of the five starting hinge structures. Of these, 31-61% were rejected because of inappropriate steric clashes to leave a total of 135,135 out of 250,000 starting models that were accepted. For deglycosylated IgG3, in which the glycan chains were omitted, 250,000 moves were attempted of which 57-85% of the models were rejected to leave 73,905 models were accepted. This is a higher rejection rate than was seen in previous Chapters 4 and 5, which is likely observed here because of disulphide bond stretching in the elongated hinge corresponding to movements in the Fc.

For the two libraries of 135,135 and 73,905 models, a scattering curve was generated for each model using the SasCalc module in SASSIE-web. SasCalc calculated the scattering curve $I(Q)$ using an all-atom expression for the scattering intensity in which the orientations of the $Q$ vectors are taken from a quasi-uniform spherical grid generated by the golden ratio (Watson et al., 2013). For X-ray modelling, consideration of the hydration shell would require the explicit addition of a monolayer of water molecules to the protein surface before calculating $I(Q)$, and would require much computational effort (Watson et al., 2013). Thus, the hydration shell was not considered here for X-rays, and was not required for neutrons. These scattering curves were compared to the SAXS and SANS experimental curves extrapolated to zero concentration, using the R-factor function in SASSIE-web. This function calculates the difference between the
modelled curve $I_{\text{Model}}(Qi)$ and the interpolated experimental curves $I_{\text{Expt}}(Qi)$, this function being analogous to that used in protein crystallography:

$$R = \frac{\sum \sqrt{\|I_{\text{Expt}}(Qi) - \eta I_{\text{Model}}(Qi)\|^2}}{\sum \|I_{\text{Expt}}(Qi)\|} \times 100$$

where $Qi$ is the $Q$ value of the $i$th data point, $I_{\text{Expt}}(Qi)$ is the experimental scattering intensity, $I_{\text{Model}}(Qi)$ is the theoretical modelled scattering intensity, and $\eta$ is a scaling factor used to match the theoretical curve to the experimental $I(0)$ (Wright et al., 2015). Lower $R$-factor values represent better fits. An iterative search to minimize the $R$ factor was used to determine $\eta$ (Wright et al., 2015). The use of $\chi^2$ analyses to evaluate the fits was not possible because this requires the experimental data points to have errors associated with them, which were not available upon interpolation of the curve. In the extrapolated experimental scattering curves, the lowest $Q$ values in the range before the fitted Guinier $R_G$ region were interpolated to zero $Q$ using MATLAB in order to satisfy the input requirement for the SasCalc module in SASSIE-web. Interpolation makes the $Q$ spacing uniform between the data points, and extrapolation extends the full $I(Q)$ curve to zero $Q$. The resulting 825 and 146 $I(Q)$ values in the $Q$ range of 0.0-1.5 nm$^{-1}$ were utilised for the SAXS and SANS curve fits respectively, and defined the $Q$ spacing for use in SasCalc and the $R$-factors. For the SANS curve fits, a correction of 2.0% of $I(0)$ was required to allow for a flat incoherent background that was attributed to the proton content of IgG3 and the heavy water dialysis buffer (Bonner et al., 2009). The 135,135 glycosylated and 73,905 deglycosylated models gave an $R$-factor vs. $R_G$ distribution that encompassed the experimental extrapolated $R_G$ value. This $R$-factor analysis was repeated for four experimental SAXS curves at different concentrations for each of glycosylated and deglycosylated IgG3 (Table 7.2). The same analysis was repeated for two SANS curves at different concentrations for each of glycosylated and deglycosylated IgG3 (Table 7.3). For each concentration, the best-fit 100 models with the smallest $R$-factors were accepted. All the accepted models were subjected to short energy minimisations to correct any broken bonds that may have arisen during the TAMC simulation. All the calculated parameters in Tables 7.2 and 7.3 were calculated both prior to and after the energy minimisation to monitor its effect on the extracted structures. None of the parameters underwent a significant
change after energy minimisation. Moreover, two best fit curves (one before and one after energy minimisation) are shown in Figures 7.10 and 7.11 to demonstrate that the short energy minimisations had a negligible effect on the theoretical SAXS and SANS curves.

Principal component analysis (PCA) provided by the Bio3d package in R (Grant et al., 2006) was used to identify the main classes of best-fit IgG3 conformations found in the 400 best-fit glycosylated and deglycosylated SAXS models (Table 7.3). A separate analysis of the 400 best-fit SANS models was performed (Table 7.5). To remove any bias in the PCA clustering of coordinate sets caused by the presence or absence of the glycans, the glycans were removed from the best-fit glycosylated models prior to generating the PCA. The mid-point structure for each PCA group was identified using a centroid model computed using R.

In order to model AUC parameters the theoretical $s_{20,w}$ values were generated for the best-fit 800 glycosylated and deglycosylated SAXS and SANS IgG3 models using HullRad (Fleming et al., 2018) (Table 7.2 and 7.3). Hullrad includes glycan residues for glycosylation, however there are inconsistencies in the Protein Database nomenclature for glycans. The nomenclature in the HullRad script was thus modified to ensure that the IgG3 glycosylation was correctly incorporated in the $s_{20,w}$ calculation.

7.3 Results

7.3.1 Purification and characterisation of glycosylated and deglycosylated IgG3

Human IgG3 was purified in high yields from human myeloma serum. In order to work with two different forms of IgG3, the deglycosylation of myeloma IgG3 was set up using PNGase F digests according to the manufacturer's protocol (section 7.2.1). The completeness of deglycosylation was verified by a combination of size-exclusion gel filtration, SDS-PAGE and mass spectrometry, and also by analytical ultracentrifugation (see below):

