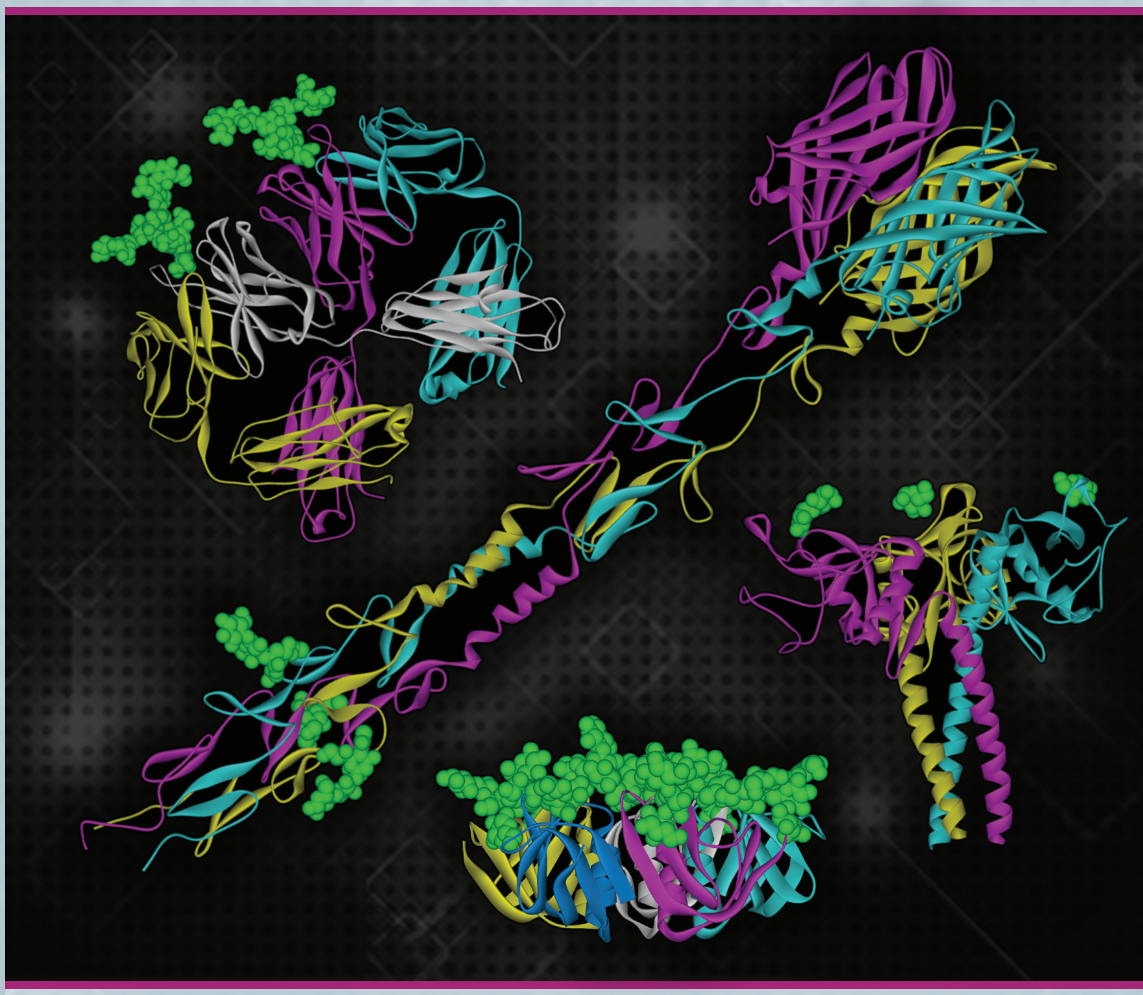
Structural Glycobiology



EDITED BY
Elizabeth Yuriev
Paul A. Ramsland

CRC Press
Taylor & Francis Group

Structural Glycobiology

Structural Glycobiology

EDITED BY
Elizabeth Yuriev
Paul A. Ramsland

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20120801

International Standard Book Number-13: 978-1-4398-5412-9 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

SECTION III Alternative Approaches for Yielding Detailed Structural Information

SECTION IV Carbohydrates in Medicine

Preface

Structural glycobiology is a rapidly progressing field of research, where the diverse structural and functional roles of carbohydrates (oligo- and polysaccharides, glycolipids, and glycoproteins) are examined using a wide variety of experimental as well as theoretical (predictive) approaches. Carbohydrates are key molecules in diverse biological processes that include, but are not limited to, metabolic pathways, cell–cell interactions, carbohydrate–protein interactions, host–pathogen interactions, and immunity.

Although there are several well-written and comprehensive textbooks on glycobiology and the chemistry, biochemistry, and microbiology of carbohydrates, no current book focuses on the specific topic of structural glycobiology. We believe that this book fills the gap by bringing together world-recognized authors to contribute chapters that cover their own specialties in the experimental, theoretical, and emerging technologies employed in this field.

In this book, individual chapters are written by expert authors who are active research scientists in the field and are specialists in key techniques that are relevant to modern structural glycobiology. The book provides concise overviews of the application of specialized technologies to the study of carbohydrates in biology, reviews of relevant and current research in the field, and is illustrated throughout by specific examples of how research investigations have yielded key structural and associated biological information on carbohydrates, glycolipids, and glycoproteins.

The topics covered in this book are broadly divided into four sections. Section I covers well-established, but often challenging, experimental approaches for structure determinations of carbohydrate–protein complexes and large glycoprotein assemblies and explores the techniques of x-ray crystallography and small-angle scattering (Chapter 1), nuclear magnetic resonance (NMR) (Chapter 2), and cryoelectron microscopy (Chapter 3). Jeffries, Farrugia, and Ramsland (Chapter 1) discuss two complementary approaches examining the high-resolution three-dimensional structures (x-ray crystallography) and the solution shapes and conformations (small-angle x-ray and neutron scattering) of carbohydrate binding proteins and glycoproteins. They outline the general features of carbohydrate–protein interactions and discuss the importance of multivalent carbohydrate binding and the role of oligomerization in carbohydrate recognition by proteins. Koharudin and Gronenborn (Chapter 2) provide an easily accessible and educational outline of how NMR can be applied to the study of protein–glycan interactions. The power of the different NMR methodologies for investigating carbohydrate binding, bound carbohydrate conformations, and detailed structures of carbohydrate–protein complexes is beautifully illustrated with carbohydrate–lectin systems. The state-of-the-art in cryoelectron microscopy (cryoEM) for studying very large assemblies of glycoproteins is presented by Zeev-Ben-Mordehai and Grünewald (Chapter 3). Recent advances in sample preparation, image collection, and processing are greatly accelerating the use of cryoEM for structural studies of glycoproteins under near native settings (e.g., in viruses,

cells, and tissues). They provide stunning examples of where cryoEM, particularly tomographic methods, are providing unprecedented structural information on biological assemblies, such as membrane channels, intracellular junctions, and viral glycoproteins.

Section II covers theoretical, or modeling-based, approaches, such as molecular mechanics, molecular dynamics, and free energy calculations (Chapter 4) and carbohydrate docking (Chapter 5). Sarkar and Pérez (Chapter 4) give an excellent overview of the computational approaches used to study protein–carbohydrate interactions. They complement the description of traditional methods with a brief foray into alternative methods used for the enhancement of conformational sampling, such as molecular robotics. Agostino, Ramsland, and Yuriev (Chapter 5) demonstrate the usefulness of molecular docking in structural glycobiology by considering recent docking validation studies on a range of protein targets. They also describe very recent developments in the modeling of water-mediated carbohydrate–protein interactions.

