

**GALECTIN-1 BINDING TO LACNAC AND 3'-O-SUBSTITUTED LACNAC LIGANDS.
A STUDY OF NON-COVALENT INTERACTIONS IN A LECTIN – CARBOHYDRATE
COMPLEX**

by

MICHAEL GARRETT FORD

(Under the Direction of Robert J. Woods and James H. Prestegard)

ABSTRACT

Galectins are a class of carbohydrate binding proteins, initially characterized by their affinity for galactose terminating ligands, that play important roles in cellular adhesion and development. There are considerable data on the types of saccharides which bind to the galectin-1 dimer, with higher affinity ligands containing the core sequence Gal- β -(1,4)-GlcNAc or LacNAc. While initially described as binding specifically to galactose-terminating ligands, a number of important 3'-O-substituted structures are found which are physiologically significant. The capping of galactose with 3'-O-SO₃ increases affinity, as does polymerization of LacNAc units as seen in polylactosamine structures found in basement membrane glycoproteins. The addition of sialic acid is another especially significant carbohydrate modification regulating the ability of glycans to bind to galectin-1. While glycans containing a Neu5Ac- α -(2,6)-LacNAc capping structure are unable to bind galectin-1, Neu5Ac- α -(2,3)-LacNAc terminated glycans can bind to galectin-1, with an affinity very similar to the LacNAc

disaccharide. The competing actions of sialyltransferases can therefore act as a regulator of galectin-1 binding. The modification of lactosamine in N- and O-linked glycans is thought to be the determining switch in the induction of apoptosis by galectin-1 in developing thymocytes. The clonal selection of immature thymocytes is dependent on this apoptotic process for the production of T-cells capable of discriminating self/non-self antigens.

In order to provide a structural framework for the observations on galectin-1 ligand binding, we have utilized an X-ray structure of the galectin-1 LacNAc complex. Structural information on the complexes of the 3'-O-substituted ligands bound to the carbohydrate recognition domain of galectin-1 is obtained from molecular dynamics studies. The resulting molecular trajectories of galectin-1 – ligand complexes are not only able to reproduce geometric features of the X-ray structure, but also predict novel contacts between 3'-O- substituents and protein residues. Predicted binding energies of conformations taken from MD simulations are derived from the contributions to the binding free energy.

The effect of 3'-O- modification of LacNAc on binding can also be studied directly, by observing the properties of the protein – ligand complexes in solution NMR spectroscopy. Monitoring chemical shift changes of backbone resonances during ligand titration is an established measure of changes in the local environment of protein residues due to ligand binding, and such experiments require the assignment of protein backbone resonances. In addition to the traditional suite of 3D experiments normally performed to assign a protein, we have undertaken an assignment strategy which incorporates additional angular information, in the form of the residual dipolar coupling.

This approach not only assigned many residues in the galectin-1 binding site, but should also serve as a general tool for the assignment of larger proteins. Titration experiments on the assigned galectin-1 protein confirm a common binding site for LacNAc and Neu5Ac-LacNAc. The results provide a structural interpretation for the large body of *in vitro* binding data by providing a model in which galectin-1 mediated crosslinking of heterogeneous cell surface oligosaccharides is responsible for a variety of adhesion phenomena ascribed to this lectin.

INDEX WORDS: Galectin, lectin, molecular dynamics, NMR, protein, carbohydrate

**GALECTIN-1 BINDING TO LACNAC AND 3'-O-SUBSTITUTED LACNAC LIGANDS.
A STUDY OF NON-COVALENT INTERACTIONS IN A LECTIN – CARBOHYDRATE
COMPLEX**

by

MICHAEL GARRETT FORD

B.A., University of Michigan, 1994

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

© 2004

Michael Garrett Ford

All Rights Reserved

**GALECTIN-1 BINDING TO LACNAC AND 3'-O-SUBSTITUTED LACNAC LIGANDS.
A STUDY OF NON-COVALENT INTERACTIONS IN A LECTIN – CARBOHYDRATE
COMPLEX**

by

MICHAEL GARRETT FORD

Major Professor: Robert J. Woods
James H. Prestegard

Committee: J. David Puett
Michael Pierce

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

Acknowledgments

I would like to thank a number of people who were invaluable to the completion of my thesis. I would first like to thank my major professors Robert Woods and James Prestegard for all of their guidance and support, without which none of this would have been possible. I would also like to thank my committee, David Puett and Michael Pierce for their valuable suggestions. I would also like to thank Kelly Moremen and Thomas Weimar for collaborations which greatly helped me complete this work. I must recognize the contributions of a number of Woods and Prestegard lab members, as well as other CCRC and BCMB members, for the immense amount of effort that they have expended on my behalf. I would like to recognize not only the technical help and valued scientific discussion, but also the genuine friendship that has made my experience at the University of Georgia so special. Jarrod Barnes, Maral Basma, Renu Kadavelraj, Karl Kirschner, Ahamed Pastheril, Sarah Tschampel, Meredith Beckham, Jorge Gonzalez-Outerino, Cathy Bougault, Ziad Eletr, Lianmei Feng, Andrew Fowler, John Glushka, Hashim al-Hashimi, Anita Kishore, Laura Morris, Ron Seidel, Greg Wiley, Homayoun Valafar, Brian Woosley and Avinash Sujan. I would also like to thank Jon Moon, Peter Horyani and Jason Sterling for sharing a beer now and then to maintain sanity. Finally I would like to thank my many friends here in Athens and especially my wife Rosie, my dog Sidney and my son Caleb for reminding me of what is truly important in life.

Table of Contents

	Page
CHAPTER	
1 Introduction	1
2 Molecular Dynamics Simulation of Galectin-1 – Ligand Complexes	28
3 Protein NMR Resonance Assignment and the Residual Dipolar Coupling....	76
4 A Combined Approach for Protein Resonance Assignment Using Residual Dipolar couplings, C α Chemical Shifts and ϕ, ψ Values.....	99
5 Order Tensor Solutions and Interpretations of Molecular Symmetry.....	118
of the Galectin-1 Dimer	
6 Mapping the Galectin-1 Binding Site and Conclusions.....	126

Chapter 1

Introduction

Historical perspective of the study of lectins

Lectins are a group of proteins that bind carbohydrates. The study of lectins began with observations that certain plant extracts and snake venoms possessed the ability to agglutinate erythrocytes. This agglutination process is now known to be a lectin mediated process of crosslinking between cell surface oligosaccharides. The first observations of agglutination and haemolysis of erythrocytes by rattlesnake venom were made over 140 years ago by Mitchell¹. In 1888, Stillmark discovered that castor bean preparations could agglutinate cells in a species specific manner, the extracts contained the lectin which came to be called ricin². The molecular basis of this blood typing system remained undetermined until the carbohydrate dependent nature of the agglutination reaction was determined by Watkins and Morgan in 1952³. The selective binding ability of carbohydrate binding proteins gave rise to the name lectin, from the Latin *legere*, meaning to pick out or choose⁴. A historical view of the study of lectins is shown in Table 1.1.

The observation that lectins agglutinate certain cell types led to one of the first tests for the carbohydrate specificity of lectins. The ability of mono- and disaccharides to inhibit the agglutination reaction was used to determine the sugar composition, linkage and anomeric specificity of lectin binding partners. The definition of lectin now

Table 1.1 Key discoveries in the lectin field.

1860 – 1886 1902	Descriptions of agglutination and lysis of erythrocytes by snake venoms [Mitchell, Flexner and Noguchi] ^{1,5}
1888	Stillmark observes agglutination by castor bean extract (ricin) ²
1935 - 1945	Legume and eel agglutinins used as blood typing reagents Establishment of the ABO blood type system
1936	Concanavilin A found to bind carbohydrates [Sumner, Howell]
1952	Carbohydrate nature of blood group determinants established ³
1954	Term lectin first used ⁴
1972	First 3D structure of a lectin, Concanavilin A ⁶
1973 - 1975	Galactose – specific lectins (discoidins) found in slime mold
1974	Rabbit hepatic lectin (Ashwell receptor, asialoglycoprotein receptor) reported ⁷
1975	Isolation of electrolectin, and subsequent discoveries of galectins in numerous vertebrates ⁸
Modified from Kilpatrick 2002 ⁹ and Rudiger et al. 2000 ¹⁰	

includes all multivalent proteins, excluding antibodies, glycosidases, and glycosyl transferases, which bind carbohydrates. Lectins have been found in some prokaryotes, all major groups of eukaryotes, and are important for a number of biological processes.

Lectin families

There are now a dozen or more lectin families, as well as numerous ‘orphan’ lectins that do not seem to fall into any particular category (reviewed in¹¹). The lectin families are grouped together based upon functional similarities as well as ligand specificity. Table 1.2 highlights the best studied lectin groups.

The I-type lectins, also members of the immunoglobulin superfamily, are found in the nervous system as well as in cells involved in the innate immune response. The majority of I-type lectins are specific for glycans containing sialic acid (the siglecs), an anionic nine carbon sugar with over 40 derivatives, with a few members binding other

glycoconjugates¹². As type I transmembrane proteins, there is considerable interest in possible signaling mechanisms mediated by the siglecs.

Table 1.2. Features of major lectin families.

Lectin family	Ligand specificity	Distinctive feature	Proposed functions
C-type lectins	varied	Ca ⁺⁺ required for binding ligand	Innate immunity, fertilization, protein folding, protein localization, cellular migration
I-type lectins	Sialic acid	Immunoglobulin fold	Cell-cell adhesion
S-type lectins (galectins)	Galactose – containing oligosaccharides	Free cysteines required for ligand binding	Development, cell-cell and cell-matrix adhesion, immune response
P-type lectins	Man-6-PO ₄	Bind phosphorylated mannose residues	Targeting of lysosomal hydrolases

The P-type lectins recognize mannose-6-phosphate residues located on hydrolases and target them to the correct lysosomal compartment. The function of these lectins is critical for the formation of functional lysosomes¹³.

The C-type lectins are the largest family of lectins, and this class of lectin includes many of the best characterized carbohydrate binding proteins. The C-type lectins require calcium for carbohydrate binding and consist of a diverse group of lectins that vary widely in their carbohydrate binding specificity, protein structure and biological function.

The homing of leukocytes to the site of inflammation is dependent on a class of C-type lectins called the selectins. The interaction between L-selectin on the leukocyte cell surface and O-linked oligosaccharides on several cellular adhesion molecules of the vascular endothelium (GlyCAM-1, CD34) is mirrored by interactions between vascular E- and P-selectin with mucin-like glycoproteins on the leukocyte cell surface (reviewed

in ¹⁴). The selectins bind carbohydrates of the sialyl-Lewis family, that are presented in a multivalent fashion on cellular adhesion molecules found on the surface of the leukocytes and endothelial cells. The collectins, another C-type lectin subfamily, are involved in the innate immune response. Mannose binding proteins bind to mannose and fucose residues found on the surface of pathogenic bacteria and can both target them for immune clearance by macrophages, as well as activate the complement pathway. These C-type lectins typically possess multiple carbohydrate recognition domains (CRDs) fused to a collagen-like domain.

The asialoglycoprotein receptor is a membrane-bound lectin that binds terminal Gal and GalNAc residues on N-glycans and targets the glycoproteins for lysosomal degradation^{15,16}. This process is thought to be critical for regulation of protein turnover for circulating glycoproteins. The turnover of improperly folded proteins is controlled by calnexin and calreticulin, two similar calcium dependent lectins that control glycoprotein transport within the endoplasmic reticulum^{17,18}. They bind to terminal glucose residues on *N*-linked oligosaccharides and prevent the exit of improperly folded proteins from the ER.

Classification of the S-type lectins (galectins)

The first galactose – specific lectin (galectin) was isolated from the electric organ of the electric eel by Teichberg in 1975⁸, with subsequent discoveries of galectins from chick, cow, rat, fish and human sources. The galectins are also known as S-type lectins, due to the presence of cysteines on the protein surface. Originally thought to be restricted to metazoans, galectins have since been found in bacteria, viruses, sponges, fungi and *Arabidopsis*¹¹. Galectins are classified based on two major criteria, namely,

binding to galactose – containing oligosaccharides, and a high degree of homology in both the overall protein structure¹⁹⁻²¹ and sequence conservation of residues in the CRD^{11,22}. Galactose is a major component of vertebrate glycans and is found in both *N*- and *O*- linked glycoproteins, as well as in glycolipids. Notably, of the fourteen vertebrate galectins characterized to date, one of them, the Charcot-Leyden crystal protein, or galactin-10, binds to mannose, not galactose²³. The CLCP is grouped together with galectins based on structural and sequence homology alone.

Physical properties of galectins

The galectins share a number of characteristics with most cytosolic proteins. They are soluble, with acetylated N-termini, lack disulfide bridges and do not possess *N*-linked glycosylation sites. However, they are not restricted to the cytoplasm and can be found in the nucleus, cell surface, and extracellular space. The export mechanism for galectins remains an open question as they do not contain peptide localization sequences commonly seen in secreted proteins. A novel apocrine export mechanism has been proposed²⁴, in which specific vesicles export galectin molecules that have accumulated at the plasma membrane.

Galectin (lectin) binding to carbohydrate ligands

It has been established that galectins, and lectins in general, have a relatively low affinity for their carbohydrate ligands. In contrast to a DNA – protein or antigen – antibody complex, which can possess a dissociation constant (K_D) in the range of low micromolar to nanomolar, lectin – carbohydrate interactions are generally in the millimolar range. There are several important biological implications for the weaker

binding of this class of proteins. In general, cell adhesion is a dynamic process that involves the forming and breaking of new adhesions during growth, development and cellular migration. Multivalency of both lectin and oligosaccharide can serve to multiply the impact of the weaker interactions through avidity effects, as will be discussed.

Galectin-1 glycoprotein ligands

In the case of mammalian glycans, the usual *in vivo* context for galactose is the disaccharide Gal- β -(1,4)-GlcNAc, also called N-acetyllactosamine or LacNAc. LacNAc is found in both N- and O-linked oligosaccharides, as well as in some glycolipids. A representation of these oligosaccharides is shown in Figure 1.1, which shows LacNAc units containing the common substituent Neu5Ac, which is alpha linked to either the 3' or 6' hydroxyl group of galactose. The biological implications of carbohydrate modification, and its effects on galectin-1 affinity, will be discussed later. In a few cases, specific glycoproteins known to bind galectin-1 have been linked to functions such as cell-cell and cell-matrix adhesion and the signaling pathways governing mitosis and apoptosis.

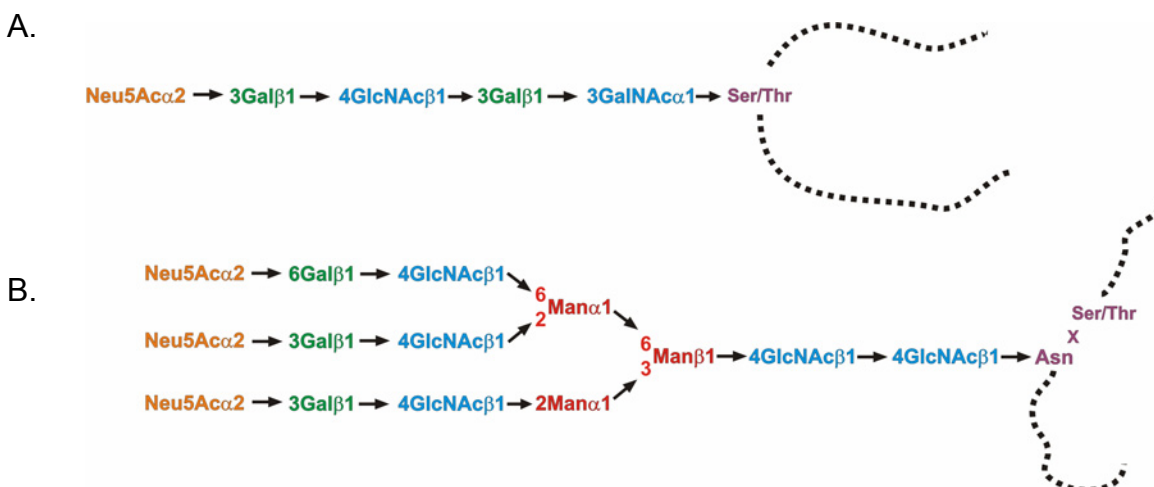


Figure 1.1 Covalent attachment of oligosaccharides in glycoproteins. A. O-linked glycans are attached to the protein by a linkage through the hydroxyl group of a serine or threonine residue. B. N-linked protein showing the attachment of a representative triantennary glycans to an asparagine side chain in the sequon Asn-X-Ser/Thr. (Modified from GlycoWord, <http://www.gak.co.jp/FCCA/glycoword/wordE.html>)

Laminin is a connective protein of the basement membrane that interacts with cell-surface receptors such as integrins to anchor cells to the basement membrane. Laminin is a glycoprotein possessing long poly-LacNAc chains which have been proposed to bind to galectin-1²⁵. Fibronectin, is a large, multidomain protein of the basement membrane that associates with a number of matrix components and also possesses polylactosamine chains capable of serving as galectin-1 ligands²⁶.

The addition of galectin-1 causes the clustering of T-cell receptors CD45 and CD43. Microdomains of these clusters are visible in immunofluorescence staining of thymocytes²⁷, and their formation is thought to influence the intracellular signaling of these receptors²⁸. The glycan chains of CD45 have been characterized²⁹ and possess LacNAc units typical of galectin-1 ligands. The chain of events relating galectin-1 to receptor crosslinking and eventual apoptosis of thymocytes is the best characterized biological system defining a precise biological function for galectin-1³⁰.

Galectin-1 mediated apoptosis is an important regulatory mechanism of developing thymocytes. The selective induction of self-reactive thymocytes prevents immune recognition of self antigens. Through the action of glycosyl transferases, the composition of *N*- and *O*-glycans is altered such that the ability to bind galectin-1, and the resulting cascade of events leading to apoptosis is switched on and off. Core 2 GnT, which transfers GlcNAc residues to the GalNAc of core 1 antigen is the glycosyltransferase critical for the creation of polylactosamine chains on *O*-linked glycans, which bind to galectin-1 with hi avidity³¹. Similarly, the action ST6 Gal I sialyltransferase, which links sialic acid with an α -(2,6) linkage to the galactose residue of LacNAc creates a trisaccharide (Neu5Ac- α -(2,6)-LacNAc) which cannot bind the galectin-1 CRD. This modification is thought to be an important regulatory switch for susceptibility to galectin-1 – induced apoptosis of immature thymocytes³².

Multivalency in lectin – carbohydrate interactions

The agglutination of cells by lectins is a direct result of multivalency, a fundamental property of lectins. The CRDs of lectins usually occur in multiple copies, either within a single polypeptide chain or as part of non-covalently associated multimers. This feature has important implications for the function of this class of proteins and extends to the carbohydrate ligands as well.

The multimeric structure of lectins is clearly evident in the structure of galectin-1, that shows that the homodimeric protein contains identical CRDs with binding sites located at opposite ends of the dimer. Intricate 'bouquet' structures, in which collagen-like stalks present CRDs in bunches, have also been observed in the structure of mannose binding protein (MBP)³³. The oligosaccharide ligand itself may also be

multivalent, with multiple glycosyl residues being presented in linear or branched chains. A schematic view of multivalency in lectin – carbohydrate interactions is shown in Figure 1.2.

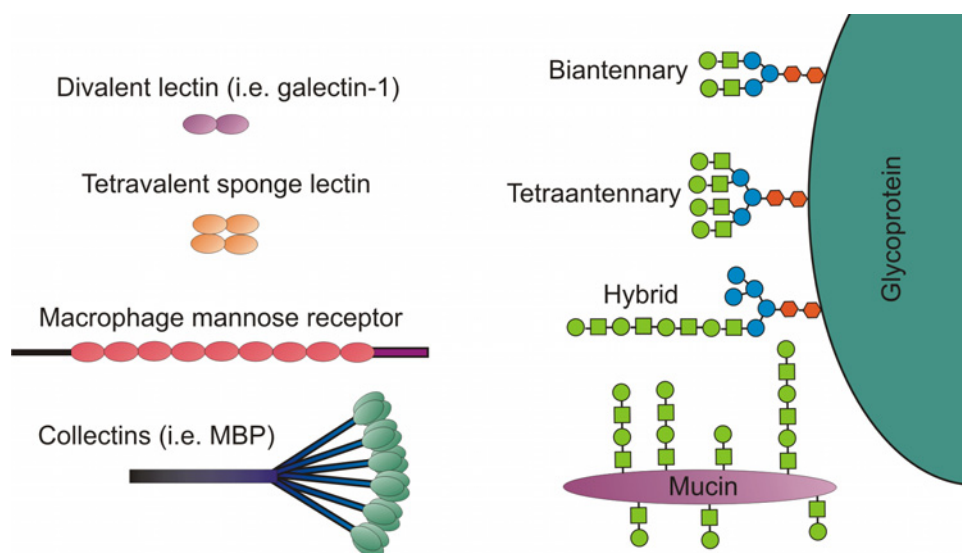


Figure 1.2 Multivalency in lectin and oligosaccharide structures. Biantennary, tetraantennary and hybrid structures are N-linked branched oligosaccharides. Ovals represent lectin CRDs, and representation of sugar units are as follows: Galactose - ●; GlcNAc - ■, ◆; Mannose - ●.