(i) Deglycosylation of IgG3 was completed after 10 hours. The elution of the IgG3 PNGase F-treated product from a gel filtration column slightly preceded that for native glycosylated IgG3 (Figure 7.3A). Both forms eluted as a main symmetrical peak at 16.43 ml and 16.38 ml for glycosylated and deglycosylated
Table 7.2 Modelling fits for the X-ray scattering and analytical ultracentrifugation data in light water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>$R_g$ before minimization (nm)</th>
<th>$R_g$ after minimization (nm)</th>
<th>$R_{xs}$ before minimization (nm)</th>
<th>$R_{xs}$ after minimization (nm)</th>
<th>L (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ before minimisation (S)</th>
<th>$s_{20,w}$ after minimisation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library of glycosylated Models</td>
<td>87200</td>
<td>4.96-10.97</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2.04-21.48</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 1.00 mg/ml</td>
<td>100</td>
<td>7.02-7.58</td>
<td>6.98-7.54</td>
<td>2.04-2.81</td>
<td>1.89-2.67</td>
<td>1.30-1.67</td>
<td>1.34-1.71</td>
<td>NA</td>
<td>2.04-2.93</td>
<td>2.09-4.44</td>
</tr>
<tr>
<td>Best fit at 1.00 mg/ml</td>
<td>1</td>
<td>7.43</td>
<td>7.38</td>
<td>2.24</td>
<td>2.10</td>
<td>1.45</td>
<td>1.45</td>
<td>28</td>
<td>2.04</td>
<td>2.80</td>
</tr>
<tr>
<td>Top 100 at 0.50 mg/ml</td>
<td>100</td>
<td>6.80-7.43</td>
<td>6.79-7.39</td>
<td>2.04-2.79</td>
<td>1.89-2.67</td>
<td>1.36-1.77</td>
<td>1.34-1.71</td>
<td>NA</td>
<td>2.66-3.58</td>
<td>2.62-4.84</td>
</tr>
<tr>
<td>Best fit at 0.50 mg/ml</td>
<td>1</td>
<td>7.09</td>
<td>7.06</td>
<td>2.26</td>
<td>2.19</td>
<td>1.57</td>
<td>1.55</td>
<td>25</td>
<td>2.66</td>
<td>2.89</td>
</tr>
<tr>
<td>Library of deglycosylated Models</td>
<td>135141</td>
<td>5.06-10.68</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.70-21.8</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 1.00 mg/ml</td>
<td>100</td>
<td>7.22-7.56</td>
<td>7.20-7.55</td>
<td>2.29-2.71</td>
<td>2.42</td>
<td>1.34-1.70</td>
<td>1.33-1.67</td>
<td>NA</td>
<td>1.70-3.23</td>
<td>1.77-3.20</td>
</tr>
<tr>
<td>Best fit at 1.00 mg/ml</td>
<td>1</td>
<td>7.36</td>
<td>7.35</td>
<td>2.26-2.69</td>
<td>2.39</td>
<td>1.33-1.67</td>
<td>1.59</td>
<td>27</td>
<td>1.70</td>
<td>1.89</td>
</tr>
<tr>
<td>Top 100 at 0.50 mg/ml</td>
<td>100</td>
<td>7.09-7.45</td>
<td>7.08-7.44</td>
<td>2.10-2.54</td>
<td>2.08-2.48</td>
<td>1.38-1.72</td>
<td>1.15-1.73</td>
<td>NA</td>
<td>3.17-3.47</td>
<td>3.12-4.52</td>
</tr>
<tr>
<td>Best fit at 0.50 mg/ml</td>
<td>1</td>
<td>7.23</td>
<td>7.22</td>
<td>2.27</td>
<td>2.25</td>
<td>1.55</td>
<td>1.55</td>
<td>27</td>
<td>3.17</td>
<td>3.19</td>
</tr>
</tbody>
</table>
Table 7.3 Principal component analysis of modelling fits for the X-ray scattering and analytical ultracentrifugation data in light water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>$R_g$ before minimization (nm)</th>
<th>$R_g$ after minimization (nm)</th>
<th>$R_{OS1}$ before minimization (nm)</th>
<th>$R_{OS2}$ after minimization (nm)</th>
<th>$\Delta L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ before minimization (S)</th>
<th>$s_{20,w}$ after minimization (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA Group 1</td>
<td>50</td>
<td>6.80-7.58</td>
<td>6.79-7.54</td>
<td>2.34-2.79</td>
<td>1.33-1.73</td>
<td>NA</td>
<td>2.68-3.58</td>
<td>2.45-4.84</td>
<td>6.18-6.52</td>
<td>6.18-6.52</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>50</td>
<td>6.80-7.58</td>
<td>6.79-7.54</td>
<td>2.34-2.79</td>
<td>1.33-1.73</td>
<td>NA</td>
<td>2.68-3.58</td>
<td>2.45-4.84</td>
<td>6.18-6.52</td>
<td>6.18-6.52</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>6.99</td>
<td>6.90</td>
<td>2.77</td>
<td>1.49</td>
<td>1.51</td>
<td>21</td>
<td>3.53</td>
<td>3.34</td>
<td>6.37</td>
</tr>
<tr>
<td>PCA Group 2</td>
<td>50</td>
<td>7.29-7.56</td>
<td>7.25-7.54</td>
<td>3.04-2.69</td>
<td>1.34-1.66</td>
<td>NA</td>
<td>1.95-3.48</td>
<td>1.88-4.44</td>
<td>5.69-6.36</td>
<td>5.74-6.36</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>19</td>
<td>7.29-7.56</td>
<td>7.27-7.54</td>
<td>2.24-2.69</td>
<td>1.37-1.53</td>
<td>NA</td>
<td>1.95-3.45</td>
<td>1.88-3.44</td>
<td>5.69-5.93</td>
<td>5.74-5.94</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>7.33</td>
<td>7.26</td>
<td>2.57</td>
<td>1.56</td>
<td>1.49</td>
<td>22</td>
<td>2.85</td>
<td>3.15</td>
<td>6.24</td>
</tr>
<tr>
<td>PCA Group 3</td>
<td>83</td>
<td>6.84-7.16</td>
<td>6.83-7.08</td>
<td>2.08-2.61</td>
<td>1.36-1.77</td>
<td>NA</td>
<td>2.20-3.58</td>
<td>2.46-4.52</td>
<td>5.85-6.61</td>
<td>5.89-6.61</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>81</td>
<td>6.84-7.10</td>
<td>6.83-7.08</td>
<td>2.08-2.61</td>
<td>1.36-1.77</td>
<td>NA</td>
<td>2.20-3.58</td>
<td>2.46-4.10</td>
<td>6.32-6.61</td>
<td>6.32-6.61</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>2</td>
<td>7.16</td>
<td>7.08</td>
<td>2.52-2.54</td>
<td>1.44-1.46</td>
<td>NA</td>
<td>3.45-3.46</td>
<td>4.51-4.52</td>
<td>5.85-5.86</td>
<td>5.89-5.90</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>6.88</td>
<td>6.87</td>
<td>2.11</td>
<td>1.53</td>
<td>1.48</td>
<td>24</td>
<td>3.54</td>
<td>4.06</td>
<td>6.60</td>
</tr>
<tr>
<td>PCA Group 4</td>
<td>38</td>
<td>7.07-7.35</td>
<td>6.99-7.27</td>
<td>2.57-2.81</td>
<td>1.41-1.67</td>
<td>NA</td>
<td>2.29-3.56</td>
<td>2.09-3.57</td>
<td>6.03-6.27</td>
<td>6.03-6.27</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>38</td>
<td>7.07-7.35</td>
<td>6.99-7.27</td>
<td>2.57-2.81</td>
<td>1.41-1.67</td>
<td>NA</td>
<td>2.29-3.56</td>
<td>2.09-3.57</td>
<td>6.03-6.27</td>
<td>6.03-6.27</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>7.24</td>
<td>7.13</td>
<td>2.81</td>
<td>1.53</td>
<td>1.55</td>
<td>23</td>
<td>2.93</td>
<td>2.36</td>
<td>6.19</td>
</tr>
<tr>
<td>PCA Group 5</td>
<td>179</td>
<td>7.09-7.56</td>
<td>7.08-7.55</td>
<td>2.10-2.71</td>
<td>1.34-1.72</td>
<td>NA</td>
<td>1.70-3.47</td>
<td>1.78-3.70</td>
<td>5.71-6.06</td>
<td>5.78-6.07</td>
</tr>
<tr>
<td>Glycosylated</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>179</td>
<td>7.09-7.56</td>
<td>7.08-7.55</td>
<td>2.10-2.71</td>
<td>1.34-1.72</td>
<td>NA</td>
<td>1.70-3.47</td>
<td>1.78-3.70</td>
<td>5.71-6.06</td>
<td>5.78-6.07</td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>7.40</td>
<td>7.39</td>
<td>2.59</td>
<td>1.50</td>
<td>1.48</td>
<td>23</td>
<td>2.23</td>
<td>2.06</td>
<td>5.83</td>
</tr>
</tbody>
</table>

305
Table 7.4 Modelling fits for the X-ray scattering and analytical ultracentrifugation data in heavy water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>$R_G$ before minimization (nm)</th>
<th>$R_G$ after minimization (nm)</th>
<th>$R_{XS-1}$ before minimization (nm)</th>
<th>$R_{XS-1}$ after minimization (nm)</th>
<th>$R_{XS-2}$ before minimization (nm)</th>
<th>$R_{XS-2}$ after minimization (nm)</th>
<th>$L$ (nm)</th>
<th>$R$-factor before minimization (%)</th>
<th>$R$-factor after minimization (%)</th>
<th>$s_{20,w}$ before minimization (S)</th>
<th>$s_{20,w}$ after minimization (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library of glycosylated Models</td>
<td>87200</td>
<td>4.95-10.94</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>8.72-32.29</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 1.85 mg/ml</td>
<td>100</td>
<td>6.33-6.93</td>
<td>6.32-6.89</td>
<td>1.73-2.37</td>
<td>1.70-2.34</td>
<td>1.25-2.09</td>
<td>1.28-2.09</td>
<td>NA</td>
<td>8.72-10.71</td>
<td>8.07-10.73</td>
<td>6.32-6.75</td>
<td>6.51-6.74</td>
</tr>
<tr>
<td>Best fit at 1.85 mg/ml</td>
<td>1</td>
<td>6.93</td>
<td>6.89</td>
<td>1.93</td>
<td>1.84</td>
<td>1.30</td>
<td>1.33</td>
<td>23</td>
<td>8.72</td>
<td>8.07</td>
<td>6.62</td>
<td>6.73</td>
</tr>
<tr>
<td>Top 100 at 0.70 mg/ml</td>
<td>100</td>
<td>6.14-6.93</td>
<td>6.13-6.89</td>
<td>1.73-2.29</td>
<td>1.69-2.29</td>
<td>1.25-2.09</td>
<td>1.28-2.09</td>
<td>NA</td>
<td>11.82-11.30</td>
<td>13.66</td>
<td>6.32-6.75</td>
<td>6.51-6.75</td>
</tr>
<tr>
<td>Best fit at 0.70 mg/ml</td>
<td>1</td>
<td>6.19</td>
<td>6.65</td>
<td>1.84</td>
<td>1.82</td>
<td>1.49</td>
<td>1.51</td>
<td>25</td>
<td>11.82</td>
<td>11.77</td>
<td>6.64</td>
<td>6.70</td>
</tr>
<tr>
<td>Library of deglycosylated Models</td>
<td>13514</td>
<td>5.04-10.62</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>5.90-28.52</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Top 100 at 1.21 mg/ml</td>
<td>100</td>
<td>6.30-6.86</td>
<td>6.27-6.81</td>
<td>1.79-2.62</td>
<td>1.76-2.59</td>
<td>1.21-1.96</td>
<td>1.21-1.97</td>
<td>NA</td>
<td>5.90-7.51</td>
<td>5.56-7.35</td>
<td>6.02-6.41</td>
<td>6.11-6.42</td>
</tr>
<tr>
<td>Best fit at 1.21 mg/ml</td>
<td>1</td>
<td>6.56</td>
<td>6.62</td>
<td>2.35</td>
<td>2.32</td>
<td>1.58</td>
<td>1.57</td>
<td>22</td>
<td>5.90</td>
<td>6.76</td>
<td>6.15</td>
<td>6.16</td>
</tr>
<tr>
<td>Top 100 at 0.65 mg/ml</td>
<td>100</td>
<td>6.30-7.04</td>
<td>6.27-7.02</td>
<td>1.79-2.62</td>
<td>1.76-2.59</td>
<td>1.21-1.95</td>
<td>1.20-1.96</td>
<td>NA</td>
<td>8.75-10.17</td>
<td>8.57-10.10</td>
<td>5.95-6.40</td>
<td>6.00-6.42</td>
</tr>
<tr>
<td>Best fit at 0.65 mg/ml</td>
<td>1</td>
<td>6.62</td>
<td>6.60</td>
<td>2.10</td>
<td>2.06</td>
<td>1.67</td>
<td>1.67</td>
<td>21</td>
<td>8.75</td>
<td>8.57</td>
<td>6.31</td>
<td>6.32</td>
</tr>
</tbody>
</table>
Table 7.5 Principal component analysis of modelling fits for the X-ray scattering and analytical ultracentrifugation data in heavy water.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Models</th>
<th>PCA Group 1</th>
<th>Glycosylated</th>
<th>Deglycosylated</th>
<th>Centroid</th>
<th>PCA Group 2</th>
<th>Glycosylated</th>
<th>Deglycosylated</th>
<th>Centroid</th>
<th>PCA Group 3</th>
<th>Glycosylated</th>
<th>Deglycosylated</th>
<th>Centroid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R_2$ before</td>
<td>$R_2$ after</td>
<td>$R_{XS,1}$ before</td>
<td>$R_{XS,1}$ after</td>
<td>$R_{XS,2}$ before</td>
<td>$R_{XS,2}$ after</td>
<td>$L$ (nm)</td>
<td>$R$-factor before</td>
<td>$R$-factor after</td>
<td>$s_{20,w}$ before</td>
<td>$s_{20,w}$ after</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>minimisation</td>
<td>minimisation</td>
<td>minimisation</td>
<td>minimisation</td>
<td>minimisation</td>
<td>minimisation</td>
<td>(nm)</td>
<td>minimisation</td>
<td>minimisation</td>
<td>minimisation</td>
<td>(S)</td>
<td>(S)</td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>19</td>
<td>6.32-7.04</td>
<td>6.32-7.02</td>
<td>1.83-2.36</td>
<td>1.79-2.36</td>
<td>1.50-1.85</td>
<td>1.50-2.36</td>
<td>NA</td>
<td>6.66-10.16</td>
<td>6.41-10.09</td>
<td>5.95-6.34</td>
<td>6.00-6.34</td>
<td></td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>6.63</td>
<td>6.63</td>
<td>2.04</td>
<td>2.00</td>
<td>1.79</td>
<td>1.80</td>
<td>21</td>
<td>7.39</td>
<td>6.93</td>
<td>6.25</td>
<td>6.32</td>
<td></td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>6.66</td>
<td>6.64</td>
<td>1.91</td>
<td>1.89</td>
<td>1.32</td>
<td>1.31</td>
<td>22</td>
<td>1.37</td>
<td>1.35</td>
<td>6.59</td>
<td>6.65</td>
<td></td>
</tr>
<tr>
<td>PCA Group 3</td>
<td>173</td>
<td>6.30-6.79</td>
<td>6.27-6.77</td>
<td>1.90-2.62</td>
<td>1.85-2.59</td>
<td>1.21-1.96</td>
<td>1.21-2.59</td>
<td>NA</td>
<td>5.91-10.17</td>
<td>5.56-10.10</td>
<td>6.10-6.41</td>
<td>6.11-6.42</td>
<td></td>
</tr>
<tr>
<td>Glycosylated</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Deglycosylated</td>
<td>173</td>
<td>6.30-6.79</td>
<td>6.27-6.77</td>
<td>1.90-2.62</td>
<td>1.85-2.59</td>
<td>1.21-1.96</td>
<td>1.21-2.59</td>
<td>NA</td>
<td>5.91-10.17</td>
<td>5.56-10.10</td>
<td>6.10-6.41</td>
<td>6.11-6.42</td>
<td></td>
</tr>
<tr>
<td>Centroid</td>
<td>1</td>
<td>6.74</td>
<td>6.66</td>
<td>1.93</td>
<td>1.88</td>
<td>1.86</td>
<td>1.83</td>
<td>21</td>
<td>7.30</td>
<td>6.41</td>
<td>6.31</td>
<td>6.32</td>
<td></td>
</tr>
</tbody>
</table>
IgG3, respectively (Figure 7.3A). Trace amounts of a possible dimer just before 15 ml were discarded. This process ensured that the IgG3 sample was monodisperse with no aggregates present immediately before analytical ultracentrifugation or scattering experiments.