Section III covers alternative techniques for yielding structural information on carbohydrates from complex biological samples (fluids/secretions, cells, and tissues). Here, the rapid advances in mass spectrometry (Chapter 6) are being complemented with glycan-based arrays (Chapter 7) for the study of carbohydrate specificity and recognition. Kolarich and Packer (Chapter 6) provide a detailed overview of the latest mass spectrometric methods for characterization of complex *N*- and *O*-linked oligosaccharides and how the exciting subfield of glycoproteomics is emerging for studying diverse protein glycoforms that are relevant to health and disease. A brief discussion of the need for uniform collection and reporting standards for glycomics-based mass spectrometric investigations is followed by a useful overview of the bioinformatics resources available to researchers using mass spectrometry in glycobiology. The contribution by Song, Smith, and Cummings (Chapter 7) illustrates glycan array technology, focusing on the Consortium for Functional Glycomics methodologies and also introduces the cutting-edge shotgun glycomics approach, which has enormous potential to accelerate research into carbohydrate-mediated interactions from diverse and complex biological samples.

Section IV deals with carbohydrates in medicine. Although carbohydrates are centrally involved in many physiological, biochemical, and cellular processes, three areas of modern medicine (organ transplantation, cancer immunotherapy, and infection treatment) have been directly impacted by our understanding of the structural role of carbohydrates in immune recognition. Brockhausen and Gao (Chapter 8) focus on a range of cancer-related structural and enzymatic glycoaberrations. Christiansen et al. (Chapter 9) deal with carbohydrate antigens implicated in organ rejection. Specifically, they discuss the biochemical, genetic, and immunological characteristics of these carbohydrates, their origins, and interactions with antibodies. Xu and Wilson (Chapter 10) highlight the role of protein–carbohydrate binding for viral adhesion and invasion. They focus on three paradigm systems of viral proteins that recognize sialic acid in cell surface glycans as receptors for viral attachment. They demonstrate, using high-resolution x-ray structures, the subtle structural variations governing the recognition process; for example, for avian versus human influenza A hemagglutinin. Gandhi and Mancera (Chapter 11) describe molecules designed

to mimic the biological activity of glycosaminoglycans (GAGs) through modifications of structure, composition, and sulfation patterns. These new generation GAG-mimetics offer rich potential as therapeutics for treating cancer, inflammation, and infection.

Although each chapter could be a useful stand-alone introduction to a specific technique or area of structural glycobiology, several themes are consistent throughout the book, namely, the role of specific proteins in carbohydrate recognition and function: lectins (Chapters 1, 2, 4, and 5), antibodies (Chapters 1, 4, 5, 9, and 10), and glycosyltransferases (Chapters 4, 5, 8, and 9). From the ligand point of view, the structure and biological roles of two particular types of carbohydrates are of interest in several areas of study: sialic acid and its derivatives (Chapters 1, 8, and 10) and glycosaminoglycans (Chapters 5, 7, 8, 10, and 11).

In summary, this book covers the experimental, theoretical, and alternative technologies that are being applied to the study of the structural basis for the diverse biological roles of carbohydrates. This should be a valuable reference for researchers, graduate students, postdoctoral scientists, and academics with an interest in glycobiology. Researchers from other fields, such as medicinal chemists, biochemists, immunologists, and microbiologists, should also find this a relevant and up-to-date reference and a suitable introduction to the field.

Editors

Elizabeth Yuriev, BSc (Hons) (Kharkov State University, Ukraine), PhD (Victoria University, Melbourne), is a senior lecturer of medicinal chemistry at the Monash Institute of Pharmaceutical Sciences, Monash University, Melbourne, Australia. She is a foundation member and the secretary of the Association of Molecular Modellers of Australasia. She is a member of the editorial board of the *Journal of Molecular Recognition* (John Wiley & Sons). Her current research is focused on protein recognition of carbohydrate and peptide ligands, GPCR modeling, and molecular docking. She has 63 publications in peer-reviewed journals/books, including 47 primary research papers, 8 invited reviews/book chapters, and 8 editorial notes/commentaries.

Paul A. Ramsland, BSc (Hons), PhD (University of Technology, Sydney), is a senior research fellow and a group leader in the Centre for Immunology at the Burnet Institute, Melbourne, Australia. In 2011, he was awarded the Sir Zelman Cowen Fellowship for excellence in medical research relating to cancer. He is an honorary principal fellow (associate professor) in the Department of Surgery Austin Health at the University of Melbourne and is an adjunct senior lecturer in the Department of Immunology at Monash University, Melbourne, Australia. He is an associate editor for the *Journal of Molecular Recognition* and *Frontiers in Immunology* (*Molecular Innate Immunity Specialty*), and a section editor for *BMC Structural Biology* as well as serving on the editorial boards of *Molecular Immunology* and *Molecular Biotechnology*. His current research is focused on examining the three-dimensional structural roles of carbohydrates and glycoproteins in immunity and infection. He has 87 publications in peer-reviewed journals/books, including 67 primary research papers, 16 invited reviews/book chapters, and 4 editorial notes/commentaries.

Contributors

Mark Agostino
Medicinal Chemistry and Drug Action
Monash Institute of Pharmaceutical Sciences
Monash University
Melbourne, Victoria, Australia

Inka Brockhausen
Division of Rheumatology
Department of Biomedical and Molecular Sciences and Department of Medicine
Queen's University
Kingston, Ontario, Canada

Dale Christiansen
Department of Surgery
Austin Health/Northern Health
University of Melbourne
Heidelberg, Victoria, Australia

Richard D. Cummings
Department of Biochemistry
School of Medicine
Emory University
Atlanta, Georgia

William Farrugia
Centre for Immunology
Burnet Institute
Melbourne, Victoria, Australia

Neha S. Gandhi
Curtin Health Innovation Research Institute
and
Western Australian Biomedical Research Institute
and
School of Biomedical Sciences
Curtin University
Perth, Western Australia, Australia

Yin Gao
Division of Rheumatology
Department of Biomedical and Molecular Sciences and Department of Medicine
Queen's University
Kingston, Ontario, Canada

Angela M. Gronenborn
Department of Structural Biology
School of Medicine
University of Pittsburgh
Pittsburgh, Pennsylvania

Kay Grünewald
Oxford Particle Imaging Centre
Division of Structural Biology
Wellcome Trust Centre for Human Genetics
University of Oxford
Oxford, United Kingdom

Cy M. Jeffries
Bragg Institute
Australian Nuclear Science and Technology Organisation
Kirrawee DC, New South Wales, Australia

Leonardus M. I. Koharudin
Department of Structural Biology
School of Medicine
University of Pittsburgh
Pittsburgh, Pennsylvania

Daniel Kolarich
Department of Biomolecular Systems
Max-Planck-Institute of Colloids and Interfaces
Potsdam, Germany

Ricardo L. Mancera
Curtin Health Innovation Research Institute
and
Western Australian Biomedical Research Institute
and
School of Biomedical Sciences
and
School of Pharmacy
Curtin University
Perth, Western Australia, Australia

Nicolle H. Packer
Biomolecular Frontiers Research Centre
Department of Chemistry and Biomolecular Sciences
Macquarie University
Sydney, New South Wales, Australia

Serge Pérez
Centre de Recherches sur les Macromolécules Végétales
Centre National de la Recherche Scientifique
and
European Synchrotron Research Facility
Grenoble, France

Paul A. Ramsland
Centre for Immunology
Burnet Institute
and
Department of Surgery Austin Health
University of Melbourne
and
Department of Immunology
Monash University
Melbourne, Victoria, Australia

Mauro S. Sandrin
Department of Surgery
Austin Health/Northern Health
University of Melbourne
Heidelberg, Victoria, Australia