Galectin structure

The X-ray determined structures of a number of galectins -1,-2,-3,-7 and -10 are currently available³⁴. A β -sandwich motif is conserved throughout the galectin family, and is even seen in a number of legume lectins³⁵ and a mammalian glycanase³⁶, a fact which highlights the evolutionary conservation of this structural feature. The β -sandwich motif consists of two antiparallel β sheets, which are curled together to form a cleft, in which the oligosaccharide ligand is bound. This fold, seen in Figure 1.3, has been seen in the structures of galectin-1 with branched biantennary glycans³⁷, and LacNAc³⁸.

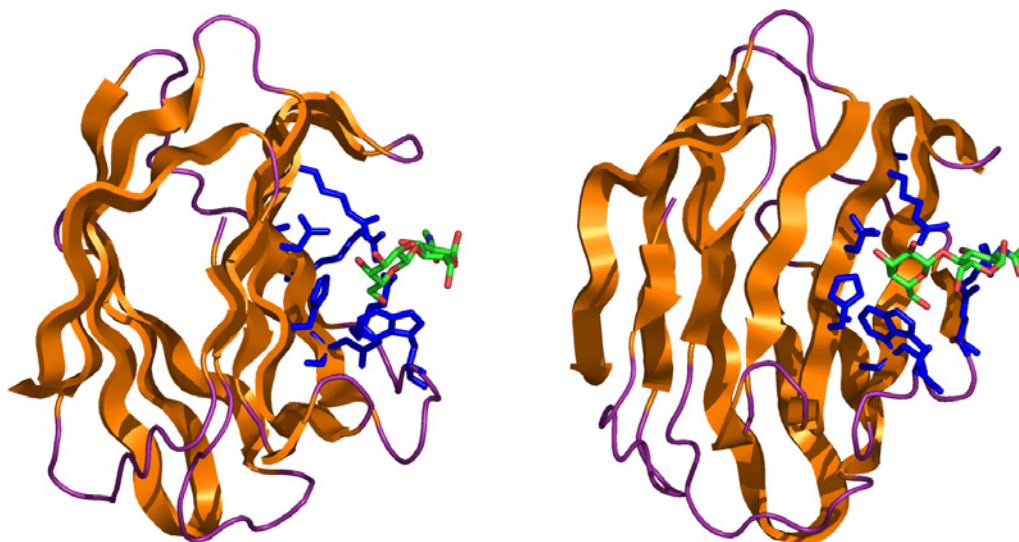


Figure 1.3. Two views of the β -sandwich structure of the galectin-1 monomer. Conserved residues which participate in ligand interactions are shown in blue, and the LacNAc disaccharide is shown in green.

The dimer interface of galectin-1 contains both the N- and C-termini, as well as an axis of pseudo C2S symmetry, as shown in Figure 1.4.

In the crystal structure of galectin-1 with LacNAc, solved by Liao et al.³⁸, a number of important protein – ligand interactions were noted. As expected, there are several interactions between residues of the galectin-1 binding cleft and the galactose of LacNAc. The O4 hydroxyl forms hydrogen bonds with the side chains of H44, N46 and R48, and these interactions are what makes galectin-1 (and the galectin family) specific for galactose. There are also hydrogen bonds between the galactose O6, and the side chains of N61 and D71. The aromatic side chain of W68 stacks with the back

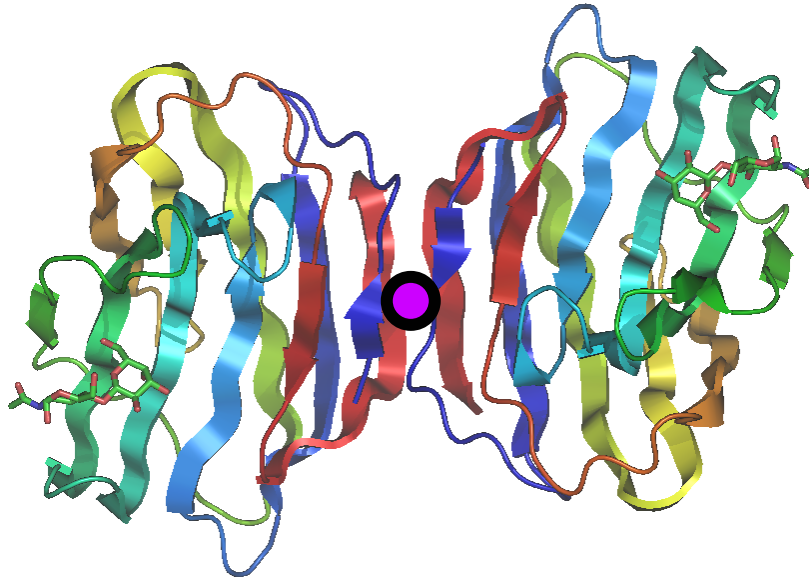


Figure 1.4. The galectin-1 dimer. The N-termini (blue) and C-termini (red) are both located at the dimer interface. The axis of symmetry is shown in purple.

face of the galactose ring, and this interaction between sugar rings and aromatic amino acids is a feature seen in many lectins³⁹ and antibodies⁴⁰. In a similar manner, the O3 of GlcNAc interacts with R48 and D71, while the N-acetyl group hydrogen bonds with R73. These protein residues are conserved within the galectin-family, and this conservation is highlighted in Figure 1.5.

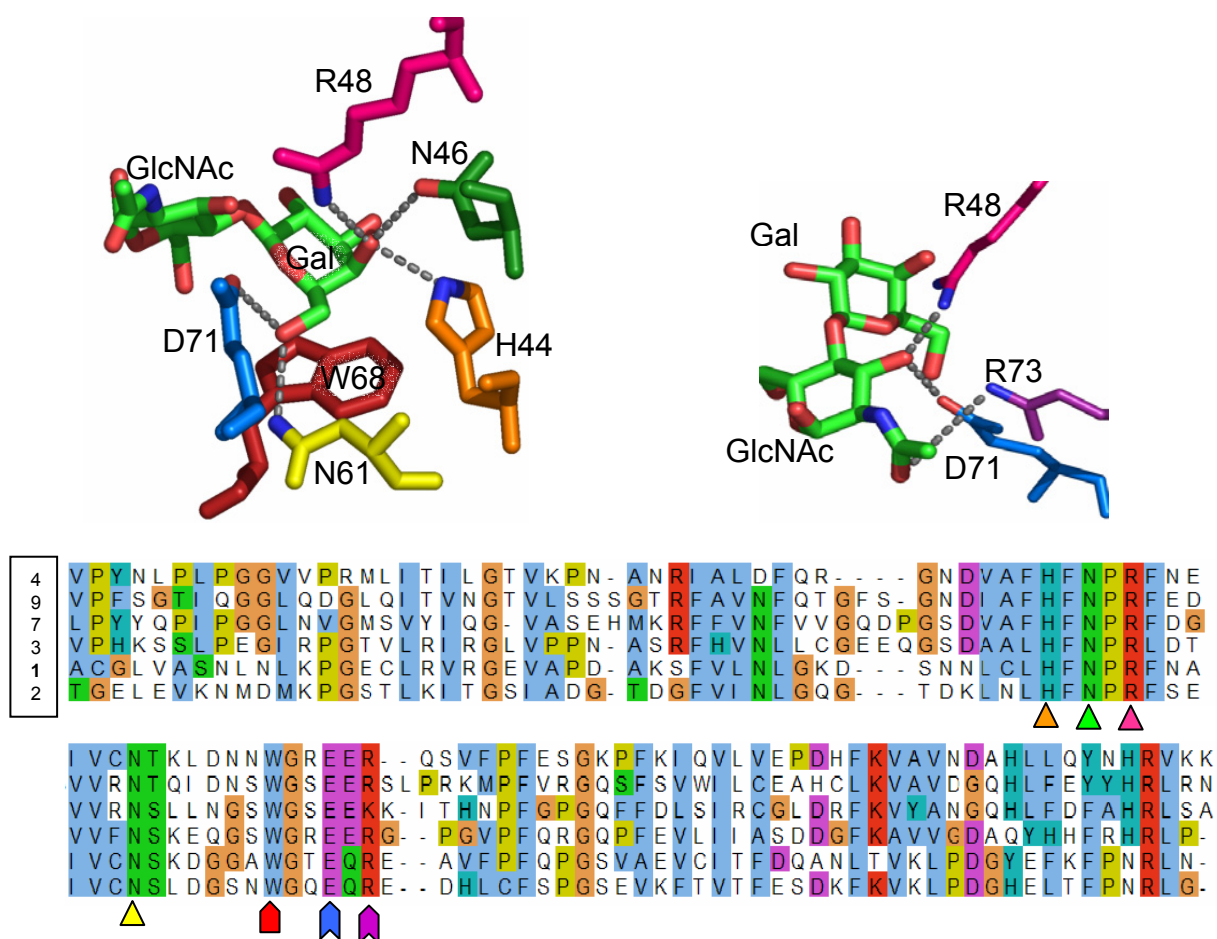


Figure 1.5. Structural recognition of LacNAc by the galectin-1 CRD. Hydrogen bonds are shown as dashed lines. Human galectin sequences showing amino acid sequence conservation of residues involved in LacNAc binding are highlighted with pointers color-coded to match the structural representation. The number of the galectin sequence is shown in the box on the upper left corner.

Galectin domain organization

Galectins can be further categorized based upon the structural organization of the protein domains. A prototype galectin consists of two identical CRDs that associate spontaneously to form dimers in solution. A tandem-repeat galectin contains two CRDs on a single polypeptide chain. Galectin-3, termed a chimera-type galectin, contains a C-terminal CRD fused to an N-terminus containing a repeating proline-rich domain,

which binds RNA, and is responsible for self-association into dimers when bound to ligand⁴¹. The representative structures of galectin folds are shown in Figure 1.6

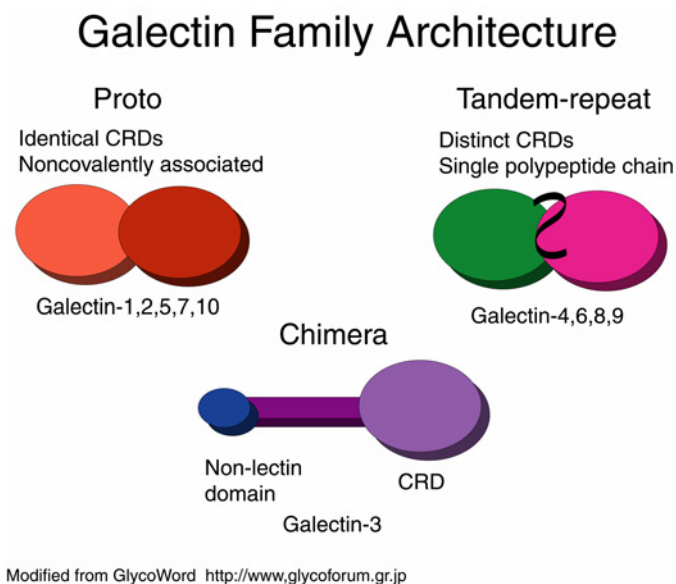


Figure 1.6. Domain architecture of the galectin family.

Proposed extended binding site of galectin-1

In addition to revealing overall protein topology and defining several critical protein – carbohydrate interactions, the X-ray structure of the galectin-1 – LacNAc complex also gives some clues as to how the CRD might accommodate long oligosaccharide chains. Several observations showed the ability of galectin-1 to bind oligosaccharides much larger than the LacNAc disaccharide, such as poly-LacNAc chains from laminin⁴² and various *N*-glycans⁴³. The LacNAc disaccharide binds at the entrance of a cleft which extends the length of the galectin-1 CRD, as seen in Figure 1.7. The extended binding site includes residues 3-5, 28-33, 38 and 119-125 of the bovine galectin-1 sequence.

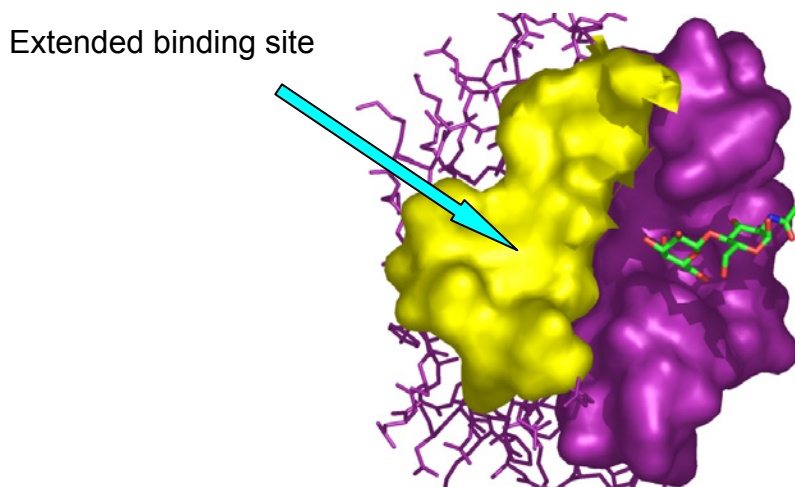


Figure 1.7. The galectin-1 CRD showing the binding groove and proposed extended binding site which accommodates 3'O-substituents of ligands such as Neu5Ac-LacNAc. The primary binding site is purple, and the proposed extended binding site is yellow.

Role of cysteines

Whether the surface cysteines of the galectins play a role in the binding of carbohydrate ligands remains unclear. Initial studies on galectins indicated a requirement for cysteines in the reduced state for optimal binding, an observation which led to their classification as S-type lectins. However, replacement of cysteine residues

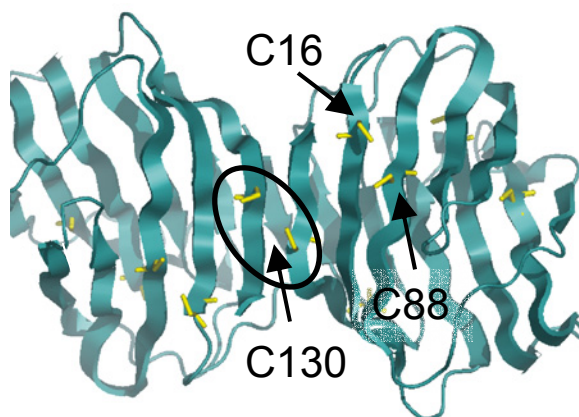


Figure 1.8. View of the galectin-1 dimer highlighting possible sites of disulfide bond formation. The C130 pair spans the dimer interface, while the C16/C88 pair is located on adjacent beta strands.

with serines indicates that only a few critical residues are necessary for activity^{44,45}. It is likely that disulfide bond formation between adjacent cysteines results in a change in structure, leading to a disruption of optimal binding geometry. It has even been suggested that the oxidized form of galectin-1 may perform an entirely different role in nerve regeneration which is completely independent of any carbohydrate binding activity^{46,47}. The crystal structure of galectins-1 shows possible cysteine crosslinking pairs, illustrated in Figure 1.8. The crosslinking of the C16/C88 pair would disrupt the beta sheet structure, while a disulfide formed from C130 residues of individual monomers would result in a covalently crosslinked dimer.

Functions of mammalian galectins

The mammalian galectins have been numbered galectin-1 through -14, and have been structurally and functionally characterized to differing degrees. Galectin-1 and galectin-3 have been the most intensely studied of the group, and consequently are the best characterized members of this lectin family. The classification scheme based on their date of discovery masks a number of characteristics that complicate the study of these proteins. Murine galectin-4 and -6 share 80% identity for the mouse, and raise cross-reactive antibodies. Similarly, galectin-5 and galectin-9 are 70% identical for the human and 85% identical for the mouse and galectin-13 is 54% identical to Charcot-Leyden crystal protein (galectin-10/CLCP). There are a number of proteins related to galectin-10/CLCP, such as GRIFIN, galectin-related interfiber protein. These complications, in addition to overlapping binding specificities and probable functional redundancy contribute to difficulties in mapping functions to individual galectins.

Table 1.3. Proposed functions for mammalian galectins

Galectin	Localization	Proposed Functions	Structure
galectin-1	Developmentally regulated expression in all major tissues ^{48,49}	Cell adhesion ^{50,51} , regulation of development ⁵²⁻⁵⁴ , apoptosis of T-cells ⁵⁵ , pre-mRNA splicing ⁵⁶	X-ray structures: LacNAc ³⁸ , lactose ⁵⁷ complex glycan ³⁷
galectin-2	Gastric epithelium, muscle ⁵⁸		X-ray structure: bound lactose ⁵⁹
galectin-3	Fibroblasts, gastric epithelium, dendritic cells	Pre-mRNA splicing ⁶⁰ , cell adhesion	X-ray structure CRD ⁶¹
galectin-4	Epithelium of colon, small intestine ⁶²	Mucosal immunity of GI	
galectin-5	Erythrocytes, hematopoietic cells ⁶³	Erythrocyte maturation	
galectin-6	Gastrointestinal epithelium ⁵²	Mucosal immunity of GI	
galectin-7	Epithelial tissue ⁶⁴		X-ray structure: Gal, GalNAc, Lactose, LacNAc ⁶⁵
galectin-8	Liver, heart, muscle, kidney and brain ⁶⁶		
galectin-9 ecalectin	Eosinophils, lymphoid tissue ⁵⁸	Eosinophil attraction and activation	
galectin-10/CLCP	Basophils, eosinophils ⁶⁷		X-ray structure ⁶⁸
galectin-11	Gastrointestinal epithelium ⁶⁹	Mucosal immunity of GI tract	
galectin-12	Adipose tissue	Apoptosis of adipocytes ⁷⁰ , cell cycle regulation	
galectin-13 placental protein 13	Placenta and fetal tissues ⁷¹	Placental development, implantation	Homology model ⁷²
galectin-14	eosinophils	Inflammatory response ⁷³	

It is clear from Table 1.3 that the diverse functions linked to members of the galectin family do share several recurring themes. The numerous links to immune function (galectin-1, -4, -6, -9, -10, -11, -14) indicate that galectins are important for several aspects of the immune response. The role of galectin-1 in T-cell apoptosis is thought to regulate thymic selection of T-lymphocytes, a key process in self recognition. The presence of galectins in leukocytes (basophils and eosinophils) points toward a direct role in the inflammatory response, which remains to be defined. Their presence in the gastric epithelium has been proposed as a component of the mucosal immune surveillance of the gastrointestinal tract.

Another major role for galectins is in the regulation of development. Galectin-13, first characterized as an important placental protein, is involved in embryonic implantation. While the glycosyl ligands for galectin-13 remain to be determined, it is believed that galectin-13 can bind to basement membrane glycoproteins such as laminin, in a manner similar to galectin-1⁷². Galectin-1 and -3 are thought to be important in cell-cell adhesions formed during the migration of particular cell types during the developmental process, as illustrated by the altered migration patterns of primary olfactory neurons⁵³ seen in mice deficient in galectin-1 and galectin-3⁷⁴. The levels of expression of galectin-1 and -3 are regulated in a manner that is specific to certain tissues or developmental stages⁷⁵.

A wide variety of *in vitro* effects has been ascribed to galectin-1. Galectin-1 has been shown to both promote^{50,51} and inhibit⁷⁶ the adhesion of cells to the extracellular matrix. These interactions occur through contacts between these galectins and proteins of the basement membrane, such as laminin and fibronectin. Related to these

observations is the evidence that galectin-1 and -3 are both factors regulating mitogenesis. Studies of the effects of exogenous galectin-1 show conflicting evidence of both promotion^{77,78} and inhibition⁷⁹ of cell growth, depending upon the system under examination. Moreover, in one human fibroblast cell line, galectin-1 has been shown to be both mitogenic and inhibitory of cell growth, depending upon concentration⁸⁰.

Galectin-1 and -7 have been shown to be factors in the promotion of apoptosis. Galectin-1 has been shown to induce apoptosis of CD45+ B-cells, a possible mechanism for clonal selection⁸¹ critical to the regulation of self/non-self recognition in the immune system⁵⁵. Galectin-7 is induced by p53 and promotes apoptosis through the JNK signaling pathway⁸². Conversely, galectin-3 has been shown to provide protection from apoptotic signals⁸³. Galectin-1 and -3 have also been shown to play roles in the splicing of pre-mRNA^{84,85}. The activity of these proteins as splicing factors is dependent on the presence of the fully active CRD, and galectins-1 and -3 appear to function in a redundant manner.

As in studies on adhesion and cellular proliferation, there are a number of contradictory reports on the links between galectin expression and malignant transformation. In general, galectin-1 is found in increased levels in tumor cells, and is correlated with changes in cell morphology^{86,87}. Conversely, in tumors of the breast and thyroid levels of galectin-3 are decreased relative to normal tissue^{88,89}. In addition to the conflicting reports coming out of these studies, the overlap in function that exists between members of the galectin family make direct correlations between protein levels and transformation difficult to establish.

Functional substitution of galectin ligands

The 3' hydroxyl of the galactose is the site of several biologically important substitutions. In poly-LacNAc chains, such as those seen in laminin, the 3' position of galactose is the linkage to the next LacNAc disaccharide. The majority of the termini of N-glycan branches contain a link to an anionic saccharide, sialic acid. Sialic acid can be either α -(2,3) or α -(2,6) linked to the galactose residue of LacNAc and only Neu5Ac- α -(2,3)-LacNAc is capable of binding galectin-1. The addition of a sulfate at the 3' position is another LacNAc modification seen in certain tissue types. A representation of several relevant substitutions of galactose is shown in Figure 1.9.