(ii) When the IgG3 samples were submitted to non-reducing and reducing SDS-PAGE analyses at equimolar concentrations, purified glycosylated and deglycosylated IgG3 showed a single band above 140 kDa on 4-12% Bis Tris NuPage gel under non-reducing conditions, which is consistent with the assumed mass of ~158 kDa for IgG3 (Figure 7.3B). Under reducing conditions two bands were present corresponding to the heavy chain (with an apparent mass of ~50-70 kDa) and the light chain (with an apparent molecular mass of ~25 kDa) (Figure 7.3B). These apparent molecular masses were as expected from the known sequence (Figure 7.2).

(iii) Liquid chromatography mass spectrometry measurements showed five peaks for glycosylated (in black) and deglycosylated IgG3 (in magenta) in a mass range of 160.2-169.5 kDa for glycosylated IgG3 and 156.6-166.0 kDa for deglycosylated IgG3 (Figure 7.3C) (Miller et al., 2008). These peaks were unaffected by longer PNGase F digestion times. The existence of five peaks may be attributed to O-linked glycans of sizes 1.9-2.8 kDa (Plomp, Dekkers, Rombouts, Visser, Koeleman, et al., 2015). The most intense IgG3 glycosylated population had an observed deconvoluted mass of 164,847 Da. For deglycosylated IgG3 (in magenta) the most intense population had an observed deconvoluted mass at 161,383 Da. The mass of each N-glycan chain was calculated by subtracting the glycosylated and deglycosylated masses and halving the outcome to give 1,732 Da. This value agrees well with an assumed N-glycan composition of Gal2Man3GlcNAc4 that gives a mass of 1622 Da.

7.3.2 Analytical ultracentrifugation of glycosylated and deglycosylated IgG3

Sedimentation velocity experiments characterised the masses and solution structures of glycosylated and deglycosylated IgG3. The SEDFIT analyses of the boundaries involved fits of as many as 600 scans, and the good agreement between the experimental boundary scans and fitted lines is clear (left, Figure 7.4A). From the $c(s)$ analyses, the molecular masses of the IgG3
Figure 7.3 Purification, SDS-PAGE and mass spectrometry of human glycosylated and deglycosylated IgG3.

A Elution peaks from a Superose 6 Increase 10/300 gel filtration column for a glycosylated IgG3 sample (black) and a deglycosylated IgG3 sample (magenta). The dashed vertical lines indicate the peak positions. B Lane 1, 3, and 5, molecular mass markers are denoted in kDa. Lanes 2 and 4, non-reduced and reduced SDS-PAGE of glycosylated IgG3 after gel filtration. Lanes 6 and 7, reduced and non-reduced SDS-PAGE of deglycosylated IgG3 after gel filtration. C Mass spectra of glycosylated (black) and deglycosylated (magenta) IgG3.
The experimentally observed sedimentation boundaries using interference optics for a concentration series of glycosylated IgG3 and deglycosylated IgG3 in histidine buffer in light water. Scans were recorded at 30,000 rpm and 20°C, from which 31-60 boundaries (black outlines) are shown from totals of up to 896 scans. The SEDFIT fits are shown in blue. The peaks in the corresponding size distribution analyses $c(s)$ revealed a monomer peak (M) at $s_{20,w}$ values of 5.78-6.33 S for glycosylated and deglycosylation IgG3 in light water. B The $s_{20,w}$ values for the monomer peaks are shown as a function of concentration for glycosylated (●), and deglycosylated (●) IgG3.

Figure 7.4 Sedimentation velocity analyses of glycosylated and deglycosylated IgG3.
monomer peak were 108 kDa (glycosylated) and 122 kDa (deglycosylated) in light water. In heavy water the molecular masses of the IgG3 monomer peak were 139 kDa (glycosylated) and 156 kDa for deglycosylated. These values were comparable with the composition-calculated masses of 158 kDa and 154 kDa for the glycosylated and deglycosylated IgG3 monomers respectively. These also agree well with the values from mass spectrometry of 165 kDa and 161 kDa for glycosylated and deglycosylated IgG3 respectively (Figure 7.3C). In the resulting size distribution analyses $c(s)$, a clear monomer peak that monitored the overall IgG3 solution structure was observed at average $s_{20,w}$ values of $5.82 \pm 0.06$ S for glycosylated IgG3 and $6.29 \pm 0.05$ S for deglycosylated IgG3 in light water (Figure 7.4A; Table 7.1). Because the samples were dilute, the expected 0.2 S reduction (3%) in the $s_{20,w}$ values after deglycosylation to follow the 4 kDa reduction (3%) in the IgG3 mass was not seen for reason of noisy data at low concentrations. These $s_{20,w}$ values were consistent with previous AUC studies of glycosylated IgG3 that reported $s_{20,w}$ values of $6.11 \pm 0.03$ S (Lu et al., 2007) and $5.90 \pm 0.02$ (Gregory et al., 1987). Variable minor peaks at 3-4 S were seen adjacent to the main monomer peak (M) (Figure 7.4A); there were attributable to minor fragmentation of IgG3 and/or artefacts caused by small sample-buffer optical mismatches. In conclusion, given that no concentration dependences were seen in Figure 7.4C, these data sets showed that IgG3 was monomeric in solution.

### 7.3.3 X-ray and neutron scattering of glycosylated and deglycosylated IgG3

The solution structures of glycosylated and deglycosylated IgG3 samples were characterised by X-ray and neutron scattering. The two methods provided different perspectives of the same solution structure. X-rays in light water buffers detect the hydration shell surrounding the protein structure, whereas neutrons in heavy water buffers see a much reduced effect of this hydration shell for reason of the different solute-solvent contrast in use (Perkins, 1986, 2001; Svergun et al., 1998).