Anita Sarkar
Centre de Recherches sur les Macromolécules Végétales
Centre National de la Recherche Scientifique
Grenoble, France

David F. Smith
Department of Biochemistry
School of Medicine
Emory University
Atlanta, Georgia

Xuezheng Song
Department of Biochemistry
School of Medicine
Emory University
Atlanta, Georgia

Ian A. Wilson
Department of Molecular Biology
and
Skaggs Institute for Chemical Biology
The Scripps Research Institute

Rui Xu
Department of Molecular Biology
The Scripps Research Institute
La Jolla, California

Elizabeth Yuriev
Medicinal Chemistry
Monash Institute of Pharmaceutical Sciences
Monash University
Melbourne, Victoria, Australia

Tzviya Zeev-Ben-Mordehai
Oxford Particle Imaging Centre
Division of Structural Biology
Wellcome Trust Centre for Human Genetics
University of Oxford
Oxford, United Kingdom

Section I

Experimental Techniques to Determine Three-Dimensional Structures

1 Crystallography and Small-Angle Scattering of Carbohydrate–Protein Complexes and Glycoproteins

Cy M. Jeffries, William Farrugia, and Paul A. Ramsland

CONTENTS

1.1 INTRODUCTION

Carbohydrate-binding proteins and glycoproteins are the focus of intense scientific investigation due to their central role in diverse biological processes that include, but are not limited to, immunity and infection, cellular adhesion, and cellular communication and signaling. Yet, our understanding of the fundamental molecular mechanisms through which carbohydrate-binding proteins and glycoproteins realize their functions still remains underdeveloped. Of note, while it is estimated that over 50% of all eukaryotic proteins are glycosylated (Apweiler et al. 1999), only around 5% of the three-dimensional (3D) structures deposited in structural databases such as the Protein Data Bank (PDB) (Berman et al. 2000) include proteins with *N*- or *O*-linked carbohydrates (often called glycans). Similarly, only around 7% of all PDB entries contain information on protein/carbohydrate systems (covalently or noncovalently bound to proteins), and there are even fewer examples of high-resolution structures where the associated carbohydrate components have been fully resolved (Lutteke 2009).

In this chapter, we discuss two highly complementary experimental approaches for probing the 3D structures of carbohydrate-binding proteins and glycoproteins: (1) x-ray crystallography that can provide high-resolution details of macromolecular 3D structures and (2) small-angle scattering that provides global structural parameters and shape information from proteins in solution. Using select examples, we summarize the structural basis for carbohydrate recognition and the role of multivalency through oligomerization as revealed by x-ray crystallography, while small-angle scattering is highlighted as a powerful strategy to probe the states and shapes of the intact glycoproteins without the conformational constraints imposed by the crystal matrix.

1.2 STRUCTURE DETERMINATION BY X-RAY CRYSTALLOGRAPHY

X-ray crystallography is a powerful means and currently the most commonly used experimental methodology for determining 3D structures of biological macromolecules. The basic approach and methodology for determining 3D structures by crystallography is similar for all biological macromolecules and has been described in detail elsewhere. For a comprehensive and easily accessible reference on macromolecular crystallography, the reader is referred to the excellent textbook on *Biomolecular Crystallography* by Bernhard Rupp (2010). Herein, we provide a brief overview of the steps involved in 3D structure determination by x-ray crystallography and some of the specific considerations required when working with carbohydrate–protein complexes and glycoproteins.

The first and most crucial step in any crystallography project is growing a single crystal that diffracts x-rays to suitably high resolution (in practice, normally between 3.0 and 1.0 Å) for 3D structure determination. This largely empirical process is achieved via screening highly purified material against numerous crystallization conditions that typically contain dehydrating or precipitating agents (e.g., polyethylene glycol and ammonium sulfate) and a variety of additives (e.g., buffers and metal ions).

The most common crystallization method is vapor diffusion where a small droplet containing the sample and a crystallization solution is equilibrated against a larger reservoir of the same crystallization media. Crystallization screening against hundreds of individual conditions is performed in parallel in multi-well plastic plates and can be highly automated using robotics. Once a crystal is obtained—often from one or a handful of specific conditions—x-ray diffraction data can be collected. However, to determine a structure, the crystallization process may require a number of rounds of optimization to produce crystals with improved diffraction intensity and resolution (for further reading, see the comprehensive textbook on *Protein Crystallization* by McPherson [1999]). Since simple carbohydrates are typically highly soluble and are relatively small, it is possible to either co-crystallize these ligands with the target protein or soak into the hydrated crystals of the carbohydrate-binding protein. The affinity of carbohydrates for protein-binding sites is generally quite low (K_d values in the millimolar to micromolar range); thus, it is often beneficial to use a molar excess (e.g., 10- to 100-fold) of the carbohydrate over the protein to ensure that high occupancy is achieved by the ligand in the carbohydrate-binding site during the crystallization or crystal-soaking process.

The natural heterogeneity of *N*- and *O*-linked glycans results in glycoproteins being mixtures of glycosylated variants or glycoforms (Marino et al. 2010). Consequently, generating crystals of glycoproteins with a well-ordered (uniform) crystal matrix and which diffract x-rays to high resolution can often be a very frustrating enterprise. Approaches to crystallize glycoproteins have included the production of recombinant proteins in bacterial systems such as *Eschericha coli* that essentially lack glycosylation machinery or in eukaryotic systems such as insect cell lines (or engineered mammalian cell lines) that add carbohydrates of reduced complexity and increased homogeneity compared to unmodified mammalian cells (Nettleship et al. 2010). Alternatively, site-directed mutagenesis of the glycoprotein can be used to remove some or all of the glycosylation motifs from the protein to obtain crystals for structure determination. Another method of increasing the quality of crystals is to truncate the carbohydrates with the most accessible technique being the removal of terminal sialic acid residues with neuraminidase (Lustbader et al. 1989). However, the removal of sialic acids seems to have been often overlooked as a simple method for generating high-quality crystals of glycoproteins. We suggest that neuraminidase treatment should be routinely trialed for crystallizing mammalian glycoproteins (particularly with proteins purified directly from primary sources) such as we found useful for generating crystals of a glycosylated antigen-binding fragment (Fab) from an IgM cryoglobulin, which was purified from the plasma of a Waldenström's macroglobulinemia patient (Ramsland et al. 2006). For a detailed example of the wide variety and potential of glycosylation modification strategies, see the excellent study by Lee and colleagues (Lee et al. 2009) who successfully determined the crystal structure of the Ebola virus trimeric spike glycoprotein (Lee et al. 2008).

Most macromolecular diffraction data is currently collected using laboratory or high-intensity synchrotron x-ray sources from crystals that have been cryoprotected at low temperature (around 100 K using liquid N_2 cooling systems) to reduce the effects of ionizing radiation and thermal damage. Diffraction data is often obtained from a single crystal that is precisely rotated (around at least one axis) to collect a

series of diffraction images that result from passing an intense x-ray beam through the crystal. The positions and intensities of numerous diffraction "spots" obtained from the crystal at each angle are integrated into a unique dataset using readily available computer algorithms. The electron density is reconstructed by combining the Fourier transformation of these diffraction data with the derived phases, which may need to be determined experimentally or calculated using molecular replacement (MR) methods that are based on fitting previously determined homologous protein structures to the experimental data (Rupp 2010).