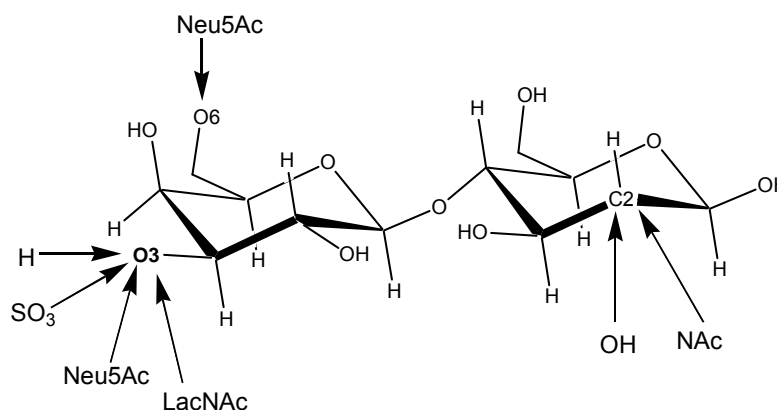


Figure 1.9. Schematic diagram of Lac(NAc), showing sites of biologically relevant chemical substitution.

Examination of the effects of 3'-O-substituents on galectin-1 binding

The 3'-O- substituents of LacNAc affect the affinity for galectin-1, the details of which will be discussed in the following chapters. Of particular interest is the description of the protein – carbohydrate interactions observed in galectin-1 – ligand complexes. While there is a structural model for the galectin-1 – LacNAc complex, (Liao et al 1994) there is currently no information on how 3'-O- substituted ligands would be accommodated within the galectin-1 CRD. By performing molecular dynamics

simulation on galectin-1 ligands, we have been able to predict interactions between 3'-O- substituted ligands and the galectin-1 binding groove, as well as obtain information on the energetic contributors to ligand binding.

The simulation of non-covalently linked protein-carbohydrate complexes remains a field with many open questions. While there have been a number of studies on the structural interactions between N-linked glycans with their attached proteins, the body of work examining lectin – carbohydrate interactions is much more sparse. Therefore it has become necessary to perform simulations under a variety of conditions to obtain trajectories that reproduce the structural and energetic properties of *in vitro* complexes. Similarly, NMR analysis of galectin-1 complexes has required a number of sophisticated labeling techniques and the development of novel strategies for assignment, as well as the use of established methods. Out of this analysis emerges a tool for the assignment of binding site residues in structurally classified proteins. In addition to these contributions to the theoretical consideration of these complexes, there are several important biological ramifications of this work.

References

1. Mitchell, S. and Reichert, E., *Researches Upon the Venoms of Poisonous Serpents*. Contributions to Knowledge. Vol. XII. 1886: Smithson.
2. Stillmark, H., *Über Ricin, ein giftiges Ferment aus dem Samen von Ricinus communis L. und einigen anderen Euphorbiaceen*, in *Schnakenburg's Buchdruckerei*. 1888: Dorpat.
3. Watkins, W. and Morgan, W., *The Structural Basis for Specificity in Human ABO(H) Blood Groups*. *Nature*, 1952. 169: p. 825.
4. Boyd, W., *The Proteins of Immune Reactions*, in *The Proteins*, H. Neurath and K. Bailey, Editors. 1954, Academic Press: New York. p. 756 - 844.
5. Flexner, S. and Noguchi, H., *Snake Venom in Relation to Haemolysis, Bacteriolysis and Toxicity*. *Journal of Experimental Medicine*, 1902. 6: p. 13837 - 13843.
6. Weimar, T., *Manuscript in preparation*.
7. Stockert, R., Morell, A., and Scheinberg, I., *Mammalian Hepatic Lectin*. *Science*, 1974. 186: p. 365 - 366.
8. Teichberg, V., Silman, I., Beitsch, D., and Resheff, G., *A β -D-Galactoside Binding Protein from Electric Organ Tissue of Electrophorus electricus*. *Proceedings of the National Academy of Sciences USA*, 1975. 72: p. 1383 - 1387.
9. Kilpatrick, D., *Animal Lectins: A Historical Introduction and Overview*. *Biochemica and Biophysica Acta: General Subjects*, 2002. 1572: p. 187 - 197.
10. Rudiger, H., Siebert, H., Solis, D., Jimenez-Barbero, J., Romero, A., von der Leith, C., Diaz-Maurino, T., and Gabius, H., *Medicinal Chemistry Based on the Sugar Code: Fundamentals of Lectinology and Experimental Strategies with Lectins as Targets*. *Current Medicinal Chemistry*, 2000. 7: p. 389 - 416.
11. Cooper, D., *Galectinomics: Finding Themes in Complexity*. *Biochemica and Biophysica Acta: General Subjects*, 2002. 1572: p. 209 - 231.
12. Angata, T. and Brinkman-Van der Linden, E., *I-type Lectins*. *Biochemica and Biophysica Acta: General Subjects*, 2002. 1572: p. 294 - 316.
13. Dahms, N. and Hancock, M., *P-type Lectins*. *Biochimica et Biophysica Acta, General Subjects*, 2002. 1572: p. 317 - 340.
14. Lasky, L., *Selectins: Interpreters of Cell-Specific Carbohydrate Information During Inflammation*. *Science*, 1992. 258: p. 964 - 969.
15. Ashwell, G. and Morell, A., *The Role of Surface Carbohydrates in the Hepatic Recognition and Transport of Circulating Glycoproteins*. *Advances in Enzymology*, 1974. 41: p. 99 - 128.
16. Breitfeld, P., Simmons, C., Strous, G., Geuze, H., and Schwartz, A., *Cell Biology of the ASGP-R System: A Model of Receptor-Mediated Endocytosis*. *International Reviews in Cytology*, 1985. 97: p. 47 - 95.
17. Trombetta, E. and Helenius, A., *Lectins as Chaperones in Glycoprotein Folding*. *Current Opinion in Structural Biology*, 1998. 8(5): p. 587 - 592.
18. Parodi, A., *Protein Glycosylation and its Role in Protein Folding*. *Annual Review of Biochemistry*, 2000. 69: p. 69 - 93.

19. Loris, R., *Principes of Structures of Animal and Plant Lectins*. Biochemica and Biophysica Acta: General Subjects, 2002. 1572: p. 198 - 208.
20. Ahmed, H., *Galectins: Conservation of Functionally and Structurally Relevant Amino Acid Residues Defines Two Types of Carbohydrate Recognition Domains*. Glycobiology, 1994. 4(5): p. 545 - 549.
21. Lobsanov, Y.D., Yuri, D., and Rini, J.M., *Galectin Structure*. Trends in Glycoscience and Glycotechnology, 1997. 45(Jan.): p. 145-154.
22. Cooper, D. and Barondes, S., *God Must Love Galectins; He Made So Many of Them*. Glycobiology, 1999. 9(10): p. 979 - 984.
23. Swaminathan, D., Leonidas, M., Savage, M., Ackerman, S., and Acharya, K., *Selective Recognition of Mannose by the Human Eosinophil Charcot-Leyden Crystal Protein (Galectin 10): A Crystallographic Study at 1.8 Å Resolution*. Biochemistry, 1999. 38: p. 13837 - 13843.
24. Cooper, D. and Barondes, S., *Evidence for Export of a Muscle Lectin from Cytosol to Extracellular Matrix and for a Novel Secretory Mechanism*. Journal of Cell Science, 1990. 110: p. 1681 - 1691.
25. Zhou, Q. and Cummings, R.D., *L-14 Lectin Recognition of Laminin and its Promotion of In Vitro Cell Adhesion*. Archives of Biochemistry and Biophysics, 1993. 300(1): p. 6-17.
26. Gu, M., Wang, W., Song, W., Cooper, D., and Kaufman, S., *Selective Modulation of the Interaction of $\alpha_7\beta_1$ Integrin with Fibronectin and Laminin by L-14 Lectin During Skeletal Muscle Differentiation*. Journal of Cell Science, 1994. 107: p. 175-181.
27. Pace, K., Lee, C., Steward, P., and Baum, L., *Restricted Receptor Segregation into Membrane Microdomains Occurs on Human T Cells During Apoptosis Induced by Galectin-1*. Journal of Immunology, 1999. 163: p. 3801 - 3811.
28. Fouillit, M., Joubert-Caron, R., Poirier, F., Bourin, P., Monostori, E., Levi-s Strauss, M., Raphael, M., Bladier, D., and Caron, M., *Regulation of CD45-Induced Signaling by Galectin-1 in Burkitt Lymphoma B Cells*. Glycobiology, 2000. 10(4): p. 413 - 419.
29. Sato, T., Furukawa, K., Autero, M., Gahmberg, C.G., and Kobata, A., *Structural Study of the Sugar Chains of Human Leukocyte Common Antigen CD45*. Biochemistry, 1993. 32: p. 12694 - 12704.
30. Brewer, C., *Binding and Cross-Linking Properties of Galectins*. Biochimica et Biophysica Acta, General Subjects, 2002. 1572: p. 255 - 262.
31. Galvan, M., Tsuboi, S., Fukuda, M., and Baum, L.G., *Expression of a Specific Glycosyltransferase Enzyme Regulates T Cell Death Mediated by Galectin-1*. J. Biol. Chem., 2000. 275(22): p. 16730-16737.
32. Amano, M., Galvan, M., He, J., and Baum, L.G., *The ST6Gal I Sialyltransferase Selectively Modifies N-Glycans on CD45 to Negatively Regulate Galectin-1-induced CD45 Clustering, Phosphatase Modulation, and T Cell Death*. J. Biol. Chem., 2003. 278(9): p. 7469-7475.
33. Weis, W. and Drickamer, K., *Trimeric Structure of a C-type Mannose-Binding Protein*. Structure, 1994. 2: p. 1227 - 1240.

34. Rini, J., *X-ray Crystal Structures of Animal Lectins*. Current Opinion in Structural Biology, 1995. 5(3331): p. 617-621.
35. Loris, R., Hamelryck, T., Bouckaert, J., and Wyns, L., *Legume Lectin Structure*. Biochimica et Biophysica Acta, 1998. 1383: p. 9 - 36.
36. Keitel, U., Simon, O., Borriss, R., and Heinemann, U., *Molecular and Active Site Structure of a Bacillus 1,3-1,4- β -Glucanase*. Proceedings of the National Academy of Sciences USA, 1993. 90: p. 5287 - 5291.
37. Bourne, Y., Bolgiano, B., Liao, D., Strecker, G., Cantau, P., Herzberg, O., Feizi, T., and Cambillau, C., *Crosslinking of Mammalian Lectin (Galectin-1) by Complex Biantennary Saccharides*. Structural Biology, 1994. 1(123331): p. 863-870.
38. Liao, D., Kapadia, G., Ahmed, H., Vasta, G., and Herzberg, O., *Structure of S-lectin, a Developmentally Regulated Vertebrate β -Galactoside-Binding Protein*. Proceedings of the National Academy of Sciences USA, 1994. 91(Feb.): p. 1428-1432.
39. Taroni, C., Jones, S., and J., T., *Analysis and Prediction of Carbohydrate Binding Sites*. Protein Engineering, 2000. 13(2): p. 89 - 98.
40. Bundle, D., Baumann, H., Brisson, J., Gagne, S., Zdanov, A., and Cygler, M., *Solution Structure of a Trisaccharide-Antibody Complex: Comparison of NMR Measurements with a Crystal Structure*. Biochemistry, 1994. 33(17): p. 5183 - 5192.
41. Wang, L., Inohara, H., Pienta, K., and Raz, A., *Galectin-3 is a Nuclear Matrix Protein Which Binds RNA*. Biochemical and Biophysical Research Communications, 1995. 217(1): p. 292 - 303.
42. Merkle, R. and Cummings, R., *Asparagine-Linked Oligosaccharides Containing Poly-N-acetylactosamine Chains are Preferentially Bound by Immobilized Calf Heart Agglutinin*. Journal of Biological Chemistry, 1988. 263(31): p. 16143-16149.
43. Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Muller, W., Yagi, F., and Kasai, K., *Oligosaccharide Specificity of Galectins: A Search by Frontal Affinity Chromatography*. Biochimica et Biophysica Acta, General Subjects, 2002. 1572: p. 232 - 254.
44. Hirabayashi, J. and Kasai, K., *Effect of Amino Acid Substitution by Site-Directed Mutagenesis on the Carbohydrate Recognition and Stability of Human 14-kDa β -Galactoside-Binding Lectin*. Journal of Biological Chemistry, 1991. 266(35): p. 23648-23653.
45. Hirabayashi, J. and Kasai, K.-i., *Further Evidence by Site-Directed Mutagenesis that Conserved Hydrophilic Residues Form a Carbohydrate-Binding of Galectin-1*. Glycoconjugate Journal, 1994. 11: p. 437 - 442.
46. Inagaki, Y., Sohma, Y., Horie, H., Nozawa, R., and Kadoya, T., *Oxidized Galectin-1 Promotes Axonal Regeneration in Peripheral Nerves but Does Not Posses Lectin Properties*. European Journal of Biochemistry, 2000. 267: p. 2955 - 2964.

47. Horie, H. and Kadoya, T., *Identification of Oxidized Galectin-1 as an Initial Repair Regulatory Factory After Axotomy in Peripheral Nerves*. Neuroscience Research, 2000. 38: p. 131 - 137.
48. Colnot, C., Ripoche, M., Fowles, D., Cannon, V., Sacaerou, F., Cooper, D., and F., P., *The Role of Galectins in Mouse Development*. Trends in Glycoscience and Glycotechnology, 1997. 9(45): p. 31 - 40.
49. Cooper, D., *Galectin-1: Secretion and Modulation of Cell Interactions with Laminin*. Trends in Glycoscience and Glycotechnology, 1997. 9(45): p. 57 - 67.
50. Tanzer, M., Chandrasekaran, S., Dean, J., and Giniger, M., *Role of Laminin Carbohydrates on Cellular Interactions*. Kidney International, 1993. 43: p. 66 - 72.
51. Puche, A. and Key, B., *Identification of Cells Expressing Galectin-1, a Galactose-Binding Receptor, in the Rat Olfactory System*. Journal of Comparative Neurology, 1995. 357: p. 513 - 523.
52. Colnot, C., *The Role of Galectins in Mouse Development*. Trends in Glycoscience and Glycotechnology, 1997. 9(45): p. 31 - 40.
53. Puche, A., Poirier, F., Hair, M., Bartlett, P., and Key, B., *Role of Galectin-1 in the Developing Mouse Olfactory System*. Developmental Biology, 1996. 179: p. 274 - 287.
54. Nowak, T., Haywood, O., and Barondes, S., *Developmentally Regulated Lectin in Embryonic Chick Muscle and a Myogenic Cell Line*. Biochemical and Biophysical Research Communications, 1976. 69: p. 650 - 657.
55. Perillo, N., Pace, K., Seilhamer, J., and Baum, L., *Apoptosis of T cells Mediated by Galectin-1*. Nature, 1995. 378(Dec. 14): p. 736 - 738.
56. Patterson, R., Dagher, S., Vyakarnam, A., and Wang, J., *Nuclear Galectins: Functionally Redundant Components in Processing of pre-mRNA*. Trends in Glycoscience and Glycotechnology, 1997. 9(45): p. 77 - 85.
57. Lobsanov, Y., Gitt, M., Leffler, H., Barondes, S., and Rini, J., *X-ray Crystal Structure of the Human Dimeric S-Lac Lectin, L-14-II, in Complex with Lactose at 2.9 Å Resolution*. Journal of Biological Chemistry, 1993. 268(363331): p. 27034-27038.
58. Gitt, M., Jordan, E., and Leffler, H., *Galectin-2, Galectins-5 and -9, and Galectins-4 and -6*. Trends in Glycoscience and Glycotechnology, 1997. 9(45): p. 87 - 93.
59. Lobsanov, Y., Gitt, M., Leffler, H., Barondes, S., and Rini, J., *X-ray Crystal Structure of the Human Dimeric S-Lac Lectin, L-14-II, in Complex with Lactose at 2.9-Å Resolution*. Journal of Biological Chemistry, 1993. 268(36): p. 27034 - 27038.
60. Dagher, S., Wang, J., and Patterson, R., *Identification of Galectin-3 as a Factor in pre-mRNA Splicing*. Proceedings of the National Academy of Sciences USA, 1995. 92(4): p. 1213 - 1217.
61. Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S., and Rini, J.M., *X-ray Crystal Structure of the Human Galectin-3 Carbohydrate Recognition Domain at 2.1 Å Resolution*. Journal of Biological Chemistry, 1998. 273(21): p. 13047-13052.

62. Oda, Y., Herrmann, J., Gitt, M., Turck, C., Burlingame, A., Barondes, S., and Leffler, H., *Soluble Lactose-Binding Lectin from Rat Intestine with Two Different Carbohydrate-Binding Domains in the Same Peptide Chain*. Journal of Biological Chemistry, 1993. 268(8): p. 5929 - 5939.
63. Gitt, M., Wiser, M., Leffler, H., Herrmann, J., Xia, Y., Massa, S., Cooper, D., Lusic, A., and Barondes, S., *Sequence and Mapping of Galectin-5, a beta-Galactoside-Binding Lectin, Found in Rat Erythrocytes*. Journal of Biological Chemistry, 1995. 270(10): p. 5032-5038.
64. Magnaldo, T., Bernerd, F., and Darmon, M., *Galectin-7, a Human 14-kDa S-Lectin, Specifically Expressed in Keratinocytes and Sensitive to Retinoic Acid*. Developmental Biology, 1995. 168: p. 259 - 271.
65. Leonidas, D., Vatzaki, E., Vorum, H., Celis, J., Madsen, P., and Acharya, K., *Structural Basis for the Recognition of Carbohydrates by Human Galectin-7*. Biochemistry, 1998. 37(40): p. 13930 - 13940.
66. Hadari, Y., Paz, K., Dekel, R., Mestrovic, T., Accili, D., and Zick, Y., *Galectin-8, a New Rat Lectin, Related to Galectin-4*. Journal of Biological Chemistry, 1995. 270(7): p. 3447 - 3453.
67. Ackerman, S., Corrette, S., Rosenberg, H., Bennett, J., Mastrianni, D., Nicholson-Weller, A., Weller, P., Chin, D., and Tenen, D., *Molecular Cloning and Characterization of Human Eosinophil Charcot-Leyden Crystal Protein (Lysophospholipase). Similarities to IgE Binding Proteins and the S-type Animal Lectin Superfamily*. Journal of Immunology, 1993. 150(2): p. 456 - 468.
68. Leonidas, D., Elbert, B., Zhou, Z., Leffler, H., Ackerman, S., and Acharya, K., *Crystal structure of Human Charcot-Leyden Crystal Protein, an Eosinophil Lysophospholipase, Identifies it as a New Member of the Carbohydrate-Binding Family of Galectins*. Structure, 1995. 3: p. 1379 - 1393.
69. Dunphy, J., Balic, A., Barcham, G., Horvath, A., Nash, A., and Meeusen, E., *Isolation and Characterization of a Novel Inducible Mammalian Galectin*. Journal of Biological Chemistry, 2000. 275(41): p. 32106 - 32113.
70. Hotta, K., Funahashi, T., Matsukawa, Y., Takahashi, M., Nishizawa, H., Kishida, K., Matsuda, M., Kuriyama, H., Kihara, S., Nakamura, T., Tochino, Y., Bodkin, N., Hansen, B., and Matsuzawa, Y., *Galectin-12, an Adipose-Expressed Galectin-Like Molecule Possessing Apoptosis-Inducing Activity*. Journal of Biological Chemistry, 2001. 276(36): p. 34089 - 34097.
71. Than, N.G., Sumegi, B., Than, G.N., Berente, Z., and Bohn, H., *Isolation and Sequence Analysis of a cDNA Encoding Human Placental Tissue Protein 13 (PP13), a New Lysophospholipase, Homologue of Human Eosinophil Charcot-Leyden Crystal Protein*. Placenta, 1999. 20(8): p. 703-710.
72. Visegrady, B., Than, N., Kilar, F., Sumegi, B., Than, G., and Bohn, H., *Homology Modeling and Molecular Dynamics Studies of Human Placental Tissue Protein 13 (Galectin-13)*. Protein Engineering, 2001. 14(11): p. 875 - 880.
73. Dunphy, J., Barcham, G., Bischof, R., Young, A., Nash, A., and Meeusen, E., *Isolation and Characterization of a Novel Eosinophil-specific Galectin*

- Released into the Lungs in Response to Allergen Challenge*. Journal of Biological Chemistry, 2002. 277(17): p. 14916 - 14924.
74. Poirier, F. and Robertson, E., *Normal Development of Mice Carrying a Null Mutation in the Gene Encoding the L14 S-type Lectin*. Development, 1993. 119: p. 1229 - 1236.
 75. Liu, F., Patterson, R., and Wang, J., *Intracellular Functions of Galectins*. Biochimica et Biophysica Acta, General Subjects, 2002. 1572: p. 263 - 273.
 76. Cooper, D., Massa, S., and Barondes, S., *Endogenous Muscle Lectin Inhibits Myoblast Adhesion to Laminin*. Journal of Cell Biology, 1991. 115: p. 1437 - 1448.
 77. Stanford, G. and Harris-Hooker, S., *Stimulation of Vascular Cell Proliferation by Beta-Galactoside Specific Lectins*. FASEB Journal, 1990. 4: p. 2912 - 2918.
 78. Moiseeva, E., Javed, Q., Spring, E., and de Bono, D., *Galectin-1 is Involved in Vascular Smooth Muscle Cell Proliferation*. Cardiovascular Research, 2000. 45: p. 493 - 502.
 79. Blaser, C., Kaufmann, M., Muller, C., Zimmermann, C., Wells, V., Mallucci, L., and Pircher, H., *β -Galactoside-Binding Protein Secreted by Activated T Cells Inhibits Antigen-Induced Proliferation of T Cells*. European Journal of Immunology, 1998. 28: p. 2311 - 2319.
 80. Adams, L., Scott, G., and Weinberg, C., *Biphasic Modulation of Cell Growth by Recombinant Human Galectin-1*. Biochimica et Biophysica Acta, 1996. 1312: p. 137 - 144.
 81. Perillo, N., Uittenbogaart, C., Nguyen, J., and Baum, L., *Galectin-1, and Endogenous Ligand Produced by Thymic Epithelial Cells, Induces Apoptosis of Human Thymocytes*. Journal of Experimental Medicine, 1997. 185: p. 1851 - 1858.
 82. Kuwabara, I., Kuwabara, Y., Yang, R., Schuler, M., Green, D., Hsu, D., and Liu, F., *Galectin-7 (PIG1) Exhibits Pro-apoptotic Function through JNK Activation and Mitochondrial Cytochrome c Release*. Journal of Biological Chemistry, 2002. 277.
 83. Yang, R., Hsu, K., and Liu, F., *Expression of Galectin-3 Modulates T-cell Growth and Apoptosis*. Proceedings of the National Academy of Sciences USA, 1996. 93: p. 6737 - 6742.
 84. Dagher, S., Wang, J., and Patterson, R., *Identification of Galectin-3 as a Factor in pre-mRNA Splicing*. Proceedings of the National Academy of Sciences USA, 1995. 92: p. 1213 - 1217.
 85. Vyakarnam, A., Dagher, S., Wang, J., and Patterson, R., *Evidence for a Role for Galectin-1 in pre-mRNA Splicing*. Molecular and Cell Biology, 1997. 17: p. 4730 - 4737.
 86. Yamaoka, K., Mishima, K., Nagashima, Y., Asai, A., Sanai, Y., and Kirino, T., *Expression of Galectin-1 mRNA Correlates With the Malignant Potential of Human Gliomas and Expression of Antisense Galectin-1 Inhibits the Growth of 9 Glioma Cells*. Journal of Neuroscience Research, 2000. 59: p. 722 - 730.