The X-ray data collection at IgG3 concentrations between 0.5-2.46 mg/ml used time frame analyses to ensure the absence of radiation damage effects. The resulting $R_g$ and $R_{xs-1}/R_{xs-2}$ values monitor the elongation of the overall IgG3
structure and its approximate cross-sectional structures respectively. Guinier analyses resulted in high quality linear plots for all samples and revealed three distinctive regions of the $I(Q)$ curves, which is seen in the scattering curves for antibodies (Pilz et al., 1970; Rayner et al., 2014, 2015). From these, the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ values from the individual scattering curves were obtained within satisfactory $Q.R_G$ and $Q.R_{XS}$ limits of 1.05-1.55, 0.52-0.71 and 0.78-1.74 respectively (Figure 7.5A). The mean X-ray $R_G$ values, that monitor the overall structure for glycosylated and deglycosylated IgG3 samples were similar, being 6.95 ± 0.06 nm and 7.02 ± 0.05 nm respectively (Table 7.1). The $R_G$ value for glycosylated IgG3 compared well with previous SAXS studies which reported $R_G$ values of 6.93 nm (Liu et al., 2019), 6.20 nm (Gregory et al., 1987) and 7.16 nm (Lu et al., 2007). The $R_{XS-1}$ values from the individual curves (Figure 7.6A) is an approximate monitor of the cross-sectional Fab and Fc arrangement in glycosylated IgG3 and deglycosylated IgG3. The mean X-ray $R_{XS-1}$ values for glycosylated and deglycosylated IgG3 were both typically 1.41 ± 0.2 nm, showing that the spatial arrangement of the Fab and Fc regions was unchanged following glycan removal. These values were notably reduced from $R_{XS-1}$ values of 2.5 (± 0.1) nm measured for both monoclonal IgG1 and IgG4 before and after deglycosylation (Spiteri et al., 2021b, 2021a). This difference is attributed to the effect of the long hinge region in IgG3 that separated the Fab and Fc regions. This also had the effect of reducing the $Q$ range that could be used for the $R_{XS-1}$ fits compared to IgG1 and IgG4. The $R_{XS-2}$ values are an approximate monitor of the mean cross-sectional dimensions of the individual Fab and Fc regions. The mean $R_{XS-2}$ values for glycosylated and deglycosylated IgG3 were similar in a range of 1.59 to 1.61 (± 0.2) nm. This indicates that the averaged cross-section of the two Fab and one Fc region were unchanged before and after deglycosylation. Similar $R_{XS-2}$ values of 1.4 (± 0.1) nm were recently reported for both monoclonal IgG1 and IgG4 respectively before and after deglycosylation (Spiteri et al., 2021b, 2021a). No concentration dependences were seen for IgG3, this being seen from the Guinier values which remained unchanged within error (Figure 7.6). These results showed that the overall and individual Fab and Fc regions were unchanged in their solution structures between 0.5-2.5 mg/ml.
The SAXS curves for glycosylated and deglycosylated IgG3 at concentrations of 0.50-2.46 mg/ml. The filled circles between the arrows represent the $Q.R_G$ and $Q.R_{XS}$ fit ranges used to determine the $R_G$ and $R_{XS}$ values. The $Q$ range used for the $R_G$ values was 0.10-0.22 nm$^{-1}$; those for the $R_{XS-1}$ and $R_{XS-2}$ values were 0.22-0.28 nm$^{-1}$ and 0.50-1.10 nm$^{-1}$ respectively.

The SANS curves for glycosylated and deglycosylated IgG3 at concentrations of 0.65-1.85 mg/ml. The $Q$ range used for the $R_G$ values was 0.10-0.22 nm$^{-1}$ and those for the $R_{XS-1}$ and $R_{XS-2}$ values were 0.23-0.28 nm$^{-1}$ and 0.50-1.10 nm$^{-1}$ respectively.

Figure 7.5 X-ray and neutron Guinier $R_G$ and $R_{XS}$ analyses for glycosylated and deglycosylated IgG3.
Figure 7.6 Concentration dependence of the SAXS and SANS Guinier analyses.

The filled symbols show the values determined from the Guinier analyses and the open symbols in the $R_G$ panels indicate those determined from the $P(r)$ analyses. The colours denote the glycosylated (black) and deglycosylated (magenta) IgG3.

A The SAXS $R_G$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for glycosylated (●, ○) and deglycosylated (●, ○) IgG3. The solid lines corresponded to the mean $R_G$ values determined from the Guinier analyses for glycosylated and deglycosylated IgG3, and the dashed lines to the mean $R_G$ values determined by $P(r)$ analyses.

B The SANS $R_G$, $I(0)/c$, $R_{XS-1}$ and $R_{XS-2}$ values for glycosylated and deglycosylated IgG3, each corresponding to a single measurement in histidine buffer in $^2$H$_2$O. The solid and dashed lines correspond to the mean values for glycosylated and deglycosylated IgG3.
The corresponding neutron scattering data sets for glycosylated and
deglycosylated IgG3 in 100% $^2$H$_2$O buffer were acquired at similar concentrations
of 0.65-1.85 mg/ml. Again, the Guinier analyses revealed high quality linear fits
for the $R_G$, $R_{XS-1}$ and $R_{XS-2}$ parameters (Figure 7.5B). A concentration dependence
was not observed for IgG3, this being seen from the $I(0)/c$ values which remained
unchanged within error (Figure 7.6B). This difference between the neutron and
X-ray data sets is attributable to the fewer data points obtained with neutrons,
leading to reduced precision in the data sets. The mean neutron $R_G$ values for
glycosylated and deglycosylated IgG3 were similar at 6.23 ± 0.04 nm and 6.29 ±
0.04 nm respectively (Figure 7.6B; Table 7.1). The mean neutron $R_{XS-1}$ values for
glycosylated and deglycosylated IgG3 were similar at 0.80 ± 0.3 nm and 1.05 ±
0.3 nm respectively (Table 7.1). The mean neutron $R_{XS-2}$ values for glycosylated
and deglycosylated IgG3 were unchanged at 1.06 ± 0.2 nm and 1.12 ± 0.3 nm
respectively. Overall, the neutron $R_G$ values confirmed the X-ray analyses,
however the neutron $R_{XS-1}$ and $R_{XS-2}$ are reduced compared to their X-ray values,
this being attributed to the reduced contribution of the hydration shell to these
measurements.

The distance distribution function $P(r)$ is derived from Fourier
transformation of the scattering curve $I(Q)$, and provides structural information in
real space on glycosylated and deglycosylated IgG3. The X-ray and neutron $P(r)$
analyses gave $R_G$ values that were similar to those from the X-ray Guinier
analyses, showing that the two analyses were self-consistent (open symbols,
Figure 7.6). The X-ray maximum lengths $L$ of glycosylated and deglycosylated
IgG3 were determined from the value of $r$ when the $P(r)$ curve intersects zero on
the $r$ axis, and were generally 28 nm (Figure 7.7A). The neutron maximum lengths
$L$ of glycosylated and deglycosylated IgG3, were 25-28 nm. These were slightly
smaller when compared to the X-ray value of 28 nm (Figure 7.7B). These
reductions in the neutron $R_G$ and $L$ values compared to the X-ray values have
been previously seen in an earlier joint SAXS and SANS studies of antibodies
(Rayner et al., 2015). From previous SAXS experiments of IgG3, the reported
maximum lengths of glycosylated IgG3 molecule were 19.5 nm (Lu et al., 2007)
and 27.8 nm (Liu et al., 2019), the latter value agrees well with experimental
values (Table 7.1).
Figure 7.7 SAXS and SANS distance distribution analyses $P(r)$ for each of glycosylated and deglycosylated IgG3.

Glycosylated and deglycosylated IgG3 are denoted in black and magenta. A The peak maxima at $M_1$, $M_2$ and $M_3$ and the maximum length $L$ are indicated by arrows. The SAXS and SANS $P(r)$ curves for glycosylated and deglycosylated IgG3 are shown at concentrations of 0.50–2.46 mg/ml. B The corresponding $P(r)$ curves for the SANS curves for IgG3 at concentrations of 0.65–1.85 mg/ml. (C,D) The concentration dependence of the $M_1$, $M_2$ and $M_3$ peaks for glycosylated (●) and deglycosylated (●) IgG3 are shown. The fitted lines are the mean values for glycosylated and deglycosylated IgG3 averaged together (solid line).
The maxima in the $P(r)$ curves corresponded to the most frequently occurring distances between scattering elements within the IgG3 structures. For the IgG3 samples, three peaks, $M1$, $M2$ and $M3$. Two peaks, $M1$ and $M2$, are characteristic for IgG1, IgG2 and IgG4 antibodies (Rayner et al., 2014, 2015; Hui et al., 2019; Spiteri et al., 2021b, 2021a). The appearance of the third peak, $M3$, is attributed to the elongated IgG3 hinge structure, and was also seen previously (Liu et al., 2019). In IgG3 the $M1$ peak corresponds primarily to the shorter distances within each Fab and Fc region, and its position is expected to be almost invariant for this reason. In IgG3, $M2$ now corresponds to the distances between pairs of Fab regions, while $M3$ corresponds primarily to the longer distances between pairs of Fab and Fc regions (Figures 7.1 and 7.7). No differences in maxima peak positions were observed between glycosylated and deglycosylated IgG3 (Figure 7.7C). For the X-ray data, the $M1$ peak was observed at $3.53 \pm 0.06$ nm for glycosylated IgG3 and $3.58 \pm 0.09$ nm for deglycosylated IgG3. The $M2$ peak was observed at $8.97 \pm 0.47$ nm for glycosylated IgG3 and $9.11 \pm 0.04$ nm for deglycosylated IgG3. While not well resolved by X-rays, the positions of the $M3$ peak varied between 16 nm to 19 nm. The $M3$ peak was observed at $17.0 \pm 1.5$ nm for glycosylated IgG3 and $17.2 \pm 1.9$ nm for deglycosylated IgG3. The neutron $P(r)$ analyses reflected the same trends as the X-ray data (Figure 7.7D) although the $M3$ peak was better resolved in the neutron $P(r)$ curves (Figure 7.7B). The neutron $M1$ peak was seen at $3.8 \pm 0.5$ nm for glycosylated IgG3 and $3.7 \pm 0.2$ nm for deglycosylated IgG3, the $M2$ peak was seen at $9.7 \pm 0.2$ nm for glycosylated IgG3 and $9.7 \pm 0.1$ nm for deglycosylated IgG3, and the $M3$ peak was seen at $14.2 \pm 0.1$ nm for glycosylated IgG3 and $14.4 \pm 0.6$ nm for deglycosylated IgG3. Both the X-ray and neutron analyses were thus consistent with each other and support the premise that the two Fab regions were well separated from the Fc region in solution.