One possible approach to solving the "phase problem" with carbohydrate-binding proteins has been proposed that uses selenium derivatives of the native carbohydrate ligands for multi-wavelength anomalous dispersion (MAD) phasing experiments. This strategy was successfully used for the three-wavelength MAD phasing of a bacterial adhesin F17-G in complex with an *N*-acetyl-d-glucosamine derivative, where the anomeric oxygen was replaced by a selenium atom (Buts et al. 2003). However, 3D structures of most carbohydrate-binding proteins have been determined from phases obtained using heavy-atom crystal derivatives, MAD phasing from selenomethionine-substituted recombinant proteins, or MR.

The final step in crystal structure determination is the iterative process of crystallographic refinement where the 3D electron density map is progressively fitted (automatically and manually) with a molecular model and a variety of parameters are optimized that describe the correlation between that 3D model and the observed experimental data (e.g., atom positions, temperature B-values, and structure factor intensities or amplitudes). Well-established computational approaches are available for crystallographic model building and refinement (Rupp 2010) and these are not described here.

1.3 CRYSTAL STRUCTURES OF CARBOHYDRATE–PROTEIN COMPLEXES

A wide range of carbohydrate-binding proteins have been now characterized by x-ray crystallography such as antibodies, lectins (from plants, fungi, and animals), carbohydrate-binding proteins of pathogens, transport proteins, and enzymes. This section illustrates the basic principles of carbohydrate recognition using select examples of crystal structures of carbohydrate–protein complexes, including anti-carbohydrate antibodies, mammalian lectins involved in innate immunity, and proteins from pathogens.

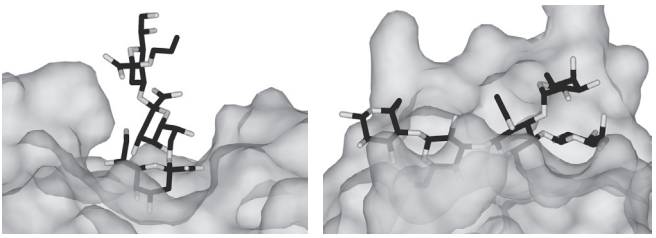
1.3.1 General Features of Carbohydrate–Protein Interactions

Carbohydrate-binding sites are generally located on the surface of proteins and form cavities or grooves. Most amino acids can participate in binding carbohydrates, although there is a frequent over-representation of amino acids with polar, charged, and aromatic side-chains lining carbohydrate-binding sites. Hydrophobic interactions primarily between aromatic residue side-chains (e.g., Tyr and Trp) and the more hydrophobic regions (faces) of carbohydrate rings are known to be important contributors to the affinity of carbohydrate–protein interactions. In addition to amino

acids, the relatively solvent-exposed binding sites contain numerous water molecules, which play a pivotal role in carbohydrate–protein interactions. Bound metal ions also provide further carbohydrate coordination centers for molecular recognition events and can act in parallel as critical structural components that help maintain the binding site shape. The multivalent binding of carbohydrates is also a frequent feature of carbohydrate-binding proteins and is typified by the repetition of carbohydrate recognition domains within a polypeptide chain and/or the oligomerization of protein subunits to generate multiple binding sites for carbohydrate recognition. Thus, a relatively low-affinity carbohydrate–protein binding site interaction is converted into a high-strength (avidity) interaction through multivalent carbohydrate binding.

1.3.2 Common Carbohydrate-Binding Modes

Two common binding modes that have been repeatedly observed in crystal structures of carbohydrate–protein complexes are end-on insertion and groove-type binding (Figure 1.1). End-on insertion involves the terminal groups of the carbohydrate ligand, normally a terminal monosaccharide unit, entering first and most deeply into the carbohydrate binding site. End-on insertion has been observed for antibodies in binding small molecules such as haptens and carbohydrates and appears to be the predominant manner in which carbohydrate epitopes are recognized by antibodies (Ramsland et al. 2003). Such binding allows the antibody to specifically interact with unique determinants (epitopes) that are presented near the terminal ends of longer carbohydrate chains conjugated to proteins or lipids. Frequently, the epitopes targeted by antibodies consist of minimal determinants, often ranging in size from disaccharides to tetrasaccharides, which are easily accommodated by the combining site that is formed by the association of the heavy and light chain variable domains (Agostino et al. 2012).



(a) (b)

FIGURE 1.1 Two common binding modes involved in carbohydrate recognition by proteins. (a) End-on insertion of a trisaccharide, Kdoα(2-8)Kdoα(2-4)Kdo, in the binding site of an antibody, Se25-2 Fab, determined at 1.49 Å resolution, PDB ID: 3SY0 (Nguyen et al. 2003). (b) Groove-type interaction of a pentasaccharide, Galβ(1-4)[Fucα(1-3)]GlcNAcβ(1-3)Galβ(1-4)Glc (lacto-N-fucopentaose III, a Lex tumor-associated antigen), in the binding site of the amino-terminal domain of human Galectin-8 at 1.33 Å resolution, PDB ID: 3AP9 (Ideo et al. 2011). Molecular surfaces are shown for the proteins and the bound carbohydrate ligands are in stick representations with carbon atoms in black and polar atoms in gray.

An example of where end-on insertion is used to recognize carbohydrates is the Se25-2 antibody that uses a germline-encoded binding site that interacts with the terminal 3-Deoxy-D-*manno*-oct-2-ulosonic acid (Kdo) residues used by certain bacteria to form lipopolysaccharides (LPS). The 1.49 Å resolution structure of Kdoα(2-8) Kdoα(2-4)Kdo in complex with Se25-2 Fab (Nguyen et al. 2003) shows how the terminal Kdo residue penetrates a cavity while antibody combining site residues could potentially participate in further interactions with the second and third carbohydrate residues in the chain (Figure 1.1a).

Lectins have also been shown to utilize end-on insertion binding for recognition, but in addition often employ groove-type binding, where an extended carbohydrate chain is bound in a solvent-filled groove (Yuriev et al. 2009). Binding of carbohydrate chains in a groove allows proteins to interact with internal carbohydrate moieties, which can partially explain the cross-reactivity with different types of complex carbohydrates by many lectins. It should be noted that antibodies can also participate in groove-type binding, but that this appears to be less frequent than for lectins and other classes of carbohydrate-binding proteins.

Galectins are a class of lectins found in mammals that form part of the innate immune system and recognize β-galactoside (βGal)-containing carbohydrates. Human Galectin-8 has been crystallized with a pentasaccharide ligand, Lacto-*N*-fucopentaose III (LNFIII, a carbohydrate containing the Lewis x, Le^x, trisaccharide epitope). The crystal structure of the Galectin-8 complex with LNFIII was determined at a resolution of 1.33 Å as well as complexes with other related carbohydrate ligands (Ideo et al. 2011). The LNFIII pentasaccharide, Galβ(1-4)[Fucα(1-3)] GlcNAcβ(1-3)Galβ(1-4)Glc, is bound in an extended conformation by an elongated groove that is open at both ends and located in the amino-terminal domain of Galectin-8 (Figure 1.1b). Conserved binding interactions (seen in the other Galectin-8 complexes) occur with the lactose Galβ(1-4)Glc disaccharide portion of the ligand (Ideo et al. 2011), while additional interactions occur with the central GlcNAc residue and the terminal βGal stacks against a Tyr side-chain in the binding site. The α1,3-linked fucose residue that is part of the terminal Le^x epitope does not contact the protein, but participates in a further stacking interaction with the terminal βGal (opposite face to the binding site Tyr residue) as expected for Lewis-type carbohydrate antigens (Yuriev et al. 2005).

1.3.3 Anchored Binding of Carbohydrate Ligands

Carbohydrates are often anchored in the binding site by tight interaction with a monosaccharide subunit of the carbohydrate chain. Two major types of anchored binding are metal ion mediated (e.g., calcium) and charge neutralization or compensation of terminal sialic acid residues (Figure 1.2).