87. Chirariotti, L., Belingieri, M., Battaglia, C., Benevenuto, G., Martelli, M., Salvatore, P., Chiappetta, G., Bruni, C., and Fusco, A., *Expression of Galectin-1 in Normal Human Thyroid Gland and in Differentiated and Poorly Differentiated Thyroid Tumors*. 1995.
88. Xu, X., el-Naggar, A., and Lotan, R., *Differential Expression of Galectin-1 and Galectin-3 in Thyroid Tumors. Potential Diagnostic Implications*. American Journal of Pathology, 1995. 147: p. 815 - 822.
89. Ohannesian, D., Lotan, D., Thomas, P., Jessup, J., Fukada, M., Gabius, H., and Lotan, R., *Carcinoembryonic Antigen and Other Glycoconjugates Act as Ligands for Galectin-3 in Human Colon Carcinoma Cells*. Cancer Research, 1995. 55: p. 2191 - 2199.

Chapter 2

Molecular Dynamics Simulation of galectin-1 – ligand Complexes

Introduction

In order to obtain atomic-level information on galectin-1 – ligand complexes for which there were no structural models, we performed molecular dynamics simulations on a series of docked complexes using the AMBER molecular dynamics suite. The galectin-1 – LacNAc complex served both as the starting point for docking of galectin-1 ligands, and as a benchmark for the performance of our simulational protocols. Starting with this galectin-1 LacNAc complex, we tested a variety of models for treatment of solvation and electrostatic forces. Having determined a viable set of simulation conditions, we ran experiments on docked complexes of 3'-O- substituted ligands bound to the galectin-1 CRD. Analysis of the resulting trajectories gave information on the protein – ligand interactions important in these complexes as well as the energetic contributions to the binding free energy. This analysis allows us to provide a structural and energetic rationale for observed differences in binding affinity, and to develop a model for the binding of galectin-1 to 3'-O-substituted ligands.

Molecular Dynamics

Molecular dynamics involves the sequential solution of a set of equations, termed the force field, which describe the forces on a set of atoms. Using an initial set of molecular coordinates, the total potential energy is described by a general equation which consists of several energetic terms. This general definition of a force field

describes the potential energy of a group of atoms as the sum of a number of geometry dependent energetic contributions as seen in Equation 2.1.

$$E_{\text{pot}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsions}} + E_{\text{vdW}} + E_{\text{electrostatic}} \quad \text{Equation 2.1}$$

The bonded terms generally consist of energetic penalties that arise from any deviations from an equilibrium value, while the van der Waals and electrostatic terms are described by an Lennard-Jones and Coulombic terms, respectively. A more complete description of these terms follows in Equation 2.2.

$$E_{\text{pot}} = \sum_{\text{bonds}} \frac{k_i}{2} (D_i - D_{\text{ref}})^2 + \sum_{\text{angle}} \frac{k_i}{2} (\theta_i - \theta_{\text{ref}})^2 + \sum_{\text{torsion}} \frac{V_i}{2} (1 + \cos(n\omega_i - \gamma_i)) + \sum_{\text{van der Waals}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{\text{electrostatic}} \left(\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \quad \text{Equation 2.2}$$

The first two terms are simple harmonic potentials, which contribute energy penalties for any deviation from reference values. The third term contains a description of the energy barrier for the torsion potential (V_n), the multiplicity of the torsion curve (the number of minima during a 360 rotation) n , a phase factor γ , and the torsion angle itself ω . The van der Waals term is often described by a Lennard-Jones potential, as seen in Equation 2.2, in which a well depth (ϵ_{ij}) and collision diameter (σ_{ij}) are used with the inter-atomic distance (r_{ij}) to approximate the energetic potential. The electrostatics are described by a Coulmbic term which includes the charge of the interacting atoms q_i, q_j and the inter-atomic distance.

Calculation of acceleration and velocity

There are several methods to obtain atomic velocities given the forces on a set of atoms. Given the total forces on a given particle at time t , the positions and velocities of that particle are then determined at some time $t + \delta t$. A general solution for this problem was described in 1967 by Verlet¹.

One of the many variations, as implemented by AMBER², is the leap-frog Verlet algorithm³ which uses several relationships to describe atomic position (x), velocity (v), and acceleration (a) with regard to time (t). Calculation of the acceleration for a given atom comes directly from the force (F), and the mass of the atom (m), as seen in Newton's second law (Equation 2.3).

$$F = ma \quad \text{Equation 2.3}$$

The position for an atom at position x , at time $(t + \delta t)$ is given by Equation 2.4

$$x(t + \delta t) = x(t) + \delta t v(t + \frac{1}{2} \delta t) \quad \text{Equation 2.4}$$

The velocities at time $(t + \frac{1}{2} \delta t)$ can be calculated from Equation 2.5.

$$v(t + \frac{1}{2} \delta t) = v(t - \frac{1}{2} \delta t) + \delta t a(t) \quad \text{Equation 2.5}$$

The leap frog algorithm then involves the following steps:

- 1) Calculation of $v(t + \frac{1}{2} \delta t)$ from $v(t - \frac{1}{2} \delta t)$ and $a(t)$
- 2) Determination of the new atomic position:

$$r(t + \delta t) = r(t) + \delta t v(t + \frac{1}{2} \delta t)$$

The velocity calculations jump over the positions, giving the name leap-frog.

Electrostatics

Coulombic evaluation of electrostatic forces

The calculation of the electrostatic interactions of a molecular complex remains one of the most important parameters in molecular dynamics calculations. In force fields using the partial atomic charge model, the evaluation of electrostatics is based on a simple Coulombic evaluation of the interaction energy between 2 charges.

$$E_{\text{elec}} = \left(\frac{1}{4\pi\epsilon_0} \right) \left(\frac{q_1 q_2}{r_{12}^2} \right) \quad \text{Equation 2.6}$$

These forces are simple to calculate for a pair of atoms, or a relatively small system. As the number of atoms increases, however it becomes prohibitively expensive to calculate all the possible individual interactions.

Non-bonded pair lists

For the evaluation of non-bonded interactions (electrostatic and van der Waals energies), during a single timestep of an MD calculation, the first step is to create a list of all of the pairwise interactions to be evaluated. The computational cost of generating this list increases as N^2 , and the generation of this list, termed the non-bonded pair list, is a major bottleneck for MD calculations. There are several strategies for handling this problem. The most simplistic approach is to introduce an arbitrary cutoff for the consideration of electrostatic forces. This approach has been widely used for the simulation of proteins and nucleic acids⁴. However, the elimination of long range electrostatic effects has been shown to affect the accuracy of accurate structural representation during MD simulations⁵, and may be handled better by an alternative Ewald summation approach⁶.

Ewald summation of electrostatic forces

The Ewald sum⁷, originally devised to examine the energetics of ionic crystals, is an alternative to the distance based cutoff. Bypassing the generation of the standard non-bonded pairlist, and force calculations, it is possible to perform MD calculations without generation of artifacts arising from a distance-base cutoff, or protein restraints. In the Ewald sum, an atom interacts with all of the other atoms within the defined periodic system (box), as well as those contained within all other periodic cells. The calculation of potential energy, due to charge-charge interactions between two atoms within the periodic box, is exactly the same as that used for the distance-based cutoff.⁸

$$E_{\text{pot}} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}^2} \right) \quad \text{Equation 2.7}$$

For a periodic system, boxes are arrayed in three dimensions, but to simplify the discussion, assume a 2 dimensional array of boxes. For a box at point n , with lattice coordinates n_x and n_y , the potential energy is shown in Equation 2.8.

$$E_{\text{pot}} = \frac{1}{2} \sum_n \sum_{i=1}^N \sum_{j=1}^N \left(\frac{q_i q_j}{4\pi\epsilon_0 |r_{ij} + n|} \right) \quad \text{Equation 2.8}$$

These equations attempt to describe the way in which the total energy contains contributions from interactions within the periodic box, as well as from periodic images and the interaction with the surrounding medium, described graphically in Figure 2.1. The calculation of the Ewald sum, as described, is computationally expensive, and its utilization in biomolecular systems requires algorithms which use several approximations in order to speed up the calculation.⁹

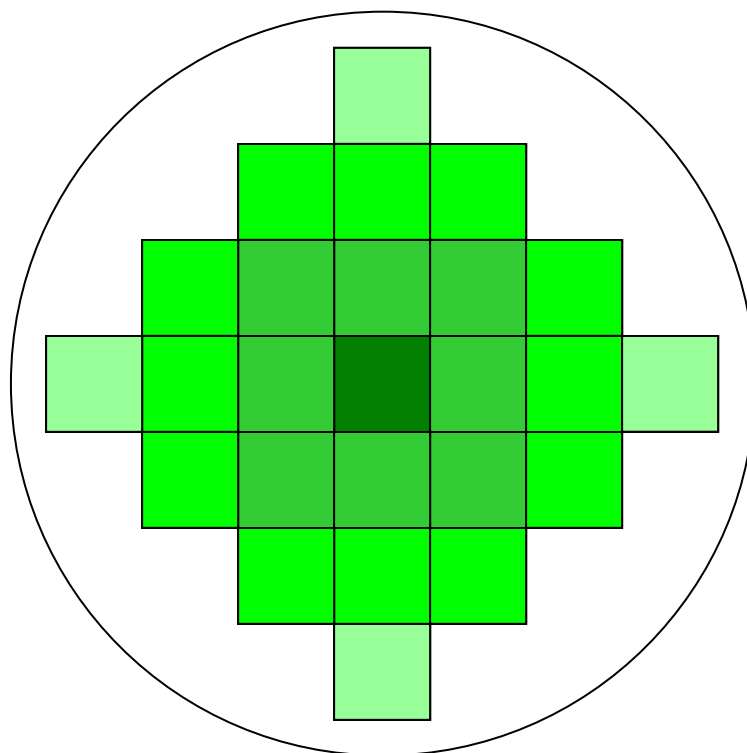


Figure 2.1 Representation of the relative contributions to the electrostatic potential as calculated in the Ewald sum. Neighboring periodic boundary units are colored according to the degree to which their interactions with atoms in the central unit contribute to the electrostatic potential, in a roughly spherical pattern. Figure modified from Leach⁸.

As implemented in AMBER, the particle mesh implementation of Ewald summation (PME),¹⁰ is used for increased computational speed ($N\log N$ vs. N^2). The Ewald sum is interpolated onto a grid, using a β -spline function, to obtain the electrostatic potential at each point, and the resulting potential is used to calculate the electrostatic forces on the atom.

Solvation models

Simulations of complex biomolecular systems require the simulation of the effects of solvent on the system. There are several ways of simulating these effects, with solvent represented either explicitly or implicitly. In an explicit representation of

solvent, solvent molecules (most often water) are represented as separate molecules in the simulation, and are equilibrated and heated with the system of interest. Explicit representation of water was used in all of the simulations performed in this work. Implicit solvation models, in which solvent effects are represented mathematically have become important in some of the energetic analysis of the explicitly solvated trajectories.

Explicit solvation

For most biomolecular simulations, water is the solvent, and this discussion will be limited to consideration of water models. The nature of the AMBER forcefield requires the use of water molecules parameterized with explicit partial charges. There are several empirical water models currently in use, which differ in their treatment of partial charges, overall geometry and ability to properly represent bulk properties. The TIP3P water model,¹¹ used in our simulations is a simple model of the water molecule, in which positive partial charges on the hydrogens is balanced by the negative charge on the oxygen. The TIP3P model works well in MD simulations and the simplicity of the model reduces the number of pairwise interactions that must be calculated for solvent - solvent interactions, compared with other models.

In addition to the choice of water model, there are several additional considerations when choosing the solvent system to be used. For simulations in which computational time is an important consideration, fewer solvent molecules allow for greater computational speed. These considerations led to the creation of several solvent models which sought to minimize the number of water molecules to be

simulated, while maintaining a system in which solvent interactions with the solute are properly represented.

For explicitly solvated simulations there are a number of ways in which the solvent can be placed around the system. One simple model is to place a sphere of solvent around the active site, centered so as to solvate the region of interest^{12,13}. These droplet simulations use a simple harmonic restraint in order to hold the water droplet in place. While there are potential artifacts associated with this solvation model¹⁴, it offers a savings in computational time compared with more rigorous solvation models. In the periodic boundary condition (PBC) system, the solvent is placed around the molecule in a defined shape, such as a cube, or truncated octahedron, and this system is packed into a repeated lattice, to remove edge effects seen in droplet simulations. We have performed simulations using both the droplet model, and a repeating cubic lattice, in order to obtain a simulation protocol which is able to reproduce experimental structural features, as well as computational efficiency. A comparison of the solvent models is shown in Figure 2.2.

Because the simulation time increases as N^2 , it remains important to choose a molecular system that contains the minimum number of particles necessary to accurately describe the behavior of the interactions under examination. While the choice of solvation model can have a large effect on the number of particles, it is probably even more important to choose a molecule, or molecular fragment of a size which is compatible with obtaining a trajectory of sufficient length (2-10ns). In some

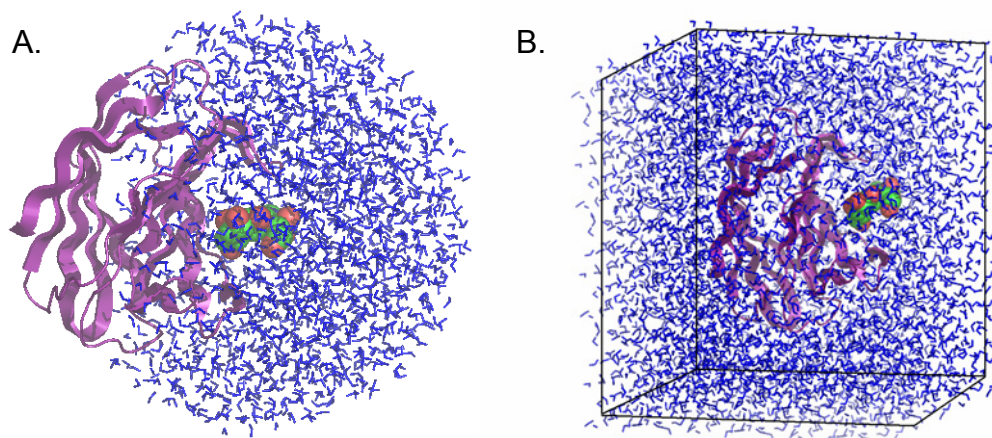


Figure 2.2 A comparison of droplet and periodic box solvation models, A. 26Å droplet of TIP3P water, centered on the ligand center of mass and B. a 45 x 55 x 60Å box of water solvating the entire galectin-1 – ligand complex.

cases, this choice is relatively easy, and the biological nature of the system under study can determine the simulation system. In a homomultimeric system, it may also be possible to simulate a single monomer. In addition, it may be sufficient to represent the dynamics of a subset of the molecule, while restraining the remainder of the molecule during the simulation. It is important to select the simulation system carefully, and to evaluate the effects of various approximations used.

Implicit solvation

In addition to the molecular representation of water molecules in MD simulations, there are other methods which seek to represent the bulk solvation effects of water, without including them explicitly. These methods of continuum solvation seek to represent the electrostatic effects of water by using a number of approximations to the dielectric of the solute and solvent. These methods can be used to derive MD trajectories of systems too large to represent with explicit solvation. In this work,

however, we have used them to obtain estimates of the free energy of solvation, a necessary component of the energetic analysis of our explicitly solvated trajectories.

In the continuum solvent models, an implicit approximation of solvent effects is used to obtain the electrostatic component of the solvation free energy. These methods treat solvent as a non-structured high dielectric medium, with a cavity of low dielectric, representing the solute (protein). The general form of this representation is shown in Figure 2.3.

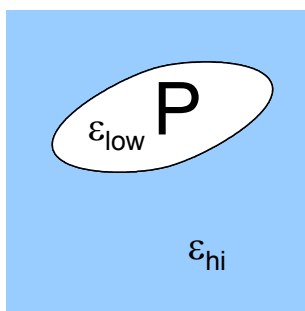


Figure 2.3 Representation of an implicitly solvated complex. The protein is treated as a region of low dielectric (2-4) while the solvent (water) is treated as a region of high dielectric (80).

The solvation free energy can be divided into several components which describe both the polar (electrostatic) and non-polar contributions. These components consist of the work needed to create a distribution of charge ΔG_{pol} and the work needed to create a cavity within the solvent ΔG_{np} , the non-polar contribution. A simple thermodynamic cycle showing the relationship of these quantities is shown in Figure 2.4.

The total solvation free energy ΔG_{solv} can be obtained from the following relationship, described in Equation 2.9.

$$\Delta G_{solv} = \Delta G_{elec,v} + \Delta G_{np} + \Delta G_{elec,wat} \quad \text{Equation 2.9}$$

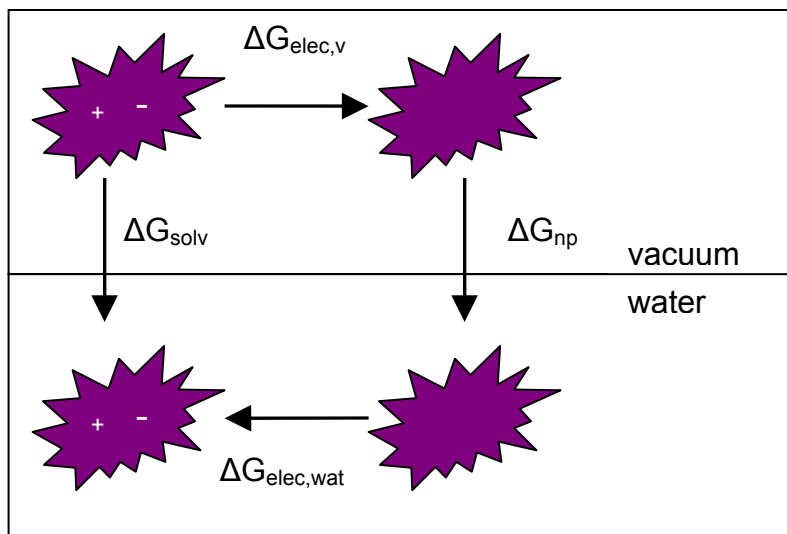


Figure 2.4 Thermodynamic cycle showing the contributions of electrostatic and non-polar contributions to the solvation free energy (modified from¹⁵).

The electrostatic components of this solvation energy are accounted for in the Poisson-Boltzman and Generalized Born continuum solvent models, described in the following section. The non-polar contribution arises from van der Waals interactions between surface groups on the solute, and solvent molecules and is proportional to the surface area of the molecule.

Poisson-Boltzman

Accurate estimations of solvation free energy can be obtained from the analytical solution of the Poisson-Boltzman equation. The Poisson equation is a simple relationship describing the variation in electrostatic potential (ϕ) of a given charge density (ρ) within a medium with a dielectric (ϵ) at a given position (r). The Poisson equation for a system of uniform dielectric is simply Coulomb's law. If, however, the dielectric varies with position, the Poisson equation takes the form seen in Equation 2.10.

$$\phi_r = -\frac{4\pi\rho_r}{\epsilon_r} \quad \text{Equation 2.10}$$

To account for salt effects, a Boltzman term is included which describes the energy required to bring an ion from infinite distance to a position r . This is shown in Equation 2.11. The variables ϕ, ϵ, κ and ρ are all dependent on the position r . Ionic strength is contained in the variable κ , and the electrostatic potential is given in units of $\frac{kT}{q}$.

$$f(\phi) = -\epsilon\kappa^2 \sinh(\phi) + \left(\frac{4\pi\rho_r}{kT} \right) \quad \text{Equation 2.11}$$

The electrostatic potential of an atomic system can be calculated using the Delphi program¹⁶, which contains an implementation of the Poisson-Boltzman equation suitable for computational studies. This approach has been utilized in a number of systems to rationalize the biological implications of charge – charge interactions in molecular function^{17,18}. The result of the Delphi programs is the electrostatic potential at a series of grid points. The solution of the Poisson-Boltzman equation is considered to be the most accurate method of solving for solvation free energy. Its computational cost, however, limits its general utility, and the development of other models has been undertaken to address this limitation.