7.3.4 Atomistic modelling of glycosylated and deglycosylated IgG3

Despite the observation by scattering of small but detectable changes in the scattering curves of monoclonal IgG1 and IgG4 following glycan removal (Spiteri et al., 2021b, 2021a), no such changes could be seen for myeloma IgG3 in the present study. The most likely reason for this is that any increase in IgG3
Fc disorder would have a minimal effect on the well-separated Fab regions for reason of the long hinge in IgG3. The atomistic modelling simulations of the glycosylated and deglycosylated IgG3 structures were used to determine best-fit molecular structures, starting from two high-resolution crystal structures for the human Fab and Fc regions (Section 7.2.4). The IgG3 sequence is that given in Figure 7.2A-G. The Fab and Fc regions were joined by a 62-residue hinge peptide (section 7.2.4; Figure 7.2E), also built using PyMOL. This hinge was independently energy minimised, then simulated in triplicate using all-atom molecular dynamics for 100 ns. From this, a PCA analysis (David et al., 2014) identified the most common hinge conformations. Five of these conformations that corresponded to each cluster mid-point in the PCA were selected and attached to the Fab and Fc regions to result in five deglycosylated full-length IgG3 models. The corresponding five glycosylated IgG3 models were created by adding complex-type biantennary glycans to the two Asn297 sidechains in the Fc region (Figure 7.1B). The five glycosylated starting structures were each subjected to energy minimisation.

Physically-realistic IgG3 models without steric clashes were created for comparison with the experimental X-ray curves. By varying the torsion angles at three flexible regions at the start and end of the two IgG3 hinges (section 7.2.4; Figure 7.1), trial IgG3 structures were created that involved movements of the two Fab and one Fc regions relative to each other. For glycosylated IgG3, 250,000 models were generated in five Monte Carlo simulations, from which 135,135 models were accepted with no steric clashes in the model. For deglycosylated IgG3, 250,000 models were generated in five Monte Carlo simulations, of which 73,905 models were likewise acceptable. As a control to ensure that no systematic trends had been overlooked, four X-ray and four neutron scattering curves for up to four concentrations were fitted for each of the four samples in question (Figure 7.12).

(a) For X-rays with both forms of IgG3, comparison of the four experimental scattering curves at 0.50-1.00 mg/ml with the 135,135 and 73,905 theoretical curves gave a goodness-of-fit $R$-factor vs $R_g$ distribution with clear minima in all eight cases (Figure 7.10A and B). The minima agreed with the
Figure 7.8 Principal component analyses (PCA) and root-mean-square deviation (RMSD) of the hinge molecular dynamics simulations (legend overleaf).
Figure 7.8 Principal component analyses (PCA) and root-mean-square derivation (RMSD) of the hinge molecular dynamics simulations (continued).

A-D The three molecular dynamics repeats for the IgG3 hinge (Section 7.3.4) were combined and analysed using PCA. The PCA groups 1, 2, 3, 4 and 5 are respectively coloured in magenta, red, yellow, green and blue. The centroid model for each cluster is represented by numbered ★ symbols.

The three molecular dynamics repeats of the IgG3 hinge grouped by PCA into four groups of PC2 vs PC1, PC3 vs PC2 and PC3 vs PC1. D, The first three eigenvalue rankings (PC1, PC2 and PC3) captured 62.3% of the variance in the molecular dynamics repeats.

E The RMSD plotted against time in the simulation, for repeat 1, 2 and 3, respectively in blue, red and yellow. The mid-point for each cluster found in the PCA is represented by numbered ★ symbols to show from where in the simulation and which repeat each mid-point structure is found.
The cartoons denote the protein backbone of the five starting IgG3 models which were used in TAMC simulations. These start models were glycosylated by the addition of two glycans at Asn\(^{297}\) (EU numbering) or Asn\(^{344}\) (continuous numbering), and then also used in TAMC.
Figure 7.10 Atomistic modelling analyses for glycosylated and deglycosylated IgG3 (legend overleaf).
Figure 7.10 Atomistic modelling analyses for glycosylated and deglycosylated IgG3 (continued).

The R-factor values for a total of 135,135 and 73,905 physically-realistic models for glycosylated and deglycosylated IgG3 are shown as circles when these were plotted against the theoretical $R_G$ values. The models were generated from five glycosylated start structures, which independently generated 34,315, 29165, 27635, 24425, and 19595 structures and these were then combined to give a total of 135,135 accepted structures. A total of 73,905 deglycosylated models were accepted from five structures which independently generated 13,345, 16,200, 7,665, 21,375, and 15,320 structures. The five starting IgG3 models (section 7.3.4) from which the TAMC structures were derived have different hinge conformations. The five sets of structures are coloured in purple, green, orange, pink and blue. (A,B) Two experimental X-ray scattering curve analyses for two concentrations of glycosylated and deglycosylated IgG3 are shown. (C,D) Two experimental neutron scattering curve analyses for two concentrations of glycosylated and deglycosylated IgG3 are shown. In all eight panels, the experimental curves were fitted to the total of 135,135 and 73,905 modelled curves. Those models with $R_G$ values closest to the experimental $R_G$ values showed the lowest goodness-of-fit R-factor as expected. The 100 best-fit models with the lowest goodness-of-fit $R$-factors are denoted by red circles. The experimental $R_G$ value is represented by the solid vertical line and the dashed vertical lines on either side of this represent the ± 2% upper and lower limits of these $R_G$ values.
experimental $R_G$ values (Figure 7.6A). The minima showed that enough trial X-ray models had been generated to result in good fits in each case. Filtering of the models to select these with the lowest $R$-factors gave the 100 best-fit models for each concentration (red, Figure 7.10). The range of the 100 $R$-factors for each of the concentrations was low at between 2.04-3.58% for the best-fit glycosylated models and 1.70-3.47% for the best-fit deglycosylated models (Table 7.2). This indicated good quality X-ray curve fits between the experimental and modelled curves.

(b) For neutrons, the same outcome was found, thus confirming the reproducibility of the curve fits. Comparisons of the same 135,135 and 73,905 theoretical curves with the neutron scattering curves at 0.65-1.85 mg/ml showed again that 100 best-fit structures could be identified at clear minima in each of the $R$-factor vs $R_G$ neutron distributions (Figures 7.10C and D). The minima agreed with the experimental $R_G$ values (Figure 7.6B). The range of $R$-factors was now higher than the X-ray values for reason of the reduced precision of the neutron scattering curves.

The eight sets of 100 best-fit X-ray and neutron models (Figure 7.10) were subjected to PCA in order to identify the resulting best-fit IgG3 conformations from the curve fits (Hui et al., 2015). The PCA determines the correlated motions of protein residues in IgG3 as linearly uncorrelated variables, each being termed a principal component (David et al., 2014). These “essential motions” were extracted from a covariance matrix of the atomic coordinates of the frames in the selected IgG3 structure set. The eigenvectors of this matrix each have an associated eigenvalue that characterises the clustering of the models based on structural coordinates (or variance). In order to eliminate bias in the PCA, the glycan chains were removed from the glycosylated IgG3 models before comparison with the deglycosylated models.

(a) For the X-ray fits, the PCA indicated a difference between the glycosylated and deglycosylated IgG3 models (black and magenta respectively, Figure 7.11A-D; Table 7.2). Thus, the best-fit 200 glycosylated and 200 deglycosylated X-ray models were each clustered into five distinct groups. The glycosylated models mostly occurred in the PCA groups 1, 2, 3 and 4, while the deglycosylated models mostly occurred in the PCA group 5 with some overlap.
Figure 7.11 Principal component analyses (PCA) of the best-fit glycosylated and deglycosylated IgG3 models (legend overleaf).
Figure 7.11 Principal component analyses (PCA) of the best-fit glycosylated and deglycosylated IgG3 models (continued).

Glycosylated models are represented by black symbols and deglycosylated models are represented by magenta symbols. The PCA groups 1, 2, 3, 4 and 5 are represented by ○, Δ, +, × and □ in that order, and the centroid model for each group is represented by blue numbered ★ symbols.

(A-D) The four sets of 100 best-fit models from the experimental X-ray scattering curves for glycosylated and deglycosylated IgG3 were grouped by PCA into four groups of PC2 vs PC1, PC3 vs PC2 and PC3 vs PC1. D, The first three eigenvalue rankings (PC1, PC2 and PC3) captured 38.2% of the variance in the 400 models.

(E-H) The four sets of 100 best-fit models from the experimental neutron scattering curves for glycosylated and deglycosylated IgG3 were grouped by PCA into four groups PC2 vs PC1, PC3 vs PC2 and PC3 vs PC1. H, The first three eigenvalue rankings (PC1, PC2 and PC3) captured 84.5% of the variance in the 400 models.
Figure 7.12 Scattering curve fits for the best-fit glycosylated and deglycosylated IgG3 models (legend overleaf).
Figure 7.12 Scattering curve fits for the best-fit glycosylated and deglycosylated IgG3 models (continued).

The experimental curves are denoted by black circles. The best-fit modelled curves are denoted by red lines, and the best-fit modelled curves after energy minimisation denoted by blue lines. The corresponding distance distribution curves $P(r)$ are shown in the top right of each panel.

A Glycosylated and B deglycosylated IgG3 X-ray curve fits are shown. In A, the glycosylated IgG3 model fits were taken from PCA group 2 (1.00 mg/ml) and PCA group 1 (0.50 mg/ml) (Table 7.3). In B, the deglycosylated IgG3 model fits were taken from PCA group 5 (1.00 and 0.50 mg/ml) in that order.