Metal ion–mediated anchoring of carbohydrate ligands is exemplified by a family of innate effector molecules called collectins, which are members of the larger group of C-type (Ca^{2+}-dependent) lectins (Seaton et al. 2010; Veldhuizen et al. 2011). Up to four Ca^{2+} ions can be bound to C-type lectin domains and not all these metal ions directly interact with carbohydrates and have been proposed to have roles in stabilizing the domains. In particular, a single Ca^{2+} is held in place in the

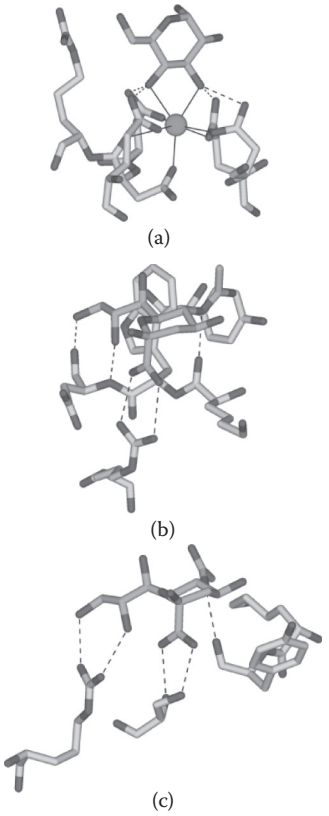


FIGURE 1.2 (See color insert.) Anchored binding of terminal carbohydrate residues. (a) Calcium-mediated coordination of a glucose residue of maltose in the binding site of human lung surfactant protein D at 1.40 Å resolution, PDB ID: 1PWB (Shrive et al. 2003). (b) Sialic acid (Neu5Ac) in the binding site of Siglec-7 at 1.90 Å resolution, PDB ID: 2DF3 (Attrill et al. 2006). (c) Sialic acid (Neu5Ac) in the binding site of rhesus rotavirus protein VP4 at 1.40 Å resolution, PDB ID: 1KQR (Dormitzer et al. 2002). Atoms are colored by type: C, yellow (protein) and cyan (carbohydrate); N, steel blue; O, red; Ca, green. Hydrogen bonds are shown as dashed black lines and metallic ion coordination bonds shown as solid black lines. Only the terminal carbohydrate residues are displayed for clarity.

carbohydrate-binding site by six coordination bonds with conserved residues (normally Asp, Glu, Asn, or Gln) and is involved in the coordination of two waters or the hydroxyl groups of bound monosaccharides. Thus, carbohydrates are anchored in the site by the strong pairing with Ca^{2+} and specificity for different carbohydrates (e.g., mannose or galactose) is determined by subtle differences in the residues in the Ca^{2+} binding pocket (Weis and Drickamer 1996). The crystal structure of human lung surfactant protein D (SP-D) has been determined in complex with maltose, Glcα(1-4)Glc, at 1.40 Å resolution (Shrive et al. 2003). The coordination of the terminal αGlc (the C2 epimer of mannose that also binds SP-D) occurs between the 3- and 4-hydroxyls and the bound Ca^{2+} ion, which provides a clear example of metal ion–mediated anchoring of carbohydrate antigens (Figure 1.2a).

Sialic acids are widely distributed in animal tissues (mostly Neu5Ac and Neu5Gc), as glycolipids (e.g., gangliosides) or at the ends of complex *N*-glycans, and are important biological ligands for many physiological recognition events and host–pathogen interactions. Their location at the termini of carbohydrate chains and the negatively charged carboxylate group make them ideal candidates for anchored binding to proteins. The most obvious anchoring mechanism is charge neutralization by formation of ion-pairs between the sialic acid carboxylate and basic residues (Arg and Lys) of the carbohydrate-binding protein. Ion-pairing between the sialic acid carboxylate anion and the guanidinium cation of an Arg residue is a key interaction of Siglec (sialic acid immunoglobulin-like lectin) receptors and is illustrated with the Siglec-7 interaction with Neu5Ac (Figure 1.2b). Siglec-7 was co-crystallized with a larger tetrasaccharide ligand, but only the terminal Neu5Ac is depicted, as this is involved in anchored binding to Siglec-7 (Attrill et al. 2006). An alternate mode of sialic acid recognition involves charge compensation of the carboxylate anion by the formation of strong hydrogen bonds with this portion of the ligand. The rhesus rotavirus VP4 carbohydrate recognition domain binds sialic acid by anchoring through multiple hydrogen bonding interactions and the sugar and the carboxylate form two hydrogen bonds with the side-chain hydroxyl and main chain amide of a serine residue in the binding site (Dormitzer et al. 2002). Thus, VP4 is an example where a protein uses hydrogen bonding for charge compensation to anchor sialic acid residues (Figure 1.2c).

1.3.4 Role of Water in Carbohydrate–Protein Interactions

Water is a critical component that both drives carbohydrate binding and contributes to the specificity of carbohydrate–protein interactions within carbohydrate binding pockets. The strength of x-ray crystallography is it has allowed investigators to show that, while most of the bulk solvent is displaced from a carbohydrate-binding site when a target carbohydrate binds, certain ordered water molecules remain and are integral to maintaining architecture and specificity of a carbohydrate-binding site. In particular, ordered water molecules frequently participate in extensive hydrogen bonding networks that form the carbohydrate–protein interaction. For example, we previously observed the involvement of seven water molecules in forming a hydrogen bonding network linking the Le^y tetrasaccharide to the binding site of a humanized antibody (hu3S193), for which the hu3S193 Fab complex with Le^y was determined at 1.90 Å resolution (Ramsland et al. 2004). The role of water in this and other Lewis carbohydrate systems was further examined by independent molecular dynamics studies (Reynolds et al. 2008), which was in general agreement that the water is directly involved both in determining specificity and maintaining the conformation of carbohydrate antigens.

Recently, Saraboji and colleagues have determined a series of ultra-high-resolution crystal structures of Galectin-3 both in its unliganded (apo) form and in complexes with lactose and glycerol (Saraboji et al. 2012). The 0.86 Å resolution crystal structure of the Galectin-3 complex with lactose contains at least 10 ordered water molecules directly engaging the carbohydrate ligand or acting as hydrogen-bonded bridges between carbohydrate- and protein-binding site residues (Figure 1.3a). When the 1.08 Å apo-structure of Galectin-8 was compared, five of the same ordered water molecules involved in bridging carbohydrate–protein were

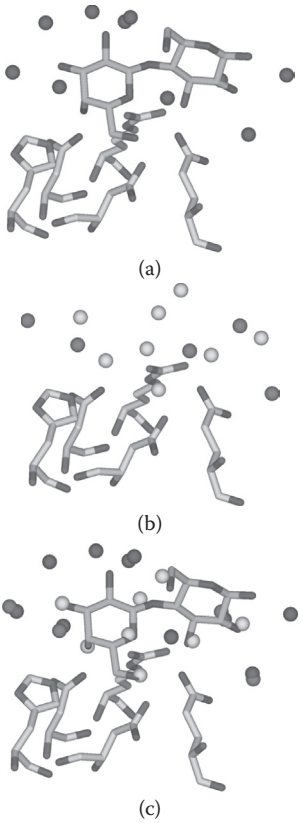


FIGURE 1.3 **(See color insert.)** Central role of water in protein recognition of carbohydrates. (a) Close-up of water (red spheres) network around the lactose interaction with Galectin-3 determined at 0.86 Å resolution, PDB ID: 3ZSJ. (b) Key binding-site waters (steel blue and light gray spheres) in the apo-structure (unliganded) of human Galectin-3 determined at 1.08 Å resolution, PDB ID: 3ZSL. (c) Structural overlay of the lactose-bound and apo crystal forms of human Galectin-3 depicted in panels A and B. Note that waters in the apo form that correspond to key binding-site waters from the lactose complex with Galectin-3 are shown in steel blue and those that superimpose with carbohydrate atoms are in light gray. The crystal structures depicted in this figure and related PDB entries are described in detail elsewhere (Saraboji et al. 2012).

maintained (Figure 1.3b, steel blue spheres). Interestingly, a further eight water molecules were observed that closely matched the positions of oxygen atoms from the bound lactose molecule (Figure 1.3b, light gray spheres). Thus, key water molecules, important for determining carbohydrate specificity, are maintained in what appears to be a pre-configured binding site ready to engage carbohydrate ligands. In addition, waters clearly occupy the same location as oxygen atoms in the bound carbohydrate (so may mimic the carbohydrate ligand), and are displaced upon entry of the carbohydrate into the binding site (see overlays in Figure 1.3c). Many other examples for the involvement of water in carbohydrate–protein interactions can be found in the literature, but are beyond the scope of this chapter.