Generalized Born

In the Born model¹⁹ of continuum electrostatics, the system is represented by a series of charged spheres embedded in a solvent of constant dielectric. In this model, the interaction of a spherical solute with radius R is dependent only on the net charge (q), and represents the energy needed to transfer a given ion from vacuum to the medium with dielectric ϵ .

$$\Delta G_{\text{elec}} = -\frac{\epsilon - 1}{2\epsilon} \frac{q^2}{R} \quad \text{Equation 2.12}$$

As implemented in AMBER, the generalized Born equation takes the form seen in Equation 2.13. In this case q_i and q_j are atomic partial charges, and κ accounts for salt effects. The f_{GB} is a function which interpolates from r_{ij} at long distances, and the Born radius, a_{ij} , at short distances.

$$\Delta G_{\text{elec}} = -\frac{1}{2} \left(1 - \frac{e^{-\kappa f_{\text{GB}}}}{\epsilon} \right) \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{f_{\text{GB}}} \quad \text{Equation 2.13}$$

$$f_{\text{GB}} = \sqrt{(r_{ij}^2 + a_{ij}^2 e^{-D})} \quad D = \frac{r_{ij}^2}{(2a_{ij})^2}$$

A critical parameter determining the accuracy of the generalized Born approach is the Born radius. As shown in Figure 2.5, the Born radius is a measure of the ‘buriedness’ of an atom within the molecule. Larger born radii result in smaller contributions to the solvation free energy. For a spherical molecule, the Born radius is approximately equal to the distance from the center of the atom to the molecular

surface. Born radii for macromolecules have been optimized to take into account AMBER partial charges²⁰.

The generalized Born approximation has been shown to compare favorably with results from the numerical results obtained from the Poisson-Boltzman equation when small molecules are considered^{20,21}. For molecules containing a significant interior volume, however, a much higher level of discrepancy has been observed²², due to the difference in treatment received for surface residues compared with those found in the interior. An overestimation of the solvation energy of deeply buried atoms is seen,

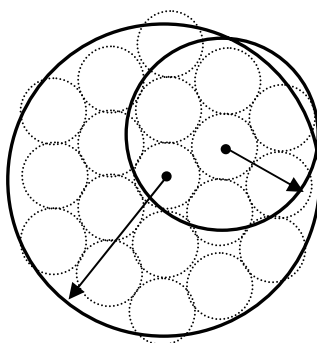


Figure 2.5 The Born radius of atoms within a spherical molecule. Two different Born radii are illustrated for two atoms (modified from²³).

which is thought to be due to the fact that the interior space between the atoms is filled with high dielectric medium. This results in an overestimation of the interior dielectric value and an underestimation of the Born radii. The calculation of solvation free energies for proteins is especially difficult due to both a large number of interior residues, as well as charged residues.

MD simulation of galectin-1 – ligand complexes

Ligands for galectin-1

In our examination of oligosaccharide binding by galectin-1, we chose to focus on a set of ligands that are likely *in vivo* ligands for galectin-1. Type II LacNAc (Gal- β -(1,4)-GlcNAc), the disaccharide core of the ligands examined in this work, is found in *N*- and O-linked glycans, as well as glycolipids. The closely related disaccharide, Type I LacNAc (Gal- β -(1,3)-GlcNAc), differs only in sugar linkage and is structurally similar to Type II LacNAc (LacNAc_{II})²⁴. As previously described, Neu5Ac- α -(2,3)-LacNAc is a potentially important biological ligand, due to its prevalence in the termini of mammalian *N*-linked glycans. 3'-O-sulfation of LacNAc has been observed in glycolipids from brain, kidney, spleen, granulocytes, stomach and intestine^{25,26}. It has been observed that affinity for galectin-1 increases with degree of LacNAc polymerization^{27,28}, until the polylactosamine chain reaches 4 repeating units in length. This observation was explored in simulations of the tetrasaccharide Gal- β -(1,4)-GlcNAc- β -(1,3)-Gal- β -(1,4)-GlcNAc or (LacNAc)₂. The ligands used in the molecular dynamics studies are shown in Figure 2.6

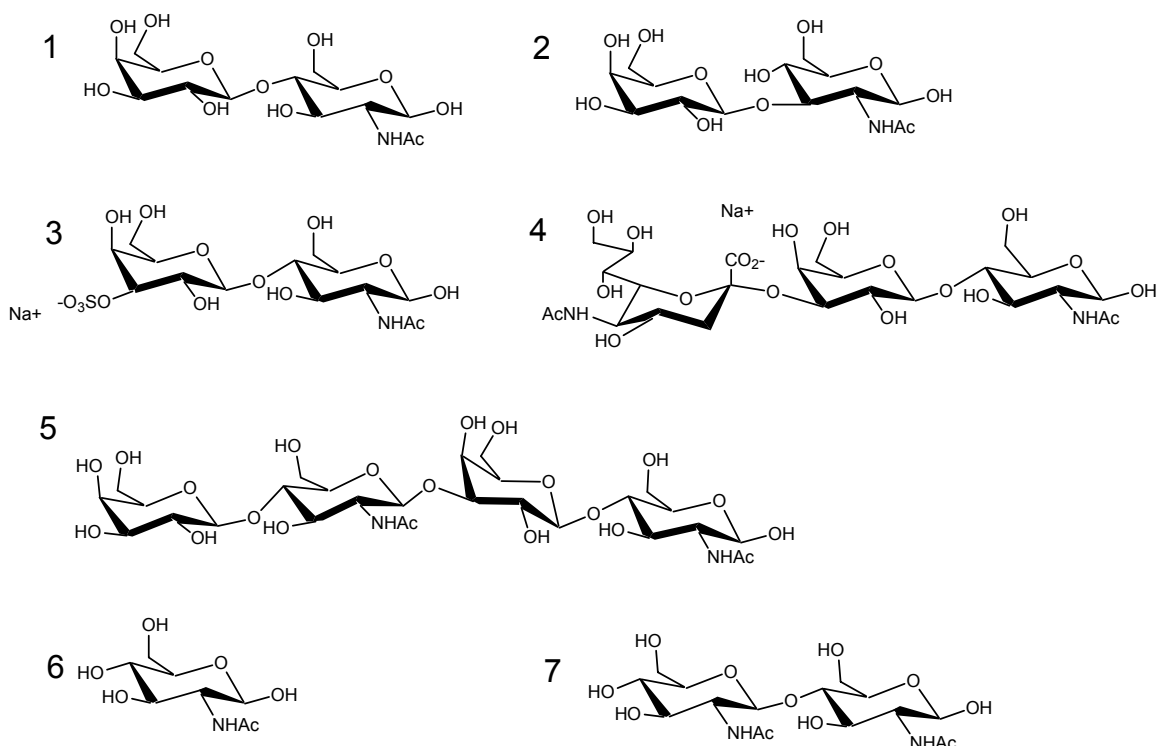


Figure 2.6 Oligosaccharide ligands used in MD simulations **1** Gal- β -(1,4)-GlcNAc (LacNAc_{II}), **2** Gal- β -(1,3)-GlcNAc (LacNAc_I), **3** 3'-O-SO₃-LacNAc, **4** Neu5Ac- α -(2,3)-LacNAc, **5** LacNAc- β -(1,3)-LacNAc (LacNAc)₂ **6** GlcNAc, **7** Glc- β -(1,4)-GlcNAc.

The affinity of oligosaccharide ligands for galectin-1 has been studied by a number of different methods. The earliest semi-quantitative tests for lectin binding involved the ability of a given saccharide to inhibit lectin mediated agglutination of erythrocytes^{29,30}. This method has largely been replaced by more sophisticated ELISA techniques,³¹⁻³³ which quantitate ligand binding through an indirect measure of binding affinity. Frontal affinity chromatography has also recently been used to quantify binding for a number of galectins to a large panel of oligosaccharides²⁷. Intermolecular binding interactions can also be quantified by direct physical methods such as isothermal titration calorimetry (ITC)³⁴ and surface plasmon resonance³⁵⁻³⁷. ITC has been widely used to study lectin – carbohydrate interactions³⁸, and has been used in several studies

of galectin-1^{39,40}. In addition, our lab has undertaken a collaboration to use SPR⁴¹ to obtain dissociation constants for the binding of galectin-1 to a number of 3'-O-substituted ligands⁴². While there is some variability in the exact values for different methods, several general trends emerge from the data, as seen in Table 2.1. To facilitate comparison between methods, all values have been normalized to lactose.

Table 2.1 Relative affinities^a of galectin-1 ligands^b used in MD simulations.

Ligand	K _a ^c	K _a ^d	Activity ^e
	SPR ⁴²	ITC ⁴⁰	ELISA ³¹
Lactose (Gal-β-(1,4)-Glc)	1.0	1.0	1.0
1 LacNAc _{II} (Gal-β-(1,4)-GlcNAc)	2.4	5.5	5.5
2 LacNAc _I (Gal-β-(1,3)-GlcNAc)		4.0	
3 3'-SO ₃ -Gal-β-(1,4)-GlcNAc	15.5		16.5
4 Neu5Ac-α-(2,3)-Gal-β-(1,4)-GlcNAc	3.3		
5 (LacNAc) ₂ (Gal-β-(1,4)-GlcNAc-β-(1,3)-Gal-β-(1,4)-GlcNAc)			
6 GlcNAc	NA	NA	NA
7 Glc-β-(1,4)-GlcNAc	NA	NA	NA

^a Relative to Gal-β-(1,4)-Glc-β-OR. ^b For SPR studies, sugar is linked to Biacore chip via spacer which is β-linked to the reducing termini of sugar ligands; other studies use the free saccharide. ^c Chinese hamster galectin-1 (C2S mutant).

^d Bovine spleen galectin-1. ^e Alkylated human galectin-1.

The binding data in Table 2.1 clearly illustrate the remarkable ability of the galectin-1 CRD to bind a variety of glycans. The similar affinities of LacNAc_{II} (**1**) and LacNAc_I (**2**), highlight the fact that the change in linkage has little effect on binding affinity for galectin-1. The increase in affinities observed for anionic ligands implicates electrostatic interactions in ligand binding. For example, 3'-O-sulfation of LacNAc increases affinity six-fold. Lower affinities seen for the 6'-O-sulfated oligosaccharide and Neu5Ac-α-(2,6)-LacNAc²⁷ (not shown) relative to the 3'-O analogs, suggest that the position of the anionic groups is important. Substitution at the O6 is likely to interfere

with interactions between galectin-1 and the O6 hydroxyl group. To determine if our simulations are able to replicate the binding selectivity of galectin-1, GlcNAc (**6**), and Glc- β -(1,4)-GlcNAc, the 4'-epimer of LacNAc (**7**), were docked using the coordinates from the galectin-1 – LacNAc structure. As these saccharides do not bind galectin-1 *in vitro*, the simulation of these complexes served as negative controls.

Ligand docking

The oligosaccharide ligands were docked in the binding site by superimposing the LacNAc component with the equivalent residues observed in the crystal structure of the galectin-1 – LacNAc_{II} complex. For the 3'-O- substituted ligands it became necessary to incorporate knowledge of glycosidic torsion angles, and experimental evaluation to obtain the initial model of the docked complex. Initial alignment of the 3'-O-substituents was achieved by evaluating the steric fit as a function of the glycosidic torsion angles. In the case of Neu5Ac- α -(2-3)-LacNAc, each of the three low energy rotamers for the α -(2,3) linkage ($\phi = +60^\circ$, -60° and 180° with $\psi = 0^\circ$) was initially examined. Analysis of these three complexes showed that the -60° structure made steric clashes with the protein and was eliminated from further study. Both the $+60^\circ$ and 180° structures were able to fit and were subjected to preliminary MD refinement, during which the $+60^\circ$ rotamer spontaneously interconverted to the 180° structure. All subsequent studies were performed with the 180° Neu5Ac- α -(2-3)-LacNAc structure. LacNAc_I, in which the β -(1,4) glycosidic linkage is replaced by a β -(1,3) linkage, was overlaid with the heavy atoms of the sugar rings of LacNAc_{II}. This change in linkage resulted in a conformation in which the *N*-acetyl group was positioned 180° away from its position in the crystal structure. For (LacNAc)₂, the disaccharide containing the

reducing terminus was superimposed with the LacNAc disaccharide from the crystal structure. The glycosidic torsion angles associated with the LacNAc at the non-reducing terminus of the tetrasaccharide were varied so as to minimize steric clashes with the protein. The resultant orientation ($\phi_{1-4} = 56.3^\circ$ $\psi_{1-4} = -4.6^\circ$ $\phi_{1-3} = 41.1^\circ$ $\psi_{1-3} = -15.4^\circ$) was consistent with expectations based on the *exo*-anomeric effect⁴³. Neu5Ac- α -(2,3)-LacNAc was neutralized with a Na^+ counterion placed 2.5 Å away from the carbon of the carboxylic acid, along the bisector of the O-C-O angle. 3'-OSO₃-LacNAc was treated in a similar manner, with the counterion placed 2.5 Å away from the sulfur atom, along the trisector of the SO₃ angle. Structures of the initial docked complexes are shown in Figure 2.7.

Simulation parameters for MD studies

Coordinates for the galectin-1 – LacNAc complex⁴⁴ were retrieved from the Protein Data Bank⁴⁵ (pdbid = 1slt). A single monomer of the dimeric complex was used in the simulations. Hydrogen atoms were added to the X-ray coordinates, and the system was solvated with the EDIT module of AMBER. All histidine residues were assumed to be neutral, and were protonated at the N ϵ position. For the water droplet, a sphere of TIP3P waters with a radius of 26Å was centered at the ligand center of mass. Diffusion of waters out of this droplet was prevented through the use of a half-harmonic potential applied at the droplet surface. For periodic boundary condition (PBC) simulations, the protein – ligand complex was placed within a theoretical box of TIP3P water with approximate dimensions of 45 x 55 x 60Å.

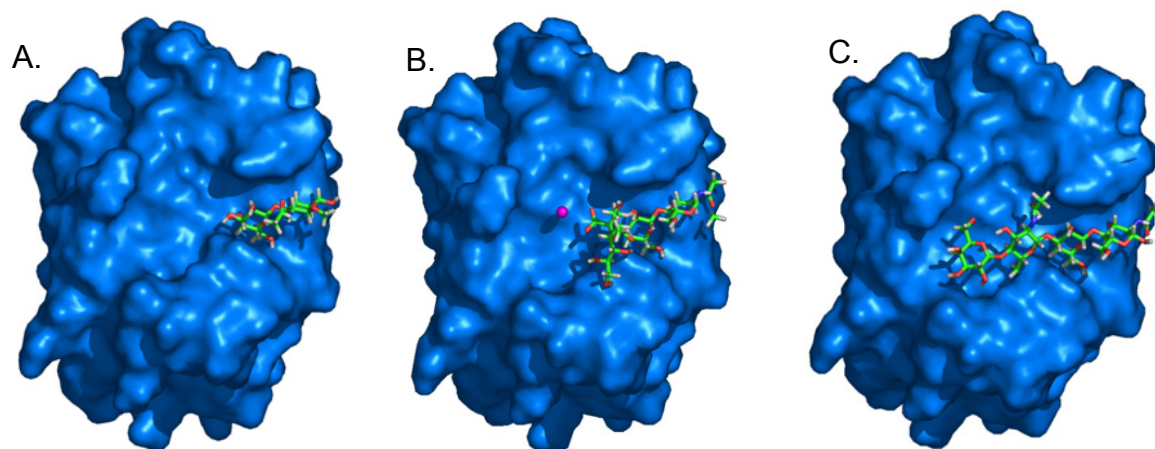


Figure 2.7 Representative docked complexes of: A. LacNAc B. Neu5Ac-LacNAc C. (LacNAc)₂. The sodium counterion is shown as a purple sphere near the carboxylic acid group of Neu5Ac in B.

All simulations on galectin-1 – ligand complexes focused on a single monomer of galectin-1. This decision was based on the symmetrical nature of the galectin-1 dimer, as well as a need to limit the size of the system under consideration, in order to be able to simulate a trajectory of sufficient length within a reasonable timeframe. To date, none of the binding studies on galectin-1 have given any indications of cooperativity in the binding of ligands by the CRD^{27,42}, the monomer is therefore a valid model for the behavior of this protein. In addition to the use of the monomer as our experimental system, we have included additional approximations in order to quickly obtain molecular trajectories of various ligands bound to galectin-1. Just as our choice of monomer decreases the simulation time by decreasing the number of atoms in the simulation, the definition of important binding site residues can also speed up MD calculations.

The binding site for galectin-1 was defined as all amino acids containing any atom within 12 Å of any atom in LacNAc_{II}, and was allowed complete motional freedom.

This 39-residue subset included most of the front β -sheet and the loops that interact with the ligand. The residues included in the binding site definition are shown in Figure 2.8. The remainder of the protein was restrained in its experimentally determined position, as taken from X-ray structure by Liao et al. Any potential artifacts arising from our choice of binding site, were explored in simulations which did not utilize any restraints on protein motions.

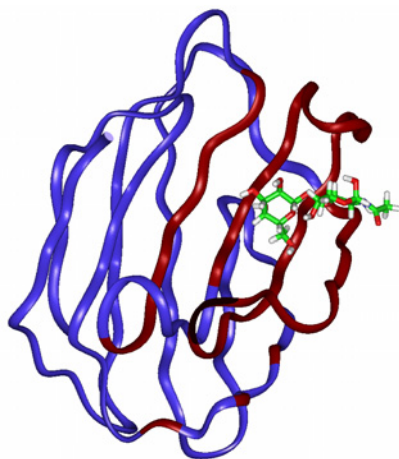


Figure 2.8 Binding site definition for the galectin-1 monomer. The binding site residues are indicated by red coloring of the protein backbone and include residues 27-32, 40-72, 107 and 110.

The simulations were performed with the all-atom AMBER force field⁴⁶, using the PARM94 parameters for proteins⁴⁷, augmented with GLYCAM parameters⁴⁸ (version 99d) for oligosaccharides. Partial atomic charges for the ligands were computed from quantum mechanical molecular electrostatic potentials as reported⁴⁹. The initial unfavorable contacts made by the solvent were removed by 1500 cycles of energy minimization; 10 cycles of steepest descent being followed by 1490 steps of conjugate gradient. The energy of the solvent molecules and binding site residues was then minimized further for 1500 steps. Energy minimization was followed by a 150ps period

of simulated annealing, during which the temperature was raised from 5 to 300K over 50ps, maintained at 300K for 50ps, and then cooled to 5K over 50ps. The energy of the whole system was then minimized, followed by heating from 5 K to 300 K over 50ps, with initial velocities assigned from a Maxwellian distribution at 5 K.

For the simulations utilizing the droplet and PBC solvation methods, an 8Å cutoff was used for calculating non-bonded interactions. 1-4 electrostatics and non-bonded interactions were scaled by the default values of 1/1.2, and 1/2.0, respectively. Production dynamics were performed at 300K using a 2 fs time-step, with the SHAKE algorithm⁵⁰ applied to constrain all hydrogen containing bonds. For the Particle Mesh Ewald (PME) simulation¹⁰, a 1Å grid spacing was used to calculate the electrostatic energies, with a fourth order spline used for interpolation.

Structural Analysis of MD trajectories of galectin-1 – ligand complexes

As a measure of the fitness of our simulation methods, we compared various structural parameters for several simulation types. Our goal was to obtain a simulation method that is able to provide both a realistic model of the behavior of the galectin-1 – LacNAc complex, as well as computational speed. For this comparison, the galectin-1 – LacNAc complex was simulated in a number of conditions; a non-periodic droplet simulation with an 8Å non-bonded cutoff, a periodic simulation with the same 8Å non-bonded cutoff as well as a periodic simulation under particle mesh Ewald (PME) conditions. These three simulations were then compared using several structural parameters of the galectin-1 – LacNAc complex.

Hydrogen bonding in the galectin-1 – LacNAc complex

One measure of the maintenance of structural integrity of this protein – ligand complex is the geometry of the protein – carbohydrate hydrogen bonds which contribute much of the binding specificity. These interactions have been characterized for the X-ray structure of galectin-1 with bound LacNAc, and are primarily interactions with the O4 and O6 of galactose and the O3 of GlcNAc. A comparison of the droplet solvation approach with both the PBC and PBC/PME methods indicated that the droplet simulation gave results comparable to both of the more rigorous treatments, while taking approximately one-third the CPU time of the PBC simulation. An analysis of hydrogen-bond distances in the galectin-1 – ligand complexes is shown in Table 2.2.

The most notable feature of the data in Table 2.2. is the high level of agreement between the values of the interatomic distances seen in the crystal structure with those seen in all of the solvated MD simulations. Nearly all of the hydrogen bonding distances between LacNAc functional groups and the side chains of binding site residues are within 0.5 Å of the experimental distances, with small standard deviations. Specifically, the key interactions between Gal-O4 and His-44 N ϵ , Asn-46 O δ 1 and Arg-48 N η 2 are noteworthy for their relatively short distances and low standard deviations, suggesting that they are very strong hydrogen bonds. The interactions of Gal-O6 (with Asn-61 N δ 2

Table 2.2 Hydrogen bond distances^a, ligand RMS values^b, and simulation parameters for three treatments of solvation and electrostatics for the complex with 1.