C Glycosylated and D deglycosylated IgG3 neutron curve fits are shown. A flat background correction of 2.0% of I(0) was applied. In C, the glycosylated IgG3 model fits corresponded to PCA group 2 (1.85 and 0.70 mg/ml) to allow for incoherent scattering (Table 7.5). In D, the deglycosylated IgG3 model fits corresponded to PCA group 3 (1.21 and 0.65 mg/ml).
with group 2. The visually-excellent In \( I(Q) \) and \( P(r) \) X-ray curve fits confirmed the validity of the modelling fits (Figures 7.12A, B). The curve fits were good from low \( Q \) to about 1.2 nm\(^{-1} \) which is where most of the intensities arise, out in the further \( Q \) range of 1.2-1.5 nm\(^{-1} \) where the intensities are lower the experimental and theoretical curves do not fit as well. Since the \( P(r) \) curve is calculated from the whole \( I(Q) \) curve, the reduced quality of the peak fits of the \( P(r) \) may be due to the poorer fitting of the \( I(Q) \) curves at the higher \( Q \) values. (b) For the neutron fits, the PCA also indicated differences between the glycosylated and deglycosylated IgG3 models (black and magenta respectively, Figure 7.11E-H; Table 7.3). The best-fit 200 glycosylated and 200 deglycosylated neutron models were clustered into three groups, with some overlap between glycosylated and deglycosylated groups. The glycosylated models were mostly in PCA groups 1 and 2, while the deglycosylated models were mostly in PCA group 3 but with some overlap with PCA group 1. Visually-good neutron \( I(Q) \) curve fits were obtained, however the fit of these curves was inferior to the fit obtained in the X-ray fits at high \( Q \) (Figures 7.12C, D). An incoherent scattering flat background correction of 2.0% of \( I(0) \) in the neutron fits accounted for a mismatch between the sample and buffer background in heavy water (Perkins et al., 2008). Interestingly, in all four neutron fits, the triple-peaked \( P(r) \) curves were well replicated in the modelled best-fit curves.

Further insights into the X-ray and neutron data and their modelling were obtained from the dimensionless Kratky analyses of \((Q.R_g)^2 . I(Q)/I(0)\) vs \( Q.R_g \) for the experimental scattering curves at the highest concentrations in use and the scattering curves from the modelled best fit structures. These plots indicate whether the macromolecule in question is globular in its structure or possesses intrinsically disordered regions (Receveur-Brechot et al., 2012). Unlike monoclonal IgG1 and IgG4 and myeloma IgG2 where two clear Kratky peaks were observed at around 2 nm and 4 nm (Hui et al., 2019; Spiteri et al., 2021b, 2021a), only a single Kratky peak close to 2 nm was observed for myeloma IgG3. For the X-ray Kratky curves (Figure 7.13A), the \( Q.R_g \) values for the experimental glycosylated IgG3 peak was 2.24, in good accord with the modelled values of 2.25 before energy minimisation and 2.23 after energy minimisation. The \( Q.R_g \) values for the experimental deglycosylated peaks of 2.26 were also in good

A X-ray experimental data (solid lines) and model fits before energy minimisation (dashed lines) and model fits after energy minimisation (dotted lines) were shown in black for glycosylated IgG3 at 1.00 mg/ml and in magenta for deglycosylated IgG3 at 1.00 mg/ml. B Neutron experimental data (solid lines) and model fits before energy minimisation (dashed lines) and model fits after energy minimisation (dotted lines) were shown in black for glycosylated IgG3 at 1.85 mg/ml and in magenta for deglycosylated IgG3 at 1.21 mg/ml.
accord with the modelled deglycosylated peaks of 2.26 before energy minimisation and 2.33 after energy minimisation. For the neutron Kratky curves (Figure 7.13B), the $Q_RG$ value for the experimental glycosylated IgG3 peak was shifted to 4.79, which were similar to the modelled peaks at 4.80 before energy minimisation and 4.80 after energy minimisation. For deglycosylated IgG3, the $Q_RG$ value for the experimental peak were 4.66, which was larger than the modelled values of 3.88 before and after energy minimisation. The existence of a single and not double Kratky peak for IgG3, also seen in other studies (Liu et al., 2019) was most likely due to the large separation between the Fab and Fc regions in IgG3, meaning that these three regions showed independent scattering properties.

As another test of the scattering modelling, the $s_{20,w}$ values for the four sets of best-fit 100 glycosylated and deglycosylated models from each X-ray concentration (Figures 7.8, 7.10) were calculated using HullRad (Fleming et al., 2018). This gave an $s_{20,w}$ range of 6.03-6.61 S for the four X-ray concentrations for glycosylated IgG3 and 5.74-6.07 S for deglycosylated IgG3 (Table 7.2). These values agreed well with the experimental $s_{20,w}$ values of 5.90-5.78 S for glycosylated IgG3 and 6.24-6.33 S for deglycosylated IgG3 (Table 7.1). These agreements corroborated the outcome of the atomistic scattering modelling, given that the mean difference between the modelled and experimental values should typically be ± 0.21 S for related macromolecules (Perkins et al., 2009). This modelling was however unable to distinguish changes before and after deglycosylation.

7.4 Discussion

This present AUC, SAXS and SANS study of human myeloma IgG3, combined with an atomistic scattering curve fit method based on molecular dynamics and Monte Carlo simulations (Perkins et al., 2008), has provided the first experimental molecular structure for a full-length IgG3 antibody. This work now completes a series of similar studies on all four human IgG subclasses IgG1-IgG4 that resulted in the first atomistic experimentally-determined solution structures for all of these and clarified their molecular functions (Hui et al., 2015, 2019; Spiteri et al., 2021b, 2021a). Unlike previous studies of the full-length IgG3 structure, this approach utilised simulations of stereochemically-correct atomistic

The black cartoon denotes the protein backbone of the starting glycosylated and deglycosylated IgG3 models. In the upper row of A and B, the Fc regions of the 100 best-fit models were superimposed onto the starting reference structure, thus focusing on variations in the Fab region in these models. In the lower row of A and B, the Fab regions of the 100 best-fit models were superimposed onto this starting reference structure, thus focusing on variations in the Fc region in these models. The blue and magenta wireframe envelopes denote the conformational space occupied by the glycosylated and deglycosylated Fab or Fc regions.

(A) Best-fit X-ray models at 1.00 mg/ml for both IgG3 forms.
(B) Best-fit neutron models at 1.83 and 1.23 mg/ml IgG3.
(C) Superimposition of the five hinge structures used for the IgG3 modelling.
(D) Cartoon to illustrate the best-fit IgG3 structure from this study.
models using a combination of molecular dynamics and Monte Carlo simulations. These served as conformational libraries of 135,135 and 73,905 models that explored a wide range of structures for full-length IgG3. When these libraries were compared with extensive experimental data sets from SAXS and SANS, it was possible to rank these models in terms of goodness of fit to the experimental scattering curves. In order to show the relative positions of the Fab and Fc regions in the top best-fit 100 models, these structures were superimposed upon each other at their hinge, then were displayed as wireframe representations rotated about their vertical axis (Figure 7.14). These views make it clear that the IgG3 solution structure involves an extended but not rigid polyproline hinge with some flexibility in this, and that the two Fab regions and the one Fc region each occupy distinct conformational space about the hinge. Unlike the cases of glycosylated and deglycosylated IgG1 and IgG4 (Spiteri et al., 2021b, 2021a), no clear differences between these two forms in IgG3 was detectable. This outcome is most likely attributable to the length of the hinge region, which may reduce the impact of any conformational differences in flexibility that was previously seen for IgG1 and IgG4 (Spiteri et al., 2021b, 2021a).

From the curve fits, the most prominent feature of the IgG3 models is the polyproline disulphide-linked hinge region (Figure 7.1A). For the curve fits, five such hinges were simulated using energy minimisation and molecular dynamics, and all showed extended but variable structures (Figure 7.14C). The maximum length of the hinge between Glu\textsuperscript{216} and Pro\textsuperscript{285} (Figure 7.2E) ranged between 7.67 nm and 10.80 nm, in good agreement with the increased length of IgG3 compared to IgG1, IgG2 and IgG4. The simulations do not predict a regular polyproline helix structure with limited flexibility, and this concurs with the scattering curves.

Experimentally, the elongated IgG3 solution structure was primarily confirmed from the Guinier scattering data, where the $R_G$ values derived from Guinier analyses were significantly larger at 6.97 nm by X-rays and 6.26 nm (Table 7.1) than the $R_G$ values reported for the other IgG subclasses, namely IgG1 at 5.02-5.32 nm, IgG2 at 5.24-5.38 nm and IgG4 at 4.77-4.94 nm (Lu et al., 2007; Rayner et al., 2014b, 2015; Hui et al., 2019a; Spiteri et al., 2021b, 2021a) (Figure 7.6). The cross-sectional radius, $R_{XS}$, that typically monitored the spatial separation of the Fab and Fc regions in IgG was smaller at 1.41 nm by X-rays for
IgG3 compared to those of the other IgG subclasses. Thus the other X-ray $R_{\text{XS-1}}$ values were 2.46-2.65 nm for IgG1 (Rayner et al., 2015; Spiteri et al., 2021b), 2.61 nm for IgG2 (Hui et al., 2019) and 2.30-2.50 nm for IgG4 (Rayner et al., 2014; Spiteri et al., 2021a). The significant decrease in the $R_{\text{XS-1}}$ value for IgG3 compared to IgG1, IgG2 and IgG4 is attributable to the separation of the Fab and Fc regions by the elongated hinge of IgG3. The $R_{\text{XS-2}}$ parameter represents the averaged spatial cross-section of each of the two Fab and one Fc regions in IgG1, IgG2 and IgG4. This was 1.61 nm in IgG3, which is similar to those of the other IgG subclasses, namely 1.35-1.43 nm for IgG1, 1.35-1.37 nm for IgG2, and 1.23-1.42 nm for IgG4. In fact, the $R_{\text{XS-1}}$ value of IgG3 is similar to the $R_{\text{XS-2}}$ values of all four IgG subclasses, suggesting that this parameter may correspond to the apparent cross-section of the two Fab regions positioned in a linear elongated arrangement shown in Figure 7.14A and B.