1.3.5 Multivalency of Carbohydrate-Binding Proteins

A quick survey of entries in the PDB that contain the keyword "carbohydrate" reveals that for antibodies and lectins there appears to be a modular type of assembly and a capacity for the protein subunits to associate as multimers/oligomers (Figure 1.4). For the antibody crystal structures, there are two major populations at around 50 kDa (Fab and Fc regions) and at 100 kDa (dimers of Fab and Fc), which indicate that in

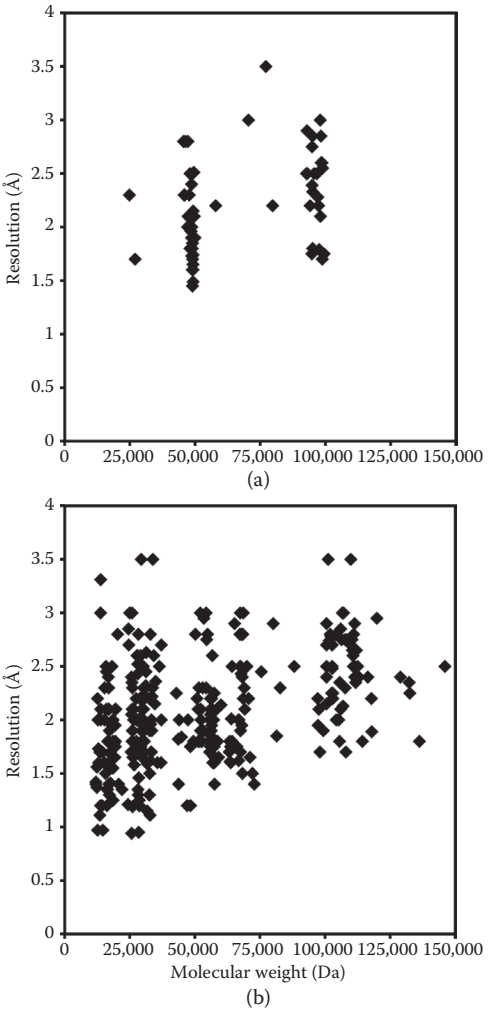


FIGURE 1.4 Distributions of resolutions and molecular weights for representative crystal structures of antibodies and lectins within the Protein Data Bank (PDB). (a) PDB entries containing the keywords "antibody" and "carbohydrate." The modular nature of antibody structures shows that most structures fall into monomers (Fab or Fc) or dimers within the crystals. (b) PDB entries containing the keywords "lectin" and "carbohydrate." Lectins also often contain repeating carbohydrate-binding domain architectures or form oligomers that appear to fall within 4–5 predominant size populations.

crystals Fab–Fab or Fc–Fc pairing is common (Figure 1.4a). For the lectins, there are at least four or five size groupings that correspond to the known tandem arrangement of carbohydrate-binding modules and stable oligomers that occur in this class of carbohydrate-binding protein (Figure 1.4b). While this survey is not comprehensive and contains some non-carbohydrate-bound structures, it is clear that there is a propensity for carbohydrate-binding proteins to form discrete oligomers in crystals. This observation is in agreement with the concept that multivalent binding to carbohydrates is often required for biological function (Dam and Brewer 2010; Weis and Drickamer 1996). The multivalency of carbohydrate–protein interactions results in high avidity and can overcome the low to modest affinities of most carbohydrate interactions with individual binding sites.

A variety of protein oligomers observed in crystals of carbohydrate–protein complexes are illustrated in Figure 1.5. The anti-carbohydrate antibody 2G12 binds to the high density of branched *N*-glycans on the human immunodeficiency virus (HIV) envelope protein, gp120. The crystal structures of the 2G12 Fab dimer, with mannose-containing saccharides, reveal how this antibody is geared toward binding the high densities of gp120 *N*-glycans (Calarese et al. 2003). The swapping of the variable domains of the heavy chains in 2G12 Fab forces a stable dimer, in crystals and the intact IgG antibody, with the carbohydrate-binding sites closely packed together to form a useful surface for multivalent carbohydrate recognition (Figure 1.5a). While this is an elegant immunological solution to carbohydrate cluster recognition, it is not widespread among antibodies. Other mechanisms such as through metal-ion-mediated Fab–Fab pairing (Farrugia et al. 2009) may be involved in carbohydrate cluster recognition by antibodies. In addition, antibodies form larger oligomers due to their covalent structure where IgG has two binding sites and polymeric IgM has at least 10 binding sites with the potential to interact with carbohydrates.

Similar to other collectins, the carbohydrate recognition domain "heads" of lung SP-D associate as a stable trimer through the presence of a coiled-coil "neck" region (Shrive et al. 2003). The carbohydrate ligands recognized by SP-D are located in binding pockets at one end of the molecule separated by around 45 to 50 Å (Figure 1.5b). The relatively flat carbohydrate-binding surface is suitable for multivalent recognition of repeating carbohydrate ligands on the surfaces of invading pathogens. Collectins can also form larger oligomers where several of the trimeric units associate to form oligomers or "fuzzy balls," which are highly multivalent macromolecular assemblies (Veldhuizen et al. 2011).

Carbohydrate-binding proteins from pathogens also tend to form as multivalent oligomers suitable for avid binding to host carbohydrate determinants. The AB_5 toxins are virulence factors of many bacterial pathogens and consist of a catalytic domain (A-subunit) and a multimeric host receptor-binding domain (pentamers of B-subunits), which display high avidity binding to glycans on target cells (Beddoe et al. 2010). The crystal structure of the pentameric B subunit of shiga-like toxin I (SLT I) of *Escherichia coli* in complex with an analog of the glycolipid Gb3 (trisaccharide epitope) shows how the AB_5 toxins can bind a large number of glycan ligands (Ling et al. 1998). Each B-subunit of SLT-1 interacts with three Gb3 molecules so that 15 glycans are bound on one face of the B5 oligomer (Figure 1.5c). A second example of a pathogen carbohydrate-binding protein is the sigma 1 (σ1) trimeric attachment