Residue	Atom	Ligand Atom	X-ray ⁴⁴	Droplet	PBC	PBC/PME
H44	Nε	Gal-O4	2.8	3.0 (0.2) ^c	3.0 (0.2)	3.4 (0.4)
N46	Oδ1	Gal-O4	3.4	3.4 (0.3)	3.4 (0.3)	3.4 (0.3)
R48	Nη2	Gal-O4	3.0	2.9 (0.2)	2.9 (0.2)	3.0 (0.2)
R48	Nη2	Gal-O5	2.9	3.0 (0.2)	2.9 (0.1)	3.0 (0.2)
N61	Nδ2	Gal-O6	2.7	3.0 (0.2)	2.9 (0.1)	3.0 (0.2)
D71	Oε1	Gal-O6	4.7	4.3 (0.6)	3.8 (0.9)	4.6 (0.2)
D71	Oε2	Gal-O6	2.8	2.9 (0.6)	3.5 (0.9)	2.7 (0.1)
R48	Nη1	GlcNAc-O3	2.8	2.9 (0.1)	2.9 (0.1)	2.8 (0.1)
R48	Nη2	GlcNAc-O3	3.2	3.3 (0.3)	3.4 (0.3)	3.3 (0.2)
R48	Nη1	GlcNAc-O4	4.2	4.1 (0.3)	4.1 (0.3)	4.3 (0.3)
D48	Nη2	GlcNAc-O4	3.6	3.6 (0.3)	3.5 (0.2)	3.5 (0.3)
D71	Oε1	GlcNAc-N	4.0	3.8 (0.5)	4.0 (0.6)	3.5 (0.2)
D71	Oε1	GlcNAc-O3	3.3	2.9 (0.3)	2.9 (0.4)	3.0 (0.2)
D71	Oε2	GlcNAc-O3	2.4	3.2 (0.5)	3.0 (0.4)	2.8 (0.2)
R73	Nη1	GlcNAc-C=O	4.5	4.1 (0.6)	4.3 (0.5)	4.5 (0.6)
R73	Nη2	GlcNAc-C=O	3.3	4.5 (0.4)	4.5 (0.4)	4.6 (0.4)
R73	Nη1	GlcNAc-O3	4.7	4.1 (0.5)	4.5 (0.4)	4.6 (0.4)
R73	Nη2	GlcNAc-O3	3.3	3.1 (0.3)	3.3 (0.3)	3.4 (0.4)
Binding site RMSD			All atoms	1.1	1.0	1.2
			Backbone atoms only	0.7	0.7	0.9
			LacNAc ring atoms only	0.9	0.7	1.1
			Number of particles	7278	16056	16056
			Simulation time (h)	168	436	604

^a Ångstroms. ^b Root-mean-squared deviation (Å) in the non-hydrogen atomic positions, relative to the X-ray structure. ^c Standard deviations in parentheses.

and Glu-71 Oε2) and GlcNAc-O3 (with Arg-48 Nη1 and Glu71 Oε1) also show values indicative of a relatively strong hydrogen bonding network.

The interatomic distances of hydrogen bonded atoms are the most commonly used structural measure of the strength of hydrogen bonds. Nonetheless, the standard deviations obtained from the MD data also help to rank the relative strengths. For example, the standard deviations observed for the interatomic distances between the Gal-O4 and both His-44 Nε and Arg-48 Nη are lower than those observed for the Gal-O6 interaction with the carboxylate oxygens of Glu-71, suggesting that the former are

stronger hydrogen bonds. Nonetheless, the overall agreement between results obtained from the droplet simulation and experiment was satisfactory, and all subsequent simulations were performed with the droplet solvation model. The droplet model also shows a considerable savings in computational time, with the droplet simulation being approximately 2.5 times faster than PBC, and 3.5 times faster than PBC-PME.

Hydrogen bonding between galectin-1 and the LacNAc core of 3'-O- substituted ligands

The agreement between the X-ray structure and the trajectory utilizing the droplet solvation model allowed us to consider a number of complexes of 3'-O-substituted ligands (**3,4** and **5**). These complexes, docked as described previously, were subjected to 2ns of MD, and the resulting trajectories were analyzed for hydrogen bonding patterns. The first consideration was the geometry of the LacNAc core shared by (**2,3,4** and **5**). The results of the hydrogen bond analysis of the LacNAc core for these ligands is shown in Table 2.3.

Interactions involving the Gal-O4 (with His-44 N ϵ , Asn-46 O δ 1 and Arg-48 N η 2), Gal-O6 (with Asn-61 N δ 2 and Glu-71-O ϵ 2) and GlcNAc-O3 (with Arg-48 N η 1 and Glu-71 O ϵ 1) were maintained for all of the ligands. Even in the case of **2**, which contains a Gal- α -(1,3)-GlcNAc linkage, many of the LacNAcII-core interactions were maintained, particularly those between Gal-O4 and His-44 and Gal-O6 and Asn-61. The loss of the contacts between galectin-1 and the GlcNAc in **2** is due to the change in linkage position, which results in a 180° rotation of the GlcNAc ring relative to its orientation in **1**. This conformation has previously been proposed from an analysis of the X-ray structure of galectin-1 bound to a complex biantennary oligosaccharide⁵¹. Notably, this

Table 2.3 Galectin-1 - LacNAc hydrogen bond distances and ligand RMSD values^a from 2 ns simulations.

Residue	Atom	Ligand Atom	Ligand				
			1	2	3	4	5
H44	Nε	Gal-O4	2.8	3.2 (0.5)	3.0 (0.2)	3.1 (0.2)	2.9 (0.1)
N46	Oδ1	Gal-O4	3.4	3.9 (1.0)	3.7 (0.5)	3.3 (0.2)	3.5 (0.2)
R48	Nη2	Gal-O4	3.0	3.8 (1.1)	3.2 (0.4)	2.9 (0.1)	2.9 (0.1)
R48	Nη2	Gal-O5	2.9	3.5 (0.8)	3.0 (0.2)	2.9 (0.1)	3.1 (0.2)
N61	Nδ2	Gal-O6	2.7	3.0 (0.2)	3.0 (0.2)	2.9 (0.1)	3.0 (0.1)
D71	Oε1	Gal-O6	4.7	5.3 (1.4)	3.9 (0.8)	4.8 (0.2)	5.3 (0.1)
D71	Oε2	Gal-O6	2.8	5.1 (0.8)	4.3 (0.7)	2.9 (0.1)	2.8 (0.1)
R48	Nη1	GlcNAc-O3	2.8	4.1 (1.2)	3.0 (0.2)	3.7 (0.2)	3.1 (0.1)
R48	Nη2	GlcNAc-O3	3.2	3.8 (0.8)	3.5 (0.3)	5.0 (0.2)	4.5 (0.2)
R48	Nη1	GlcNAc-O4	4.2	3.6 (0.6)	3.9 (0.3)	3.7 (0.2)	3.6 (0.2)
R48	Nη2	GlcNAc-O4	3.6	3.2 (0.4)	3.5 (0.3)	3.4 (0.2)	3.3 (0.2)
D71	Oε1	GlcNAc-N	4.0	-----	4.0 (0.5)	2.9 (0.2)	3.0 (0.1)
D71	Oε1	GlcNAc-O3	3.3	-----	2.9 (0.5)	2.7 (0.1)	2.7 (0.1)
D71	Oε2	GlcNAc-O3	2.4	-----	3.9 (0.7)	5.1 (0.6)	6.6 (0.2)
R73	Nη1	GlcNAc-CO	4.5	-----	4.0 (0.5)	4.7 (0.4)	7.7 (0.2)
R73	Nη2	GlcNAc-CO	3.3	-----	4.4 (0.4)	4.5 (0.3)	2.8 (0.1)
R73	Nη1	GlcNAc-O3	4.7	-----	3.7 (0.3)	3.3 (0.3)	3.1 (0.1)
R73	Nη2	GlcNAc-O3	3.3	-----	3.0 (0.1)	2.9 (0.1)	2.8 (0.1)
RMSD	CRD All atoms			1.2	1.2	1.4	0.8
	CRD Backbone			0.9	0.7	0.9	0.5

^aRoot-mean-squared deviation (Å) in the non-hydrogen atomic positions, relative to the initial structure.

conformational change has no significant change in affinity relative to **2** (see Table 2.1).

The maintenance of the core galectin-1 – LacNAc hydrogen bonds is one indication of the ability of these interactions to anchor a number of oligosaccharide ligands in the CRD. In addition to this interface between the LacNAc core and the CRD of galectin-1, it is possible to observe the presence of hydrogen bonds between 3'-O-substituents and the proposed extended binding site of galectin-1. Hydrogen bonds

between the extended portions of the non-Gal-terminating ligands and the CRD are presented in Table 2.4.

Table 2.4 Predicted hydrogen bond distances between extended ligands and galectin-1 CRD.

Ligand	Atom	Residue	Atom	Distance
4	Neu5Ac-O9	N33	Nδ2	4.7 (1.4)
	Neu5Ac-O1A ^a	H52	Nε	4.4 (0.5)
	Neu5Ac-O1B	H52	Nε	3.1 (0.5)
	Neu5Ac-O7	W68	Nε1	3.3 (0.4)
5	Gal ₂ -O3 ^b	D38	Oδ2	3.8 (0.8)

^a O1A and O1B are the carboxylate oxygen atoms in Neu5Ac, with O1A being more deeply buried in the protein binding site. ^b Gal₂ is the galactose of the non-core LacNAc in **5**.

For the complex with **4**, the interactions between His-52, and the carboxylic acid of Neu5Ac, as well as those between Trp-68 and the glyceryl side chain of Neu5Ac are the primary interactions between the CRD and the sialic acid. In the terminal LacNAc unit of **5**, the strongest hydrogen bonds were observed between Asp38 and Gal2-O3, as well as between Asp38 and the amide proton of the N-acetyl group in GlcNAc₂. The binding of **1** to the CRD, and the interactions between the extended portions of **4** and **5** are similar, in that the protein-carbohydrate interactions seen are confined to one face of each oligosaccharide. In addition, both interactions have at most only one or two strong hydrogen bonds between the CRD and extended sugars.

Aromatic ring stacking

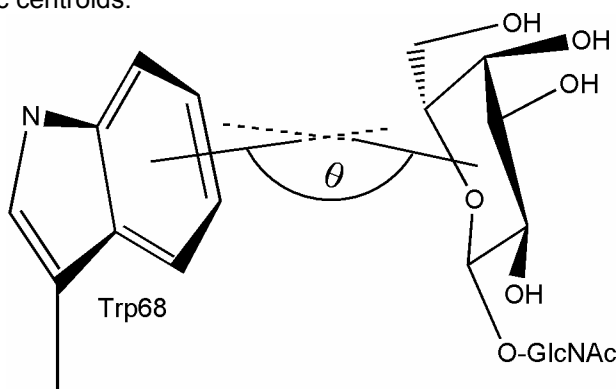
In addition to hydrogen bonding between the protein and oligosaccharide ligands, hydrophobic interactions between sugar rings and aromatic amino acid side chains in the CRDs of lectins and anti-carbohydrate antibodies are commonly observed^{52,53}. The

geometries of the aromatic stacking interactions, between the conserved Trp-68 and the galactosyl ring in each complex, are presented in Table 2.1. We have characterized this interaction by the angle (θ) between the normals to the planes defining the galactosyl ring and the six membered ring of tryptophan. For a perfectly parallel stacking arrangement θ would have a value of 180° . In the simulations, each ligand formed a stacking interaction with Trp-68, with an overall average θ value of 139° , which compares favorably with the X-ray value of 142° .

Table 2.5 Ring stacking between Trp-68 and the galactose ring of core LacNAc.

Parameter	X-ray	1	2	3	4	5
θ^a	142	131	133	132	144	158
R^b	5.1	6.0	5.6	5.4	5.8	5.5

^aThe angle in degrees between the surface normals to the plane of the ring of Trp-68 and galactose. For a perfectly planar system the angle would be 180° . ^bDistance in Ångstroms between the geometric centroids.



Further characterization of these interactions may be obtained by measuring the distance R between the geometric centroids of the pyranosyl and aromatic rings. The average value of R for each ligand was 5.7\AA ; slightly longer than that present in the X-ray structure (5.1\AA), presumably reflecting the influence of internal motions.

Structural analysis of bound and free galectin-1 ligands

The average values for the glycosidic torsion angles of each ligand, for both the protein-bound and free states, are presented in Table 2.6. As is typical for oligosaccharide – protein complexes^{54,55}, the glycosidic linkages of the free oligosaccharides exhibited greater ranges of motion than when bound to galectin-1. The ϕ and ψ torsion angles of the β -(1,4) linkage of the LacNAcII core remained within 15° of the crystal values of 52° and 13°, respectively, for all ligand complexes.

Table 2.6 Average glycosidic torsion angles for ligands in the protein-bound and free states.

Ligand	Linkage		Torsion Angle	
			Bound	Free
1	β -(1,4) ₁	ϕ^a	45.9 (9.3)	46.6 (11.2)
		ψ	15.7 (8.5)	0.6 (12.8)
3	β -(1,4) ₁	ϕ	38.9 (11.4)	47.7 (11.8)
		ψ	17.6 (9.7)	-6.2 (13.2)
4	α -(2,3) ₂	ϕ^b	158.8 (12.9)	-169.7 (12.4)
		ψ	8.1 (11.5)	-13.8 (11.9)
	β -(1,4) ₁	ϕ	45.1 (10.3)	45.1 (11.9)
		ψ	14.2 (9.5)	-6.3 (12.5)
5	β -(1,4) ₃	ϕ	51.1 (11.9)	40.8 (14.7)
		ψ	13.1 (15.3)	-5.4 (15.7)
	β -(1,3) ₂	ϕ	68.6 (14.9)	41.7 (13.8)
		ψ	29.2 (19.1)	11.8 (45.2)
	β -(1,4) ₁	ϕ	55.0 (9.0)	46.9 (12.1)
		ψ	4.0 (7.8)	-3.3 (12.3)

^a ϕ and ψ values for β -(1,3) and (1,4) linkages defined as: H1-C1-Ox-Cx and C1-Ox-Cx-Hx, respectively. ^b ϕ and ψ values for α -(2,3) linkage defined as C1-C2-O3-C3 and C2-O3-C3-H3, respectively.

The solution and bound conformation of the LacNAc core showed little variation in ϕ , adopting the conformation preferred on the basis of the exo-anomeric effect⁵⁶. The ψ angle, however, consistently displayed a modest distortion from the crystal structure

geometry of approximately 15° . The large standard deviations seen in the β -(1,3) linkage of **5** reflect a highly flexible linkage, consistent with predictions based on gas-phase energy calculations for this linkage⁵⁷.

Negative controls

In order to determine the extent to which the simulations were able to discriminate between high and very low affinity ligands⁵⁸, galectin-1 complexes with a monosaccharide, GlcNAc **6** and the 4'-epimer of LacNAc, Glc- β -(1,4)-GlcNAc **7** were examined. Over the course of the simulations, both negative controls diffused out of the binding site within 1500 ps, reaching positional RMSD values of 9Å relative to their initial positions, in 680 and 1403 ps, respectively, as seen in Figure 2.9.

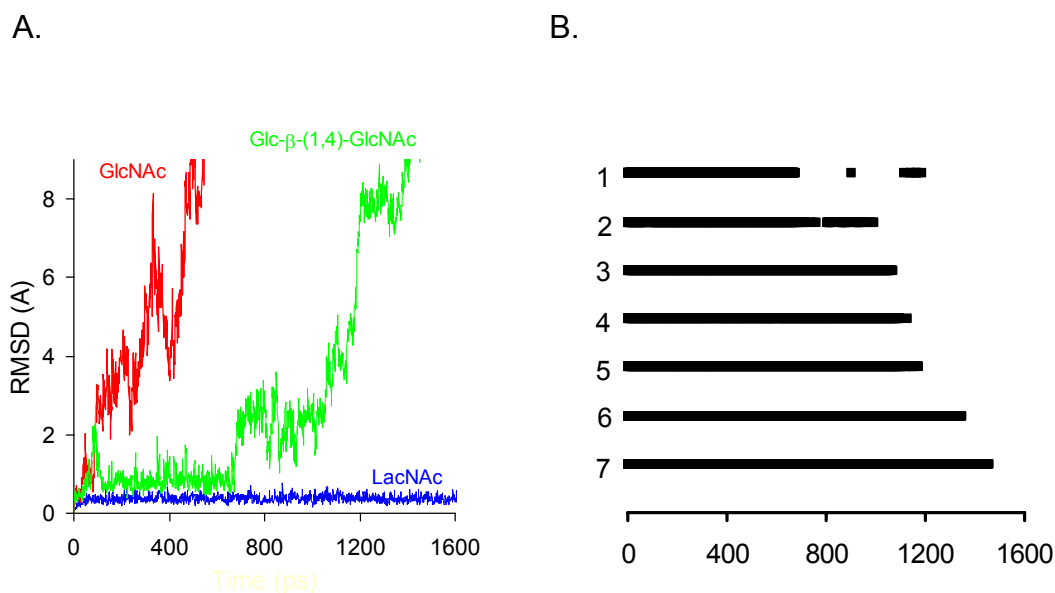


Figure 2.9 RMSD, relative to initial position, for LacNAc, GlcNAc and Glc- β -(1,4)-GlcNAc as a function of simulation time (A). Persistence of hydrogen bonds (B) ($r_{\text{H-bond}} < 4.0$ Å) during the diffusion of Glc- β -(1,4)-GlcNAc (**7**) from the CRD. Lanes in panel B correspond to the following hydrogen bonds: 1) Glu-71 O ϵ 2 – GlcNAc-O3, 2) Arg-73 N η 2 – GlcNAc-O3, 3) Arg-48 N η 1 – GlcNAc-O5, 4) Arg-48 N η 2 – GlcNAc-O4, 5) N η 2 – GlcO5, 6) His-44 N ϵ – Glc-O4, 7) Asn-61 N δ 2 – Glc-O6.

The diffusion may be characterized in terms of the specific sequence of hydrogen bond breakage, and in the case of **7**, began with the disruption of hydrogen bonds between GlcNAc-O3 and both Glu-71-Oε2 and Arg73-Nη2, at approximately 680 ps. This was followed by the loss of three hydrogen bonds between Arg-48 and ligand atoms GlcNAc-O3, GlcNAc-O4, and Glc-O5 between 1075 and 1112 ps. Hydrogen bonding pairs His-44 Nε – Glc-O4 and Asn-61 Nδ2 – Glc-O6 persisted until 1348 and 1457 ps, respectively (see Figure 3b). In contrast to the case of the disaccharide, **6** did not show any persistent hydrogen bonds prior to diffusing out of the CRD.

Electrostatic potential of the galectin-1 surface

The interaction of galectin-1 with anionic ligands such as 3'-O-SO₃-LacNAc and Neu5Ac-α-(2,3)-LacNAc is of particular importance in this work. There are several software packages that allow the creation of a qualitative picture of the electrostatic potential at the protein surface. The GRASP program⁵⁹ was used on the monomer of galectin-1, and the resulting representation of galectin-1 electrostatics is shown in Figure 2.10.

It is evident from Figure 2.10 that there is a region of positive electrostatic potential that correlates well with the position of the 3'OH of LacNAc, the site of attachment for the sulfate of (**3**) and carboxylic acid of (**4**). This type of electrostatic complementarity has been known to be important for protein – nucleic acid⁶⁰ and protein – protein complexes⁶¹, and serves as a long-range attractive force for ligand binding.

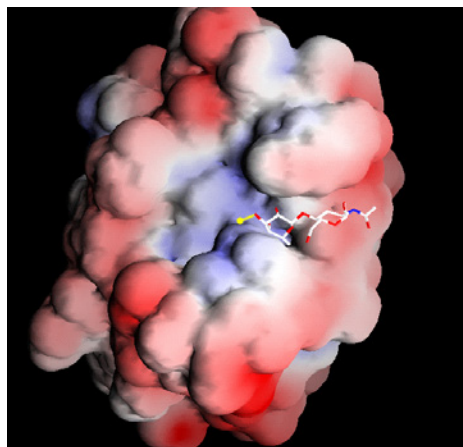


Figure 2.10. The electrostatic potential of the galectin-1 CRD. GRASP representation of the electrostatic potential of galectin-1 mapped onto the solvent accessible surface. Regions of positive potential are shown in blue, and negative potential is shown in red. The values set for the coloring scale are -3.6 and $+3.7$ k_bT , respectively. LacNAc is shown in stick form. The sulfate group of 3'-O-SO₃ is shown in yellow.

Energetic analysis of Molecular Dynamics trajectories

While the analysis of the molecular conformations from MD trajectories can provide a great deal of information on the molecular interactions present in a bound complex, it is also desirable to obtain quantitative energetic information from these simulations. To achieve this goal we employed the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) approach. This approach analyzes snapshots taken from the MD trajectory for various contributions to the binding free energy. For this type of analysis, the solvated trajectory is stripped of all solvent, energetic contributions to the binding free energy are evaluated individually, and then combined to obtain an estimate of the binding free energy. This approach has been used to study a variety of biological processes, including DNA solvation⁶², protein – ligand interactions^{63,64}, protein folding⁶⁵, and carbohydrate – protein interactions⁶⁶. The MM-GBSA method is faster computationally, and results have been seen to compare well with the more rigorous

MM-PBSA method⁶⁷, which differs in the method used (Poisson-Boltzman) to obtain the estimate of solvation free energy.