Further evidence for the formation of an elongated IgG3 solution structure compared to IgG1, IgG2 and IgG4 was obtained from the distance distribution function $P(r)$. The maximum length $L$ of IgG3 was significantly larger at 25-28 nm (Figure 7.7) than the $L$ values for the other IgG subclasses, IgG1 namely 16-17 nm for IgG1, 17 nm for IgG2 and 15-16 nm for IgG4. The extra length $L$ for IgG3 suggested that the IgG3 hinge is about 10 nm long. In addition, the appearance of the $P(r)$ curve for IgG3 showed three maxima $M1$, $M2$ and $M3$, whereas the other IgG subclasses only showed two maxima $M1$ and $M2$. The three maxima in the $P(r)$ curves for IgG3 were replicated by the atomistic modelling (Figure 7.12). In molecular terms, the $M1$ maximum corresponded to the distances within a single Fab or Fc region, while the $M2$ and $M3$ maxima corresponded to the separation between the two Fab regions and the separations of the Fab-Fc regions respectively. $M2$ and $M3$ are thus conformational monitors of the overall IgG3 structure. The different $M3$ values seen by X-rays and neutrons suggested that, while the extended IgG3 hinge can adopt a range of conformations, the IgG3 hinge remained extended in solution. Kratky plots are often used as a measure of structural flexibility (Receiveur-Brechot et al., 2012). For IgG3, only one peak was now visible (Figure 7.13), instead of the two peaks seen for IgG1, IgG2 and IgG4, and no effect was seen following glycan removal. The best-fit theoretical models showed good agreement with the experimental SAXS curve, again
indicating that the best-fit atomistic modelling was able to replicate this. No increase in flexibility was detected after glycan removal, unlike IgG1 and IgG4. A previous IgG3 study analysed the Kratky plots for IgG1, IgG2 and IgG3 to propose that IgG3 has increased flexibility when compared to IgG1 and IgG2 (Liu et al., 2019).

The experimental AUC data for IgG3 provided another clear indication that its solution structure is more elongated than those of IgG1, IgG2 and IgG4. from AUC, SAXS and SANS provided excellent insight into the glycosylated and the deglycosylated structure of human IgG3. The AUC data showed that IgG3 was monomeric in solution up to 2 mg/ml, The lower sedimentation coefficients s20,w of 6.0 S compared to IgG1 (6.5 S), IgG2 (7.2 S) and IgG4 (6.5 S) demonstrated that IgG3 has a more elongated structure than that of IgG1, IgG2 and IgG4 (Hui et al., 2015; Spiteri et al., 2021b, 2021a).

The present application of atomistic scattering modelling resulted in molecular structures that were able to reproduce all the major features of the SAXS, SANS and AUC data collection. The data sets were fitted to a library of 135,135 and 73,905 trial models. The best-fit X-ray R-factors were low at 2.04% and 1.70% (Table 7.2). This R-factor improvement is attributed to the improved signal-noise ratio of the scattering curves from the B21 instrument at Diamond. The lowest R-factors for the neutron fits were higher for reason of the weaker signal-noise ratios, these being 8.72% and 5.90% (Table 7.3). One noteworthy aspect of the modelling study using SasCalc (Watson et al., 2013) is that an atomistic representation of hydration shells was not made because this was computationally expensive. SAXS sees these hydration shells as they have higher scattering densities similar to those of the protein, compared to bulk water (Perkins, 1986, 2001; Svergun et al., 1998). The hydration shell is less visible by SANS because its neutron scattering density is reduced and is similar to that of the heavy water buffer in use. The complementary fits of the SAXS and SANS data sets are thus a control that similar IgG3 models fit the scattering curves. Of note is that the previous use of the SCT/SCTPL modelling approach (available in SASSIE-web) for X-ray and neutron fits explicitly incorporated hydration shells in a coarse-grained approach.
IgG3 is a potent antibody with a high affinity for several FcRs and C1q. The absence of the Fc glycan increased the conformational space occupied by the Fc region in studies of deglycosylated IgG1 and IgG4 (Spiteri et al., 2021b, 2021a). These studies concluded that the Asn\textsuperscript{297} glycans limit the conformational flexibility of the Fc region, stabilising it in a conformation that makes receptor binding more likely (Lu et al., 2007). In this IgG3 study, conformational change in the Fc region was not detectable after deglycosylation (Figure 7.12D). The longer hinge may enable the Fc glycans to play a more local role in stabilising the C\textsubscript{H}2 domain pair, and the loss of FcγR receptor binding in aglycosylated IgG3 was noted (Walker et al., 1989). The elongated hinge may stabilise the IgG3 Fc region by allowing any Fab motions to take place independently of the Fc region, which is not possible in the other IgG subclasses as they have shorter hinges. If the Fab and Fc regions are more independent of each other for reason of a longer hinge, this would explain the increased FcγR receptor and complement C1q affinity to IgG3, and why IgG4 with the shortest hinge in the IgG subclasses has the lowest affinity to the FcγR receptors.
Chapter 8 Discussion
8.1 Prologue

Human IgG plays a critical role in the adaptive immune response and are also widely exploited as biotherapeutics (Chapter 1). Despite this, the effect of the conserved Fc glycan, found in all IgGs at Asn²⁹⁷, on the structure, function and stability is still not fully understood (Chapter 2). Several functional studies for all IgGs have demonstrated that receptor binding is abrogated for deglycosylated or aglycosylated IgGs. The IgGs have different effector functions owing to difference in their binding affinities for the various FcγRs, with IgG3 being the most potent activator of FcγRs and complement, followed by IgG1 and IgG2, with IgG4 being considered an anti-inflammatory antibody due to its reduced binding affinity to these ligands. Due to their large size and flexibility, several structural methods such as crystallography and nuclear magnetic resonance are too difficult to use. To date there is only one high resolution crystal structure of full-length IgG1 (PDB ID: 1HZH) (Saphire et al., 2001) and two more of IgG4 (PDB ID: 5DK3) (Scapin et al., 2015) and (PDB ID: 6GFE) (Blech et al., 2019). Previous studies into the role of the Fc glycan are predominantly limited to studies on the IgG1 Fc region, alone, rather than the full-length IgG, so the role of the glycan on metrics such as the conformational flexibility of IgG was impossible to study. As most structural studies only focus on IgG1, it is difficult to form a complete picture for the role of the Fc glycan. These gaps in knowledge have been addressed in this work by the studying of three of the four IgG subclasses, IgG1 (Chapter 4), IgG4 (Chapter 5) and IgG3 (Chapter 7) using a combined approach of analytical ultracentrifugation, X-ray and neutron scattering and atomistic modelling (Chapter 3). Moreover, to understand the significance of the glycans, the thermostability of IgG1 and IgG4, both widely used as biotherapeutics, was studied using thermal unfolding experiments followed by fluorescence and small angle scattering which further clarified the role of the Fc glycans (Chapter 6). In the following sections are summaries of the results described in Chapter 4-7.

8.2 Solution structure of glycosylated and deglycosylated IgG1

IgG1 is comprised of two Fab regions connected to a Fc region through a 15-residue hinge peptide. In this thesis, glycosylated and deglycosylated
monoclonal human IgG1 (designated as A33) was subjected to a comparative multidisciplinary structural study of both forms. Following deglycosylation using PNGase F, analytical ultracentrifugation showed that IgG1 remained monomeric and the sedimentation coefficients $s_{20,w}^0$ of IgG1 decreased from 6.45 S by 0.16-0.27 S. This change was attributed to the reduction in mass following glycan removal. X-ray and neutron scattering revealed changes in the Guinier structural parameters after deglycosylation. While the radius of gyration $R_G$ was unchanged, the cross-sectional radius of gyration, $R_{XS-1}$, increased by 0.1 nm and the commonly occurring distance peak $M2$ of the distance distribution curve $P(r)$ increased by 0.4 nm. These changes revealed that the Fab-Fc separation in IgG1 was perturbed following deglycosylation. To explain these changes, atomistic scattering modelling based on Monte Carlo simulations resulted in 123,284 and 119,191 trial structures for glycosylated and deglycosylated IgG1 respectively. From these, 100 X-ray and neutron best-fit models were determined. For these, principal component analyses identified five groups of structural conformations that were different for glycosylated and deglycosylated IgG1. The Fc region in glycosylated IgG1 showed a restricted range of conformations relative to the Fab regions, while the Fc region in deglycosylated IgG1 showed a broader conformational spectrum. These more variable Fc conformations account for the loss of binding to the FcγRs in deglycosylated IgG1.

8.3 Solution structure of glycosylated and deglycosylated IgG4

IgG4 has the shortest hinge with 12 residues. To unravel the role of the IgG4 Fc glycan, glycosylated and deglycosylated monoclonal human IgG4 (designated as A33) was subjected to a comparative multidisciplinary structural study. Following deglycosylation using PNGase F, analytical ultracentrifugation showed that IgG4 remained monomeric and the sedimentation coefficients $s_{20,w}^0$ decreased from 6.52 S by 0.27 S that reflected its reduction in mass following glycan removal. X-ray and neutron scattering revealed negligible changes in the Guinier structural parameters after deglycosylation. In the distance distribution curve $P(r)$, the commonly occurring distance peaks $M1$ and $M2$ remained unchanged following deglycosylation. To further understand the role of the IgG4 glycans, atomistic scattering modelling based on Monte Carlo simulations
resulted in 111,382 and 117,135 trial structures for glycosylated and deglycosylated IgG4 respectively. In the resulting 100 X-ray and neutron best-fit models, principal component analyses identified four groups of structural conformations that showed differences between glycosylated and deglycosylated IgG4. The Fc region in the glycosylated IgG4 models showed a relatively restricted range of conformations, while the deglycosylated structures showed a slightly broader range of conformations. The difference in range of conformations between glycosylated and deglycosylated IgG4 is more subtle than the differences seen in IgG1. These slightly more variable Fc conformations in deglycosylated IgG4 account for its loss of binding to the FcγRs.