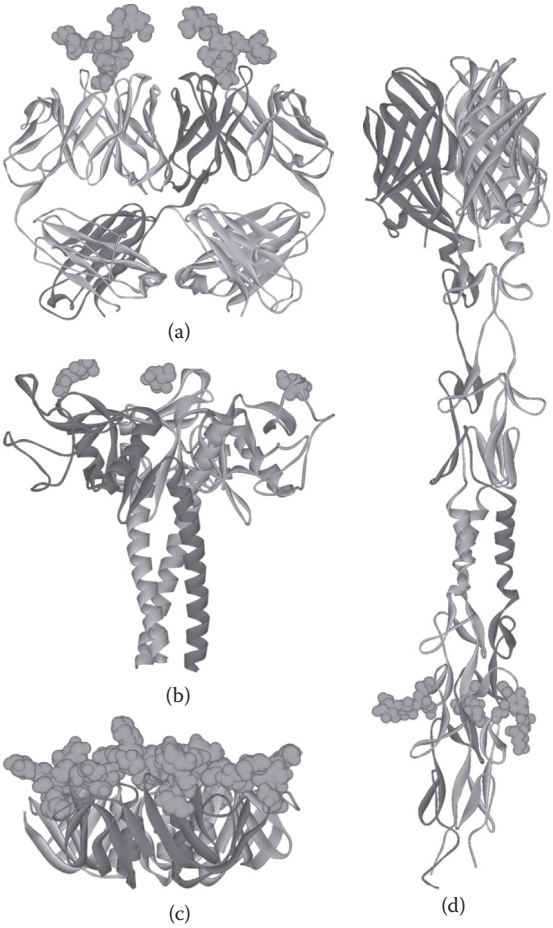


FIGURE 1.5 (See color insert.) Examples of crystal structures of protein oligomers involved in binding to carbohydrates. (a) The domain-swapped 2G12 Fab dimer with two bound high-mannose-branched complex oligosaccharides ($Man_9GlcNAc_3$) determined at 3.0 Å resolution, PDB ID: 1OP5 (Calarese et al. 2003). (b) Trimeric head and neck regions of human SP-D in complex with maltose determined at 1.40 Å resolution, PDB ID: 1PWB (Shrive et al. 2003). (c) Pentameric B subunit of shiga-like toxin I (SLT-I) of *Escherichia coli* in complex with an analog of the glycolipid Gb3 (globotriaosyl ceramide) determined at 2.80 Å resolution, PDB ID: 1BOS (Ling et al. 1998). (d) Trimeric reovirus attachment protein, sigma 1 (σ1), in complex with 3′-sialyllactose, Neu5Acα(2-3)Galβ(1-4)Glc, determined at 2.25 Å resolution, PDB ID: 3S6X (Reiter et al. 2011). Proteins are shown as ribbon-style representations and separate polypeptide chains in the oligomers are colored differently. Carbohydrate ligands are displayed as space-filling spheres in green.

protein of reovirus, which binds to sialyated glycans on host cells. The crystal structure of σ1 in complex with 3′-sialyllactose (Reiter et al. 2011) shows an unusual arrangement where the carbohydrates are bound to the extended trimeric "stalk" regions rather than the ends of the globular head domains of the attachment protein (Figure 1.5d). In the intact σ1 protein, the stalk is further extended by a long trimeric

coiled-coil region placing the carbohydrate-binding sites around the middle of this approximately 400 Å long pathogen cell attachment protein (Reiter et al. 2011).

Clearly, the diversity of oligomeric assemblies within host and pathogen carbohydrate-binding proteins (illustrated here by only a few examples) highlights the need to study the full-length proteins in their near physiological states. Such structural problems often require techniques complementary to x-ray crystallography such as small-angle scattering (of x-rays and neutrons) in aqueous solutions, which is discussed in the second part of this chapter (Sections 1.4 and 1.5).

1.4 PROBING MACROMOLECULAR STRUCTURES BY SMALL-ANGLE X-RAY AND NEUTRON SCATTERING

Small-angle x-ray and neutron scattering (SAXS and SANS) are powerful techniques that can complement x-ray crystallography to obtain global size and shape information from biological macromolecular systems in solution (Jeffries and Trewhella 2012; Mertens and Svergun 2010; Neylon 2008; Svergun 2010). Importantly, SAXS and SANS experiments require minimal amounts of purified material, can be performed in dilute solutions, and the conditions can be readily adjusted to simulate physiological environments (e.g., ionic strength and pH). Thus, conformations of macromolecules can be probed and cross-checked against the restricted and often single conformation observed in crystals. Structural parameters such as the radius of gyration (R_g), maximum dimension (D_{max}), and the probable distribution of atom-pair distances ($P(r)$ vs. r) of, or within, a macromolecule can be evaluated from the small-angle scattering data. Furthermore, it is possible to monitor (1) the oligomeric state of a protein, (2) how changes in sample environment or ligand binding can influence global protein conformation (He et al. 2009, 2003), and (3) how the formation of higher-order macromolecular complexes effects the overall structure of a macromolecule in solution (Wall et al. 2000).

Both SAXS and SANS are based on a difference existing either between the average electron density of a macromolecule relative to a supporting solvent (SAXS), or between the "isotopic composition/density" per unit volume of a macromolecule relative to a supporting solvent (SANS; in particular, hydrogen, ^{1}H, content per unit volume). This difference is known as contrast ($\Delta\rho$, where $\Delta\rho_{macromolecule} = \rho_{macromolecule} - \Delta\rho_{solvent}$). Assuming that contrast is present, and given that the total population of macromolecules within a sample is monodisperse, the rate at which scattering intensities decrease with increasing angle (after having subtracted solvent scattering contributions) will reflect the distribution of distances between scattering centers within a single particle. Models can be generated and fitted against the solvent-subtracted data to yield information on the overall volume and shape of a macromolecule in solution as well as the spatial dispositions of subunits within oligomers or even macromolecular complexes comprised of different subunits.

One of the advantages of SANS over SAXS (Lakey 2009) is that the isotopic composition/density, specifically the ratio of ^{1}H to deuterium (^{2}H) per unit volume in the solvent, can be easily altered so that $\Delta\rho$ can either be maximized to increase scattering signals from a macromolecule, or be reduced to minimize scattering signal contributions (Jeffries and Trewhella 2012). Therefore, SANS is particularly useful

for investigating higher-order macromolecular complexes. Relative to each other, different classes of macromolecules—proteins, lipids, DNA, and carbohydrates—have different ^{1}H content per unit volume and hence different neutron scattering "power." Therefore, it is possible to selectively "match out" neutron scattering contributions of one component of a complex relative to another by altering the contrast through adjustments of $^{1}H_2O$:$^{2}H_2O$ ratios in the solvent. Consequently, using SANS with contrast variation, shape information can be obtained from a whole complex as well as the shapes of the individual components within a complex: each different class of macromolecule will have a different match point at a particular $^{1}H_2O$:$^{2}H_2O$ ratio where their scattering contributions are minimized from the overall scattering profile. However, depending on the mass ratios of the components within a complex (scattering intensity is proportionate to the volume of a particle squared) the separation of each component's match point maybe narrow. The quality of the SANS scattering signal can be drastically improved by deuterating one component of a complex so as to radically alter the ratio of ^{1}H per unit volume. The selective deuteration of one component of a complex is very powerful for probing the shape of protein–protein complexes in solution as the incorporation of nonexchangeable ^{2}H into one protein enables a contrast difference to be set up between the ^{1}H and deuterated components of a complex and subsequent separation of their match points in solution (Jeffries and Trewhella 2012).

Aside from obtaining R_g, D_{max}, and $P(r)$ versus r, that in themselves can provide valuable insights into the solution states of macromolecules, what is of most interest to structural biologists is obtaining a sensible consensus model, or series of models, that best fit their scattering data. Advances in computational methods (Petoukhov and Svergun 2007) have seen the routine application of restoring molecular volumes and shapes of a macromolecule from SAXS and SANS data using easy-to-use *ab initio* methods (Franke and Svergun 2009) that do not require any prior knowledge regarding the structure of a macromolecule of interest. When combined with models derived from x-ray crystallography, NMR, or homology modeling (Petoukhov and Svergun 2005, 2006; Svergun et al. 1995) SAXS and SANS begin to open new frontiers with respect to building representative protein structures that cannot otherwise be accessed using x-ray crystallography alone. SAXS and SANS are especially powerful for the analysis of modular proteins with inherent structural flexibility, large macromolecular multi-component complexes, assemblies, and glycoproteins.