As in any binding interaction, the binding free energy can be derived from the free energies of the reaction components as seen in Equation 2.14.

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{protein}} - \Delta G_{\text{ligand}} \quad \text{Equation 2.14}$$

These terms may be obtained from a single trajectory, if it is assumed that the MD trajectory of the complex accurately represents the conformations of both the isolated protein and ligand. It is also possible to use separate trajectories of the complex, protein and ligand. There are both enthalpic and entropic contributions to ΔG_{bind} , shown in Equation 2.15.

$$\Delta G_{\text{bind}} = \Delta H_{\text{bind}} - T\Delta S_{\text{bind}} \quad \text{Equation 2.15}$$

The MM-GBSA approach derives these quantities separately. In general, the MM-GBSA method derives the binding free energy as shown in Equation 2.16, with the computational framework shown in Figure 2.11.

$$\Delta G_{\text{bind}} = E_{\text{MM}} + \Delta G_{\text{solv}}^{\text{GBSA}} - T\Delta S_{\text{MM}} \quad \text{Equation 2.16}$$

The molecular mechanical energy, E_{MM} can be obtained directly from the *sander* module of AMBER as described in Equation 2.2. Similarly the entropic term, S_{MM} can be derived from normal mode analysis of representative MD snapshots. The solvation free energy was derived with MM-GBSA, which uses the Generalized Born model to calculate the free energy of solvation.

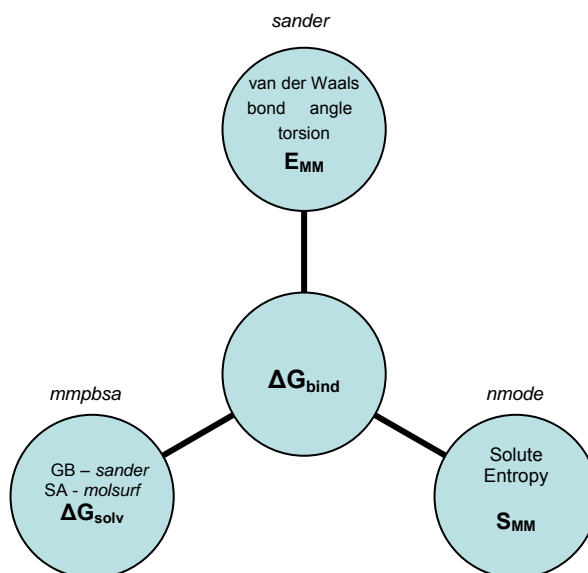


Figure 2.11 Computational framework of the MM-GBSA method. The various contributions to estimations of ΔG_{bind} are shown with the programs or routines used to generate the values.

Nonpolar contribution to the free energy of solvation, ΔG_{solv}

While the calculation of the electrostatic contributions to ΔG_{solv} has been detailed, these polar interactions are not sufficient to obtain an estimate of the free energy of solvation. The nonpolar contributions to ΔG_{solv} are comprised of both a van der Waals and a term describing the size of the cavity made by the molecule, with both terms being dependent on the solvent accessible surface area. These components are especially important for the calculation of the free energy of solvation for proteins, as they are not adequately described simply as a collection of ions, as they have regions that are non polar. It is intuitive that the van der Waals contribution to this energetic component should be related to the surface area, as the surface area determines how

many solvent molecules are able to interact with the protein. The van der Waals term is combined with a term which describes the amount of energy needed to create a cavity within the high dielectric medium and the entropic penalty for reorganization of the solvent molecules. The sum of these terms is equal to the surface area multiplied by an empirically derived constant, which is obtained from experimental data on the energies associated with the transfer of alkanes to water. The surface area is readily calculated by the *mol/surf* module of AMBER, based largely on the Connolly method of calculating surface area⁶⁸.

$$\Delta G_{\text{cavity}} + \Delta G_{\text{vdw}} = kA$$

Equation 2.17

Normal mode analysis

The normal mode analysis (NMA) is a method of extracting the lowest frequency motions in a molecule as a means of obtaining estimates of the entropy of the molecular system. The motions of interest in this analysis are low frequency motions which describe changes in global structure.

There are two general steps in this method, the first is a minimization of a structure (i.e. MD snapshot) and mapping of the energy surface. The next step is to find transition points in the energy surface, and approximate minima by a harmonic function⁶⁹, as seen in Figure 2.12.

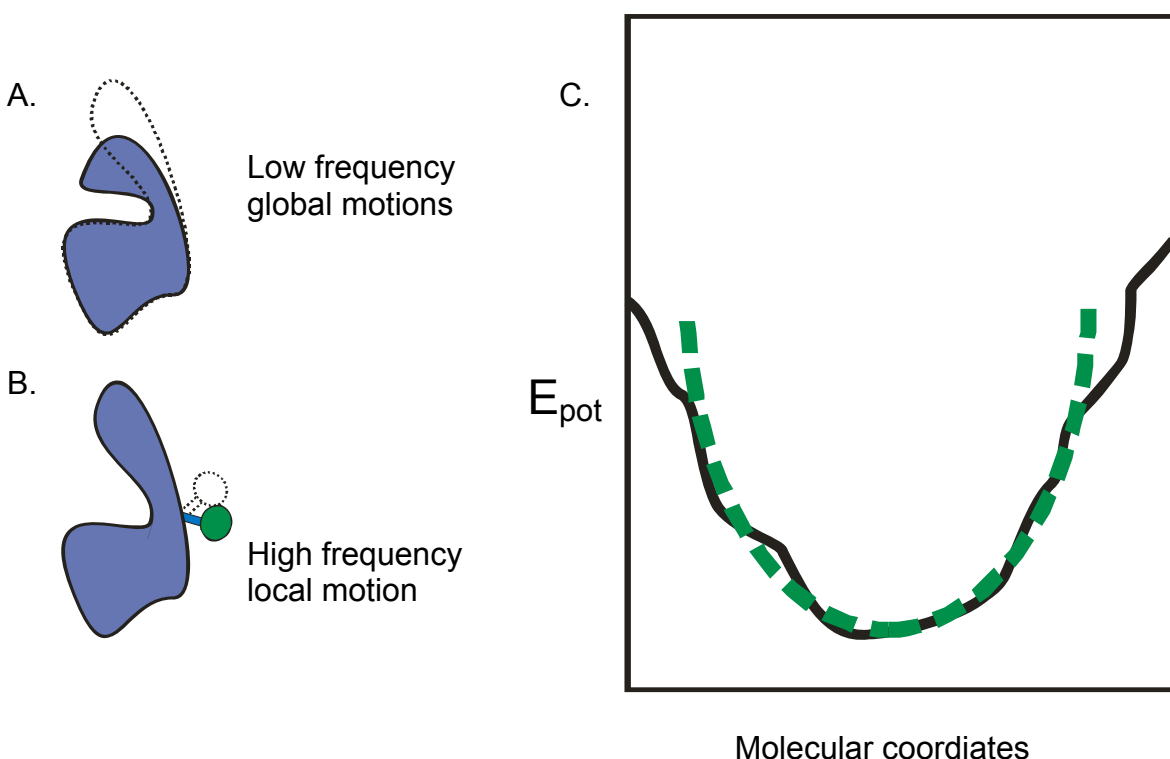


Figure 2.12 Central concepts of normal mode analysis. Representation of protein motions showing both low frequency global motions, A. and high frequency local motions B. C. A 2 dimensional representation of a protein energy surface and corresponding harmonic approximation.

The simplest model of normal mode analysis involves analogy to Hooke's Law. The relationship between a force exerted by a spring to its change in position and a force constant (k), is seen in Equation 2.18.

$$F = -k(x) = \frac{dV}{dx}; x = r - r_{eq} \quad \text{Equation 2.18}$$

This equation can be integrated to give Equation 2.19, giving the potential energy.

$$V = \frac{1}{2} kx^2 \quad \text{Equation 2.19}$$

Using Newton's second law gives Equation 2.20.

$$F = ma = m \frac{d^2x}{dt^2} \quad \text{Equation 2.20}$$

With substitution of the force from Hooke's Law into Equation 2.20, the position at time t can be derived from the amplitude of the motion, A , as well as the frequency of the motion, ν , as seen in Equation 2.21.

$$x(t) = A \sin(2\pi\nu t) \quad \text{Equation 2.21}$$

These frequencies can be related to a vibrational entropy using Equation 2.22. N_A is Avogadro's number, h is Plank's constant, k is Boltzman's constant, T is the temperature and R is the gas constant.

$$S_{\text{vib}} = -R \ln(1 - e^{-h\nu/kT}) + \frac{N_A h \nu e^{-h\nu/kT}}{T(1 - e^{-h\nu/kT})} \quad \text{Equation 2.22}$$

In the case of a protein, this vibrational entropy is a configurational entropy for a given minimized structure. It is important to note that proteins are non-ideal targets for this type of analysis, due to the complexity of the energy surface, and the large frequency range of protein motions⁶⁷. Nevertheless, this method has been used to obtain estimates of the entropy of peptides and proteins⁷⁰, and there are ongoing efforts to improve these estimates by accounting for the non-harmonic nature of energy minima, which utilize the entire MD trajectory^{66,71}. These calculations are very computationally expensive for large systems, and currently can only be performed on a few representative structures.

The energetic analysis of the galectin-1 – ligand trajectories was performed using the following parameters. For the 2ns trajectories, snapshots of the coordinates were taken every 10ps. The resulting 200 snapshots were analyzed with the modified

generalized Born solvation model, modified for use with PARM 94²⁰ to obtain the energetic contributions from solvation. Average molecular mechanical energies were also computed from the same set of 200 snapshots. The normal mode analysis was performed on 10 snapshots, corresponding to 200ps intervals. The energies of the structures used in the normal mode analysis were minimized to within a cutoff of 10^{-4} kcal/(mol·Å). No distance cutoff was applied to non-bonded interactions. The resulting enthalpic and entropic terms were combined to give estimates of the binding free energies.

The accuracy of this approach is enhanced when there is negligible difference between the bound and free conformations for the ligand, as was the case in general here. The overestimation of the absolute binding free energies was due primarily to the use of a vacuum dielectric constant when computing the interior electrostatic interactions ($\epsilon_{\text{int}} = 1$). In calculations employing non-polarizable force fields, or in cases where the protein is not given complete conformational freedom, larger interior dielectric values have been shown to perform well⁷². Further, it should be recalled that the simulations employ the monomer subunit, whereas the experimental data are for binding to the dimer. Presented in Table 2.7 are component energies for each ligand computed with $\epsilon_{\text{int}} = 1$, as well as total binding energies computed with $\epsilon_{\text{int}} = 4$ ⁷³.

Table 2.7 Energy^a component analysis for MD trajectories.

	1	2	3	4	5
$\langle \Delta E_{\text{elec}} \rangle$	-67.5 (5.7) ^b	-30.8 (11.2)	-59.9 (9.7)	-53.2 (8.3)	-103 (15.0)
$\langle \Delta E_{\text{vdw}} \rangle$	-17.5 (3.8)	-17.0 (3.4)	-23.9 (3.2)	-24.4 (3.6)	-37.9 (4.1)
$\langle \Delta E_{\text{MM}} \rangle$	-84.9 (5.9)	-47.9 (12.7)	-83.8 (9.5)	-77.6 (9.7)	-141 (15.4)
$\langle \Delta G_{\text{np}} \rangle$	-3.7 (0.3)	-2.7 (0.7)	-4.2 (0.2)	-4.7 (0.3)	-6.4 (0.5)
$\langle \Delta G_{\text{pol}} \rangle$	15.4 (4.4)	6.5 (8.5)	1.2 (6.8)	6.0 (10.3)	23.9 (12.5)
$\langle \Delta G_{\text{solv}} \rangle$	11.7 (4.5)	3.8 (8.7)	-3.0 (6.8)	1.4 (10.2)	17.5 (12.4)
$\langle \Delta G_{\text{elec,tot}} \rangle$	-52.2 (6.9)	-24.3 (12.7)	-58.7 (8.3)	-47.1 (10.7)	-79.1 (13.7)
$\langle \Delta G_{\text{tot}} \rangle$	-73.3 (7.3)	-44.1 (14.9)	-86.7 (7.8)	-76.2 (11.8)	-123.5 (13.8)
$\langle -T\Delta S \rangle$	18.3 (3.7)	18.2 (4.0)	22.5 (4.3)	30.1 (5.4)	31.6 (2.5)
ΔG_{bind}					
$(\epsilon_{\text{int}} = 1)^c$	-55.0 (8.3)	-25.9 (7.2)	-64.2 (9.0)	-46.1 (7.7)	-91.9 (8.8)
$(\epsilon_{\text{int}} = 4)$	-16.0 (2.6)	-7.7 (1.7)	-20.2 (2.7)	-11.0 (1.6)	-32.8 (2.6)

^akcal/mol. ^bStandard deviations in parentheses. ^c $\langle \Delta E_{\text{elec}} \rangle$ and $\langle \Delta G_{\text{pol}} \rangle$ respond inversely to the dielectric constant.

$$E_{\text{elec}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}}$$

$$E_{\text{MM}} = E_{\text{elec}} + E_{\text{vdW}}$$

$$\Delta G_{\text{elec,tot}} = E_{\text{elec}} + \Delta G_{\text{polar}}$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{non-polar}} + \Delta G_{\text{polar}}$$

$$\Delta G_{\text{tot}} = E_{\text{MM}} + \Delta G_{\text{solv}}$$

Several features are evident from the energetic analysis. Regardless of dielectric constant, the binding free energies correctly rank the affinities of the ligands with the exception that ligand **4** is predicted to bind more weakly to galectin-1 than **1**. Nevertheless, **3** is clearly predicted to bind better to galectin-1 than **1**, while **2** is correctly predicted to be the poorest ligand. It is impossible to discriminate between ligands **1**, **3** and **4** on the basis of net molecular mechanical energies $\langle \Delta E_{\text{MM}} \rangle$ alone, with values being indistinguishable within error limits. Instead, distinguishing the relative affinities of these ligands also requires the consideration of estimated solvation free energies $\langle \Delta G_{\text{solv}} \rangle$ and entropic contributions $\langle T\Delta S \rangle$. For example, the enhanced affinity of the sulfated ligand arises not just from direct electrostatic interactions $\langle \Delta E_{\text{elec}} \rangle$

with the receptor. A complete picture of the overall electrostatic contribution $\langle \Delta G_{\text{elec,tot}} \rangle$ requires that the interaction energies associated with solvation $\langle \Delta G_{\text{solv}} \rangle$ be included. Both LacNAc ligands (**1** and **2**), display essentially identical favorable net van der Waals interactions $\langle \Delta E_{\text{vdw}} \rangle$ and net entropic penalties $\langle -T\Delta S \rangle$. The weaker interaction with **2** appears to arise primarily from poor intermolecular and solvation electrostatic energies $\langle \Delta G_{\text{elec,tot}} \rangle$.

The much higher affinity predicted for **5** is the result of an interplay of electrostatics, in which a much more favorable intermolecular electrostatic term $\langle \Delta E_{\text{elec}} \rangle$ overcomes a less favorable contribution from the polar component of solvation energy $\langle \Delta G_{\text{pol}} \rangle$. Given the increased chain length of **5**, and the presumed resultant increase in flexibility, it is also probable that the conformational entropy term is underestimated in this analysis. The predicted increase in affinity with increasing level of LacNAc polymerization is consistent with previous observations of galectin-1 binding, which suggested that an increase in polylactosamine chain length, leads to an increase in binding affinity²⁸. In addition, recent measurements of galectin-1 affinities⁷⁴ show a 3-fold increase in binding affinity for Gal- β -(1,4)-GlcNAc- β -(1,3)-Gal- β -(1,4)-Glc, relative to lactose.

Conclusions

This work provides a structural interpretation for observed promiscuity in ligand binding for the CRD of galectin-1. Several conclusions about the specificity of galectin-1 ligand binding emerge from the data presented. The first observation that comes from the analysis of ligand binding to galectin-1 is that there exists a core interaction with LacNAc that is capable of anchoring ligands within the galectin-1 CRD. The

maintenance of geometry for these ligands, regardless of substituents at the 3'OH, indicates that the observed differences in affinities arise primarily from the properties of the interaction between the CRD and these groups linked to the 3'OH.

The examination of electrostatics contains information from the overall picture of the electrostatics of the protein surface, which clearly contains a region of electrostatic complementarity. In addition to this large scale picture, MMGBSA analysis of the MD trajectories gives estimates of binding free energies which correctly rank the binding affinities of the ligands examined, with the exception of Neu5Ac-LacNAc.

It is important to remember that the ligands examined in this work, are found *in vivo* as fragments of larger glycans. For example **1,3** and **4** would be linked to larger glycans structures through their reducing termini. This could obviously affect the conformational properties of these oligosaccharide fragments and their interaction with the galectin-1 CRD. Similarly **5** would often be found as part of a longer polylactosamine chain. It remains to be seen whether galectin-1 is capable of binding to internal LacNAc residues as well as terminal oligosaccharides.

Similarly, even if the MD simulation of the galectin-1 monomer were able to exactly predict the geometry of ligand interaction, there are properties of the dimer that may affect results of the energetic analysis. It is probable that the electrostatics of the dimer differ from that seen in the monomer, a difference that could account for some of the discrepancy seen in the free energy analysis. Another possible source of error seen in the MMGBSA analysis is the fact that, while it has parameterized for AMBER, the set of molecules included in this work did not include ions or carbohydrates. It is possible that the ionic groups found in 3'-O- substituents are not accounted for properly in the

calculations of the solvation free energy. Nevertheless, a correct qualitative ranking of the ligands was achieved.

This work utilizes an initial structural model of a lectin – oligosaccharide complex to attempt to develop a complete picture of the structural and energetic components of the binding of a number of substituted oligosaccharides. The set of anchoring interactions between galectin-1 and LacNAc are maintained for a series of these ligands. The qualitative and quantitative pictures developed from analysis of electrostatics gives further insight into the binding contributions from the 3'-O-substituents. It is also important to note that the predictions derived from the binding geometries of these ligands can be directly tested by NMR analysis of *in vitro* complexes.

References

1. Verlet, L., *Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*. Physics Reviews, 1967. 159: p. 98 - 103.
2. Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A., and Haak, J., *Molecular Dynamics with Coupling to an External Bath*. Journal of Chemical Physics, 1984. 81: p. 3684 - 3696.
3. Hockney, R., *The Potential Calculation and Some Applications*. Methods in Computational Physics, 1970. 9: p. 136 - 211.
4. Garemyr, R. and Elofsson, A., *Study of the Electrostatics Treatment in Molecular Dynamics Simulations*. Proteins: Structure, Function and Genetics, 1999. 37: p. 417 - 428.
5. Schreiber, H. and Steinhauser, O., *Molecular Dynamics Studies of Solvated Polypeptides: Why the Cut-off Scheme Does Not Work*. Chemical Physics, 1992. 168(1): p. 75 - 89.
6. York, D., Darden, T., and Pedersen, L., *The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals: A Comparison of the Ewald and Truncated List Methods*. Journal of Chemical Physics, 1993. 99: p. 8345 - 8348.
7. Ewald, P., *Die Berechnung Optischer und Elektrostatischer Gitterpotentiale*. Annals of Physics, 1921. 64: p. 253 - 287.
8. Leach, A., *Molecular Modeling, Principles and Applications*. 1st ed. 1996, Essex, England: Addison Wesley Longman Ltd.
9. Darden, T., York, D., and Pedersen, L., *Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems*. Journal of Chemical Physics, 1993. 98(12): p. 10089 - 10093.
10. Essmann, U., Perera, L., Berkowitz, M., Darden, T., Lee, H., and Pedersen, L., *A Smooth Particle Mesh Ewald Method*. Journal of Chemical Physics, 1995. 103(19): p. 8577 - 8593.
11. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R., and Klein, M., *Comparison of Simple Potential Functions for Simulating Liquid Water*. Journal of Chemical Physics, 1983. 79(2): p. 926-935.
12. Adams, D., *Alternatives to the Periodic Cube in Computer Simulation*. CCP5 Quarterly, 1983. 10: p. 30 - 36.
13. Guenot, J. and Kollman, P.A., *Molecular Dynamics Studies of a DNA-Binding Protein: 2. An Evaluation of Implicit and Explicit Solvent Models for the Molecular Dynamics Simulation of the Escherchia coli trp Repressor*. Protein Science, 1992. 1(9): p. 1185 - 1205.
14. Sankararamakrishnan, R., Konvicka, K., Mehler, E., and Weinstein, H., *Solvation in Simulated Annealing and High-Temperature Molecular Dynamics of Proteins: A Restrained Water Droplet Model*. International Journal of Quantum Chemistry, 2000. 77: p. 174 - 186.
15. Sitkoff, D., Sharp, K., and Honig, B., *Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models*. Journal of Physical Chemistry, 1994. 98: p. 1978 - 1988.