### 8.4 Thermostability of glycosylated and deglycosylated IgG1 and IgG4

The role of the Fc glycan in mediating the thermostability of intact IgG1 and IgG4 has not been fully explored using structural methods. Monoclonal human IgG1 A33 and IgG4 A33 was thus subjected to a multidisciplinary structural strategy. Thermal unfolding curves, gave two melting temperatures $T_{\text{m,app} \, 1}$ and $T_{\text{m,app} \, 2}$, for the native antibody which were reduced by 8.6°C and 3.8°C, respectively after glycan removal, using PNGase F and reduced by 6.4°C and 0.5°C for IgG4. X-ray and neutron solution scattering showed that, after deglycosylation, both the radius of gyration and maximum length of IgG1 and IgG4 from the distance distribution curve $P(r)$ increased at a lower temperature for IgG1 and IgG4. This indicated that the Fc glycans protects the IgG structure from unfolding and aggregation. Despite very high sequence homology (96.05% similarity) IgG1 was more stable than IgG4, with a difference in $T_{\text{m,app} \, 1}$ and $T_{\text{m,app} \, 2}$ of 7.1°C and a 8.1°C, respectively. This difference could be due to IgG4 possessing a shorter hinge, which reduces the global conformational stability of the structure. Interestingly, IgG4 was overall less sensitive to glycan removal, compared to IgG1 (Chapters 4 and 5).
8.5 Solution structure of glycosylated and deglycosylated IgG3

Human immunoglobulin IgG3 is unique in possessing a long hinge region and high affinities for its FcγR receptors and complement C1q. Owing to this hinge length, the full-length molecular structure of IgG3 remains elusive, and the role of the conserved Fc glycans in this is unknown. To address these issues, glycosylated and deglycosylated human myeloma IgG3 was subjected to multidisciplinary solution structure studies. By analytical ultracentrifugation, the elongated structure of IgG3 was revealed from sedimentation coefficients $s_{20,w}^{0}$ of 5.82-6.29 S for glycosylated and deglycosylated IgG3. X-ray and neutron scattering showed that the Guinier $R_G$ values were 6.95 nm for glycosylated IgG3 and were unchanged after deglycosylation, again indicating an elongated structure. The distance distribution function $P(r)$ of both forms of IgG3 showed a maximum length of 25-28 nm and three maxima termed $M1$, $M2$ and $M3$. The molecular structure of IgG3 was determined using atomistic modelling based on molecular dynamics simulations of the IgG3 hinge and Monte Carlo simulations to identify physically-realistic arrangements of the Fab and Fc regions. This resulted in libraries containing 135,135 and 73,905 glycosylated and deglycosylated IgG3 structures respectively. Comparisons with the X-ray and neutron scattering curves gave 100 best-fit models for each of the two forms of IgG3 that accounted for the experimental scattering curves. These models revealed the first molecular structures for full-length IgG3. The structures exhibited relatively restricted Fab and Fc conformations for IgG3, and their extended structures provide an explanation for the enhanced effector functions of IgG3 relative to the other IgG subclasses.

8.6 Conclusions and future work

This thesis has investigated three IgG subclasses and demonstrated that the Fc glycans plays very different structural roles in all three subclasses. For IgG1 the Fc glycan plays a clear role in maintain a conformationally compacted Fc region. This compaction of the Fc region might explain the high affinity binding of IgG1 to FcγRs and its ability to bind to complement component C1q. In IgG4 the Fc glycan plays a more subtle role and only slightly constricts Fc conformations. Overall, the IgG4 Fc is more conformationally labile than the IgG1
Fc region. This difference may provide an explanation for the reduced binding affinity of IgG4 to most FcγRs and its inability to activate complement. IgG3 on the other hand had the most conformationally compacted Fc region and glycan removal did not alter this. This conformational restriction of the IgG3 Fc region may be owing to the elongated hinge region that allows for more independent motions of the Fab and Fc regions, i.e., movements in the Fab regions does not force complementary motions of the Fc regions, in the same way as one may expect when a shorter hinge is present. The more conformationally-constricted IgG3 Fc region may provide a rationale of its enhanced binding affinities to FcγRs and C1q, which are higher than for even IgG1 (Chapter 1).

To follow this work future studies can aim to address a few key issues. Firstly, one of the main limitations of this work was modelling the heterogeneity of the Fc-glycan, which is a combinatorial issue that quickly becomes computationally expensive. To overcome this challenge investment in creating computational workflows that can automate the generation of different glycan sequences and subsequent simulations will allow the community to better investigate the functional role of specific glycan moieties such as fucosylation. Simulations should also extend to modelling the antibodies with different glycosylation patterns in complex with the Fc receptors to develop a better understanding of the role that the Fc-glycan plays in mediating the antibody’s affinity to the Fc receptors. Modelling of hundreds of different start models will become increasingly more accessible given recent upgrades to the research computing facilities at UCL. As the computational capacity available increases and improves, future studies should focus on all-atom molecular dynamics, rather than coarse graining methods, to be able to probe the interaction of residues of the Fc region and the hinge region with the Fc-glycan. These simulations should extend over time scale that are long enough to be able to capture the large conformational space that a flexible protein can take on in solution.

SAS data has proven very powerful in the studies presented in this thesis when combined with computational modelling. Methods of fitting the scattering curve should continue to improve, methods such as ensemble fitting to find the representative scattering curve should be used alongside extracting the best fitting models as it can identify conformations that may not be represented in the
best fitting structures but could be intermediate structures that are important for the function of the protein. This is especially important for investigating intrinsically disordered systems with a high degree of conformational lability.

Finally studying IgGs within an immune complex, when bound to the Fc receptors and antigens has still been a challenge to the community however with continued advances to biophysical techniques using structural techniques such as cryogenic electron microscopy and atomic force microscopy may prove to be insightful in better unravelling the role of the differing hinge lengths and could also have an impact antibody engineering to optimise or selecting IgG structure for specific antigens.
References


Arnold, O., Bilheux, J. C., Borreguero, J. M., Buts, A., Campbell, S. I., Chapon, L., Doucet, M., Draper, N., Ferraz Leal, R., Gigg, M. A., Lynch, V. E.,


*Diamond Annual Review 2019* (2019). Available at:


Frank, M., Walker, R. C., Lanzilotta, W. N., Prestegard, J. H. and Barb, A. W.


ISIS Neutron and Muon Source (no date a) *How ISIS works*. Available at: https://www.isis.stfc.ac.uk/Pages/How-ISIS-works.aspx (Accessed: 23 January 2021).

ISIS Neutron and Muon Source (no date b) *Sans2d*. Available at: https://www.isis.stfc.ac.uk/Pages/Sans2d.aspx (Accessed: 23 January 2021).

ISIS Neutron and Muon Source (no date c) *Zoom*. Available at: https://www.isis.stfc.ac.uk/Pages/Zoom.aspx (Accessed: 23 January 2021).


Drugs, 7, 607–614. doi: 10.1517/13543784.7.4.607.


Kenniston, J. A., Taylor, B. M., Conley, G. P., Cosic, J., Kopacz, K. J., Lindberg,


Journal of Chemical Theory and Computation, 12, 405–413. doi: 10.1021/acs.jctc.5b00935.


Reichert, J. M. (2021) Antibody therapeutics approved or in regulatory review in the EU or US, *The Antibody Society*. Available at:


Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G.


Subedi, G. P. and Barb, A. W. (2019) CD16a with oligomannose-type N-glycans is the only ‘low-affinity’ Fc γ receptor that binds the IgG crystallizable fragment with high affinity in vitro, *Journal of Biological Chemistry*, 293, 16842–16850. doi: 10.1074/jbc.RA118.004998.


Unchained labs (2016) UNit Brochure, *Unchained labs*. Pleasanton, California, USA.


Zhang, H. and Dalby, P. A. (2020) Stability enhancement in a mAb and Fab


Publications


Selected Oral Presentations

Analytical Ultracentrifugation 2019, Christchurch, New Zealand, August 2019. Atomistic solution structures of glycosylated and deglycosylated human IgG1 reveal that the conserved C_{H2} glycans limit the degree of Fc mobility.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,

London Synthetic Biology Network, Imperial College London, United Kingdom, December 2019.
Atomistic solution structures of glycosylated and deglycosylated human IgG1 reveal that the conserved C\textsubscript{H}2 glycans limit the degree of Fc mobility.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,

Neutron Scattering Group Early Careers Meeting, Virtual meeting, September 2020.
Atomistic solution structures of glycosylated and deglycosylated human IgG1 reveal that the conserved C\textsubscript{H}2 glycans limit the degree of Fc mobility.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,

Atomistic solution structures of glycosylated and deglycosylated human IgG1 reveal that the conserved C\textsubscript{H}2 glycans limit the degree of Fc mobility.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,

Selected Posters Presentations
Collaborative Computation project for advanced analyses of structural data in chemical biology and soft condensed matter meeting (CCP-SAS), University of Cardiff, United Kingdom, June 2017.
Atomistic solution structures of glycosylated and deglycosylated human IgG1 reveal that the conserved C\textsubscript{H}2 glycans limit the degree of Fc mobility.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,
UK Neutron and Muon Science and User Meeting, University of Warwick, June 2017.
Solution structures of glycosylated and deglycosylated human IgG1 and IgG4.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,

Network in Bioprocessing meeting, University of Warwick, October 2017.
Solution structures of glycosylated and deglycosylated human IgG1 and IgG4.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,

Knowledge Transfer Network Meeting, University of Manchester, United Kingdom, July 2018.
Atomistic solution structures of glycosylated and deglycosylated human IgG1 reveal that the conserved C_{H2} glycans limit the degree of Fc mobility.
Valentina A. Spiteri, James Doutch, Robert R. Rambo, Jayesh Gor, Paul A. Dalby & Stephen J. Perkins,

Solution structures of glycosylated and deglycosylated human IgG1 reveal that the conserved C_{H2} glycans limit the degree of Fc mobility.
Won a Best Poster Prize.

BioProNET 2 Scientific Meeting. Virtual meeting. 20th November 2020.
Solution structure of deglycosylated human IgG1 shows the role of C_{H2} glycans in its conformation
Appendix I