1.5 EXAMPLES OF GLYCOPROTEINS STUDIED BY CRYSTALLOGRAPHY AND SMALL-ANGLE SCATTERING TECHNIQUES

An increasing number of structural biologists are beginning to integrate SAXS and SANS as complementary techniques into their research programs that focus on the molecular foundations of glycoprotein structure and function. As SAXS and SANS can be performed on natively glycosylated proteins and usually require similar or smaller amounts of sample when compared to crystallography (using

synchrotron-SAXS it is possible to obtain quality x-ray scattering data from as little as 10 μL of protein at 1–5 mg·mL^{-1} in 1–10 seconds), then, at the very least, scattering techniques provide an invaluable complementary tool to probe the native global states and the shapes of glycoproteins in solution.

1.5.1 Crystallography of AMIGO-I Combined with SAXS of AMIGO-II and -III

Kajander et al. (2011a) provided an elegant example of where SAXS has been employed in combination with crystallography to probe the structures of a group of related glycoproteins. For example, the transmembrane AMIGO proteins that are required for regulating neuronal growth, mobility, and adhesion (Kuja-Panula et al. 2003), which share a similar leucine-rich repeat (LRR) domain that is responsible for modulating protein–protein interactions (Chen et al. 2006; Kajava and Kobe 2002). The crystal structure was determined for the glycosylated neuronal protein AMIGO-I and this was used to develop SAXS-based models for the related and more heavily glycosylated protein AMIGO-II as well as AMIGO-III (Kajander et al. 2011a).

AMIGO-I crystallized as a horseshoe-shaped dimer with a twofold rotational axis between each AMIGO-I monomer. The LRR domains of opposing sub-units pack together to form the base of the horseshoe while separate LRRCT capping and immunoglobulin (Ig) domains extend from the base to form the individual arms of the U-shaped horseshoe (Figure 1.6a). Interestingly, SAXS data revealed that AMIGO-II and AMIGO-III have significantly increased radii of gyration when compared to the AMIGO-I crystal structure (AMIGO-I, R_g crystal = 30 Å; AMIGO-II, R_g = 46 Å; AMIGO-III, R_g = 40 Å). These results indicate that all three related proteins have very different mass distributions and hence different domain orientations with respect to each other. Indeed, the resulting estimates of D_{max} derived from the $P(r)$ versus r profiles of AMIGO-II and AMIGO-III show them to be significantly more extended than the AMIGO-I crystal form (AMIGO-I, D_{max} crystal = 100 Å; AMIGO-II, D_{max} = 170 Å; AMIGO-III, D_{max} = 135 Å). Unfortunately, the crystallizable AMIGO-I isoform could not be studied by SAXS (for a direct comparison to SAXS of AMIGO-II and AMIGO-III) due to nonspecific protein aggregation issues. However, rigid-body modeling of AMIGO-II and AMIGO-III against their respective SAXS datasets, using the crystal structure of AMIGO-I as a rigid-body template, shows that the best-fit models are those where the position of the LRR domains at the dimer interface are reasonably preserved across all three proteins and that the structural extension involves the arms of the U-shaped molecule opening up through large movements of the C-terminal Ig domains (Figure 1.6a). Since the Ig domain of the AMIGO proteins is a primary site of glycosylation, one of the interesting hypotheses from the crystallographic/SAXS study is that the flexibility between the Ig domains and LRRs of AMIGO-I, AMIGO-II, and AMIGO-III is modulated by glycosylation, which may affect the orientation of the LRRs within the intracellular space to facilitate the formation of AMIGO LRR-LRR mediated intercellular *trans*-dimers and, consequently, promote neuronal cell adhesion (Kajander et al. 2011a).

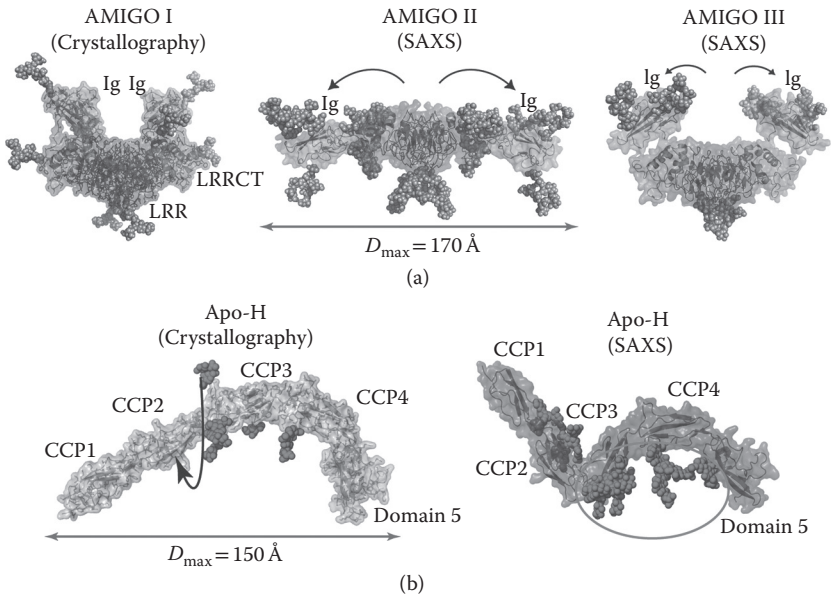


FIGURE 1.6 **(See color insert.)** Crystallography in combination with solution SAXS of glycoproteins: AMIGO proteins and Apo-H. (a) x-ray crystal structure of AMIGO-I and a comparison with the AMIGO-II and AMIGO-III models determined from solution SAXS data. The LRR domains (blue) self-associate to drive dimerization between AMIGO monomers and are capped off with the LRRCT domains (red) from which extends an Ig domain (teal). The spatial orientation of the Ig domains differs between each of the three AMIGO isoforms. The *N*-glycans of the AMIGO proteins were modeled and are represented by gray spheres. The images were generated from PDB coordinates kindly provided by Dr. Tommi Kajander (Kajander et al. 2011a). (b) The crystal structure of Apo-H, PDB ID: 1C1Z (Schwarzenbacher et al. 1999) shows that the CCP modules and domain 5 of the protein adopt a J-shaped conformation (left), while SAXS reveals that Apo-H undergoes a conformational shift in solution to form an S-shape and that the *N*-glycans (red spheres) appear to form a carbohydrate patch (marked with an ellipse) along one side of the proteins. The SAXS data is represented as a model derived using the crystallographic coordinates (PDB ID: 1C1Z) as well as information and a figure presented in the original Apo-H paper (Hammel et al. 2002).

1.5.2 Different Conformations of Apolipoprotein H (β_2-Glycoprotein I) Observed by Crystallography and SAXS

Apolipoprotein H (Apo-H or β_2-glycoprotein I) is a highly glycosylated protein (~19% w/w) that comprises five domains: four complement control protein (CCP) domains and a unique fifth domain, domain V (Bouma et al. 1999; Schwarzenbacher et al. 1999). Apo-H is involved in triggering the blood coagulation cascade (Brighton et al. 1996) and has apparent high affinity for heparin, cell membranes, macrophages, and phospholipids (Balasubramanian and Schroit 1998; Del Papa et al. 1998; Schousboe and Rasmussen, 1988). Complexes of Apo-H and phospholipids have been suggested to act as antigens for autoimmune phospholipid antibodies that are