16. Honig, B. and Nicholls, A., *Classical Electrostatics in Biology and Chemistry*. Science, 1995. 268: p. 1144 - 1149.
17. Diraviyam, K., Stahelin, R.V., Cho, W., and Murray, D., *Computer Modeling of the Membrane Interaction of FYVE Domains*. Journal of Molecular Biology, 2003. 328(3): p. 721-736.
18. Murray, D., McLaughlin, S., and Honig, B., *The Role of Electrostatic Interactions in the Regulation of the Membrane Association of G Protein beta gamma Heterodimers*. J. Biol. Chem., 2001. 276(48): p. 45153-45159.
19. Born, M., Z Physik, 1920. 1: p. 45.
20. Jayaram, B., Sprous, D., and Beveridge, D., *Solvation Free Energy of Macromolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field*. Journal of Physical Chemistry B, 1998. 102: p. 9571 - 9576.
21. Still, W., Tempczyk, A., Hawley, R., and Hendrickson, T., *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics*. Journal of the American Chemical Society, 1990. 112: p. 6127 - 6129.
22. Qiu, D., Shenkin, P., Hollinger, F., and Still, W., *The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii*. Journal of Physical Chemistry A, 1997. 101: p. 3005 - 3014.
23. Onufriev, A., Bashford, D., and Case, D., *Modification of the Generalized Born Model Suitable for Macromolecules*. Journal of Physical Chemistry B, 2000. 104: p. 3712 - 3720.
24. Lemieux, R.U., Bock, K., Delbaere, L.T.J., Koto, S., and Rao, V.S., *The Conformations of Oligosaccharides Related to the ABH and Lewis Human Blood Group Determinants*. Canadian Journal of Chemistry, 1980. 58: p. 631-653.
25. Nagai, K., Roberts, D., Toida, T., Matsumoto, H., Kushi, Y., Handa, S., and Ishizuka, I., *Mono-sulfated Globopentaosylceramide from Human Kidney*. The Journal of Biological Chemistry, 1989. 264(27): p. 16229 - 16237.
26. Roberts, D. and Ginsburg, V., *Sulfated Glycolipids and Cell Adhesion*. Archives of Biochemistry and Biophysics, 1988. 267(2): p. 405 - 415.
27. Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Muller, W., Yagi, F., and Kasai, K., *Oligosaccharide Specificity of Galectins: A Search by Frontal Affinity Chromatography*. Biochimica et Biophysica Acta, General Subjects, 2002. 1572: p. 232 - 254.
28. Merkle, R. and Cummings, R., *Asparagine-Linked Oligosaccharides Containing Poly-N-acetylglucosamine Chains are Preferentially Bound by Immobilized Calf Heart Agglutinin*. Journal of Biological Chemistry, 1988. 263(31): p. 16143-16149.
29. Teichberg, V., Silman, I., Beitsch, D., and Resheff, G., *A β -D-Galactoside Binding Protein from Electric Organ Tissue of Electrophorus electricus*. Proceedings of the National Academy of Sciences USA, 1975. 72: p. 1383 - 1387.

30. Nowak, T., Haywood, O., and Barondes, S., *Developmentally Regulated Lectin in Embryonic Chick Muscle and a Myogenic Cell Line*. Biochemical and Biophysical Research Communications, 1976. 69: p. 650 - 657.
31. Allen, H., Ahmed, H., and Matta, K., *Binding of Synthetic Sulfated Ligands by Human Splenic Galectin-1, a β -Galactoside-Binding Lectin*. Glycoconjugate Journal, 1998. 15: p. 691-695.
32. Ahmed, H., Allen, H., and DiCioccio, R., *Binding of Hydroxylysine-Linked Saccharides by Galaptin, a Galactoside-Binding Animal Tissue Lectin*. Carbohydrate Research, 1991. 213: p. 321 - 324.
33. Ahmed, H., Pohl, J., Fink, N., Strobel, F., and Vasta, G., *The Primary Structure and Carbohydrate Specificity of a β -Galactosyl-Binding Lectin from Toad (*Bufo arenarum* Hensel) Ovary Reveal Closer Similarities to the Mammalian Galectin-1 Than to the Galectin from the Clawed Frog *Xenopus laevis**. Journal of Biological Chemistry, 1996. 271: p. 33083 - 33094.
34. Lopez, M. and Makhatadze, G., *Isothermal Titration Colorimetry*. Methods in Molecular Biology, 2002. 173: p. 121 - 126.
35. Shinohara, Y., Kim, F., Shimizu, M., Goto, M., Tosu, M., and Hasegawa, Y., *Kinetic Measurement of the Interaction Between an Oligosaccharide and Lectins by a Biosensor Based on Surface Plasmon Resonance*. European Journal of Biochemistry, 1994. 223: p. 189 - 194.
36. Yamamoto, K., Ishida, C., Shinohara, Y., Hasegawa, Y., Konami, Y., Osawa, T., and Irimura, T., *Interaction of Immobilized Recombinant Mouse C-type Macrophage Lectin with Glycopeptides and Oligosaccharides*. Biochemistry, 1994. 33: p. 8159 - 8166.
37. Rich, R. and Myszka, D., *Survey of the 1999 Surface Plasmon Resonance Biosensor Literature*. Journal of Molecular Recognition, 2000. 13(6): p. 388 - 407.
38. Dam, T. and Brewer, C., *Thermodynamic Studies of Lectin-Carbohydrate Interactions by Isothermal Titration Calorimetry*. Chemistry Reviews, 2002. 102(2): p. 387 - 429.
39. Gupta, D., Cho, M., Cummings, R.D., and Brewer, C.F., *Thermodynamics of Carbohydrate Binding to Galectin-1 from Chinese Hamster Ovary Cells and Two Mutants. A Comparison with Four Galactose-Specific Plant Lectins*. Biochemistry, 1996. 48(35): p. 15236 - 15243.
40. Schwarz, F., Ahmed, H., Bianchet, M., Amzel, L.M., and Vasta, G., *Thermodynamics of Bovine Spleen Galectin-1 Binding to Disaccharides: Correlation with Structure and its Effects of Oligomerization at the Denaturation Temperature*. Biochemistry, 1998. 37: p. 5867-5877.
41. Welford, K., *Surface Plasmon-Polaritons and Their Uses*. Optical and Quantum Electronics, 1991. 23: p. 1-27.
42. Weimar, T., *Manuscript in preparation*.
43. Thogersen, H., Lemieux, R.U., Bock, K., and Meyer, B., *Further Justification for the Exo-Anomeric Effect. Conformational Analysis Based on Nuclear Magnetic Resonance Spectroscopy of Oligosaccharides*. Canadian Journal of Chemistry, 1982. 60: p. 44-57.

44. Liao, D., Kapadia, G., Ahmed, H., Vasta, G., and Herzberg, O., *Structure of S-lectin, a Developmentally Regulated Vertebrate β -Galactoside-Binding Protein*. Proceedings of the National Academy of Sciences USA, 1994. 91(Feb.): p. 1428-1432.
45. Berman, H.M., Westbrook, J., Feng, Z., Gilliland G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E., *The Protein Data Bank*. Nucleic Acids Research, 2000. 28: p. 235 - 242.
46. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Ferguson, D.M., Radmer, R.J., Seibel, G.L., Singh, U.C., Weiner, P.K., and Kollman, P.A., *AMBER 5.0*. 1997.
47. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*. Journal of the American Chemical Society, 1995. 117: p. 5179 - 5197.
48. Woods, R., Dwek, R., and Edge, C., *Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development*. Journal of Physical Chemistry, 1995. 99(11): p. 3832-3846.
49. Woods, R.J. and Chappelle, R., *Restrained Electrostatic Potential Atomic Partial Charges for Condensed-Phase Simulations of Carbohydrates*. Journal of Molecular Structure: THEOCHEM, 2000. 527(1-3): p. 149-156.
50. Ryckaert, J., Cicottie, G., and Berendsen, H., *Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes*. Journal of Computational Physics, 1977. 23: p. 327 - 341.
51. Bourne, Y., Bolgiano, B., Liao, D., Strecker, G., Cantau, P., Herzberg, O., Feizi, T., and Cambillau, C., *Crosslinking of Mammalian Lectin (Galectin-1) by Complex Biantennary Saccharides*. Structural Biology, 1994. 1(123331): p. 863-870.
52. Rini, J., *X-ray Crystal Structures of Animal Lectins*. Current Opinion in Structural Biology, 1995. 5(3331): p. 617-621.
53. Vyas, N., *Atomic Features of Protein-Carbohydrate Interactions*. Current Opinion in Structural Biology, 1991. 1: p. 732 - 740.
54. Jimenez-Barbero, J., Asensio, J.L., Canada, F.J., and Poveda, A., *Free and Protein-Bound Carbohydrate Structures*. Current Opinion in Structural Biology, 1999. 9: p. 549 - 555.
55. Bush, A., Martin-Pastor, M., and Imberty, A., *Structure and Conformation of Complex Carbohydrates of Glycoproteins, Glycolipids, and Bacterial Polysaccharides*. Annual Reviews in Biophysics and Biomolecular Structure, 1999. 28: p. 269-293.
56. Wolfe, S., Rauk, A., Tel, L.M., and Csizmadia, I.G., *A Theoretical Study of the Edward-Lemieux Effect (the Anomeric Effect). The Stereochemical Requirements of Adjacent Electron Pairs and Polar Bonds*. Journal of the American Chemical Society, 1971: p. 136 - 145.

57. Imberty, A., Mikros, E., Koca, J., Mollicone, R., Oriol, R., and Perez, S., *Computer Simulation of Histo-blood Group Oligosaccharides: Energy Maps of All Constituting Disaccharides and Potential Energy Surfaces of 14 ABH and Lewis Carbohydrate Antigens*. Glycoconjugate Journal, 1995. 12: p. 331 - 349.
58. Ahmed, H., Allen, H.J., Sharma, A., and Matta, K.L., *Human Splenic Galactin: Carbohydrate-Binding Specificity and Characterization of the Combining Site*. Biochemistry, 1990. 29: p. 5315-5319.
59. Nicholls, A., Sharp, K., and Honig, B., *Protein Folding and Association: Insights From the Interfacial and Thermodynamic Properties of Hydrocarbons*. Proteins: Structure, Function and Genetics, 1991. 11: p. 281-296.
60. Misra, V., Hecht, J., Yang, A., and Honig, B., *Electrostatic Contributions to the Binding Free Energy of the lambda cl Repressor to DNA*. Biophysical Journal, 1998. 75(5): p. 2262 - 2273.
61. Sheinerman, F., Norel, R., and Honig, B., *Electrostatic Aspects of Protein-Protein Interactions*. Current Opinion in Structural Biology, 2000. 10: p. 153 - 159.
62. Cheatham, T., Srinivasan, D., Case, D., and Kollman, P.A., *Molecular Dynamics and Continuum Solvent Studies of the Stability of PolyG-PolyC and PolyA-PolyT DNA Duplexes in Solution*. Journal of Biomolecular Structure and Dynamics, 1998. 16: p. 265 - 280.
63. Kuhn, B. and Kollman, P.A., *Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models*. Journal of Medicinal Chemistry, 2000b. 43: p. 3786 - 3791.
64. Kuhn, B. and Kollman, P.A., *A Ligand that is Predicted to Bind Better to Avidin than Biotin: Insights from Computational Alanine Scanning*. Journal of the American Chemical Society, 2000a. 122: p. 3909 - 3916.
65. Lee, M., Duan, Y., and Kollman, P.A., *Use of MM-PB/SA in Estimating the Free Energies of Proteins: Applications to Native, Intermediates and Unfolded Villin Headpiece*. Proteins, 2000. 39: p. 309 - 316.
66. Bryce, R., Hillier, I., and Naismith, J., *Carbohydrate-Protein Recognition: Molecular Dynamics Simulations and Free Energy Analysis of Oligosaccharide Binding to Concanavalin A*. Biophysical Journal, 2001. 81(September 2001): p. 1373 - 1388.
67. Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P., and Case, D., *Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices*. Journal of the American Chemical Society, 1998. 120(37): p. 9401 - 9409.
68. Connolly, M., *Analytical Molecular Surface Calculation*. Journal of Applied Crystallography, 1983. 16: p. 548 - 558.
69. Nguyen, D. and Case, D., *On Finding Stationary States on Large-Molecule Potential Energy Surfaces*. Journal of Physical Chemistry, 1985. 89: p. 4020 - 4026.

70. Levy, R., Karplus, M., Kushick, J., and Perahia, D., *Evaluation of the Configurational Entropy for Proteins: Application to Molecular Dynamics Simulation of an α -Helix*. *Macromolecules*, 1984. 17: p. 1370 - 1374.
71. Karplus, M. and Kushick, J., *Method for Estimating the Configurational Entropy of Macromolecules*. *Macromolecules*, 1981. 14: p. 325 - 332.
72. Lazaridis, T., *Binding Affinity and Specificity from Computational Studies*. *Current Organic Chemistry*, 2002. 6: p. 1319-1332.
73. Archontis, G., Simonson, T., and Karplus, M., *Binding Free Energies and Free Energy Components from Molecular Dynamics and Poisson-Boltzmann Calculations. Application to Amino Acid Recognition by Aspartyl-tRNA Synthetase*. *Journal of Molecular Biology*, 2001. 306: p. 306-327.
74. Baruffi, M., Bochkareva, E., Rogers, K., Stowell, S., Bochkarev, A., and Cummings, R., *Crystal Structures of Human Galectin-1 with Low Molecular Weight Ligands at 1.6-1.9Å, and Analysis of the Ligand Binding Affinity to Galectin-1 Determined by Isothermal Titration Calorimetry*. *Glycobiology*, 2002. 12(10): p. Abstract 177.

Chapter 3

Protein NMR Resonance Assignment and the Residual Dipolar Coupling

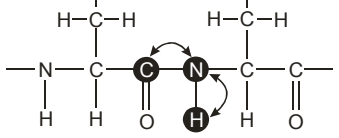
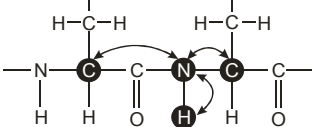
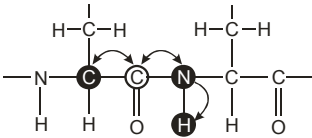
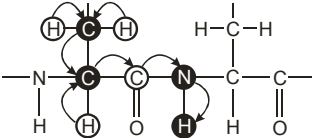
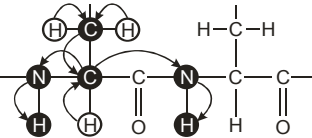
Conventional NMR assignment strategies

The assignment of individual ^1H , ^{13}C and ^{15}N resonances to specific atoms of the protein backbone is a laborious process which typically requires five or more 2D and 3D experiments¹. The goal is to sequentially connect carbon and nitrogen resonances, and then use the characteristic chemical shift ranges of alpha and beta carbons to make amino acid type assignments. The resulting assignments are not only an important step in solving a solution structure by NMR, but are also of great intrinsic value in the examination of protein interactions with ligands, protein complexes and drug design. Increasingly, as the number of structures available continues to grow, we will face the 'inverse assignment problem'. The derivation of assignments given a structure from X-ray crystallography or structure prediction methods would be of great utility for those seeking to use NMR tools on structurally defined protein targets².

2D and 3D NMR spectra used in making assignments

The experiments used in sequential assignment of protein backbone resonances correlate the amide proton, its attached nitrogen, and then the alpha and beta carbons. In addition there are many additional experiments that also correlate carbonyl carbons and side chain carbons³. A schematic diagram of these experiments is shown in Table 3.1.

Table 3.1 Traditional NMR experiments for protein structure determination. (Adapted from Cavanagh et al., 1996 by L. Morris)

Experiment	Correlations observed	Magnetization transfer	J Couplings
HNCO ⁴	$^1H_i^N - ^{15}N_i - ^{13}C_{i-1}'$		$^1J_{NH}, ^1J_{NC'}$
HNCA ⁴	$^1H_i^N - ^{15}N_i - ^{13}C_i^\alpha$ $^1H_i^N - ^{15}N_i - ^{13}C_{i-1}^\alpha$		$^1J_{NH}, ^1J_{NC^\alpha}, ^2J_{NC^\alpha}$
HN(CO)CA ⁵	$^1H_i^N - ^{15}N_i - ^{13}C_{i-1}^\alpha$		$^1J_{NH}, ^1J_{NC'}, ^1J_{C^\alpha C'}$
CBCA(CO)NH ⁶	$^{13}C_i^\beta - ^{13}C_i^\alpha - ^{15}N_{i+1} - ^1H_{i+1}^N$		$^1J_{CH}, ^1J_{C^\alpha C^\beta}, ^1J_{C^\alpha C'}$ $^1J_{NC'}, ^1J_{NH}$
CBCANH ⁷	$^{13}C_i^\beta / ^{13}C_i^\alpha - ^{15}N_i - ^1H_i^N$ $^{13}C_i^\beta / ^{13}C_i^\alpha - ^{15}N_{i+1} - ^1H_{i+1}^N$		$^1J_{CH}, ^1J_{C^\alpha C^\beta}, ^1J_{NC^\alpha}$ $^2J_{NC^\alpha}, ^1J_{NH}$

The HNCA experiment described here gives both the correlations to the I and I-1 C α resonances. The experiment is partnered with the HN(CO)CA experiment in which the NH resonances are correlated to only the I-1 C α resonances. This experiment can determine which of the two resonances belongs to the same residue as the NH pair. Once the N, H, C α (I) and C α (I-1) resonances for the protein are determined, it is possible to begin the assembly of connected fragments.

Obtaining sequential connectivity of backbone resonances

The first goal with this set of data is to correlate C α I-1 chemical shifts from one residue to the C α I chemical shift of another to link sequential residues. This is illustrated in Figure 3.1, which shows these links such as the C α I-1 of peak 77a and the C α I of peak 24b. Similar matching of C β data serves to confirm sequential connections.

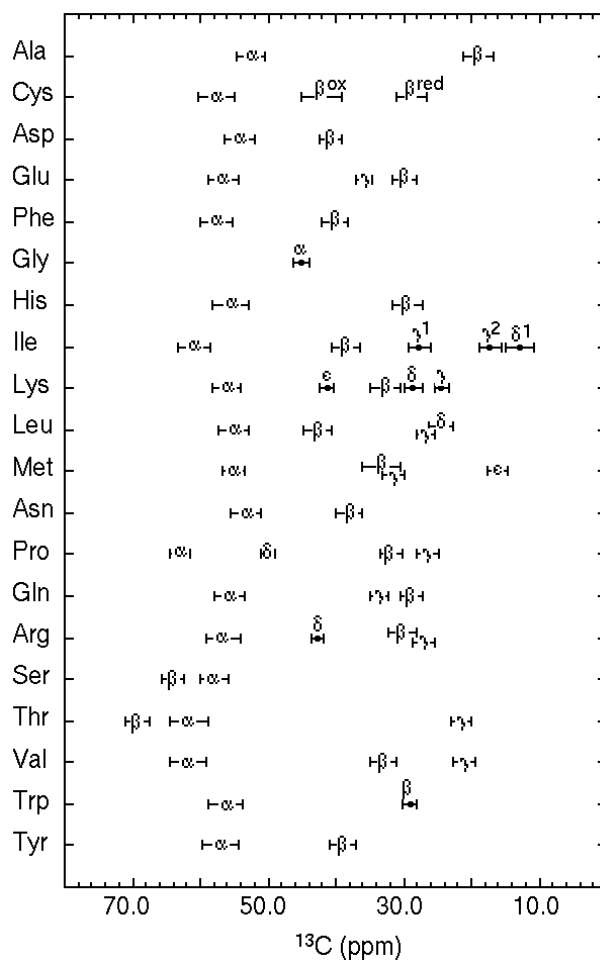
Peak ID	^1H	$^{13}\text{C}\alpha$ (I)	$^{13}\text{C}\alpha$ (I-1)	^{15}N	C β (I)	C β (I-1)
24b	7.619	52.759	55.968	124.581	43.62	41.23
77a	8.579	58.827	52.741	115.185	28.89	43.9
63c	8.58	53.598	58.832	118.807	20.35	28.6
75	7.861	53.676		115.706	45.75	20.03
53	9.461	55.442	53.682	120.152	42.5	45.7

Figure 3.1 Assembled fragment of backbone chemical shifts. Peak ID is an arbitrary identifier of a spin system consisting of the ^1H , ^{15}N , $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ resonances. Values are in parts per million (ppm).

The assembly of long fragments is challenging with incomplete data sets, such as those obtained on larger proteins. This is due to incomplete C α data, which makes fragment assembly more difficult, which in turn complicates the assignment process. Shorter fragments are much more difficult to assign unambiguously, due to the fewer number of diagnostic chemical shift values critical to the process of amino acid type assignment.

Amino acid type assignment

It has been well established that the chemical shifts of C α and C β resonances are diagnostic of certain amino acids⁸. The assignment of assembled fragments to specific stretches of amino acids can be achieved using these characteristic C α and C β shifts, as seen in Figure 3.2.

Figure 3.2 Characteristic chemical shifts of alpha and beta carbons in proteins⁹.

The ability to unambiguously assign a fragment depends on the length of the fragment, as well as the presence of amino acids with unique chemical shifts (i.e. Ala, Gly, Thr, Ser). The resulting possibilities for the fragment from galectin-1 shown previously are seen in Figure 3.3.

The final step in traditional assignment of protein resonances takes the amino acid possibilities and scans them against the primary sequence of the protein. Amino acids such as the alanine seen in this example serve as starting points with which tentative assignments can be evaluated, and result in the assignment of this fragment to

Peak ID	^1H	$^{13}\text{C}\alpha$ (I)	$^{13}\text{C}\alpha$ (I-1)	^{15}N	$\text{C}\beta$ (I)	$\text{C}\beta$ (I-1)	AA Assignment
24b	7.619	52.759	55.968	124.581	43.62	41.23	L,D
77a	8.579	58.827	52.741	115.185	28.89	43.9	C_{red} , E,Q,H,R,W
63c	8.58	53.598	58.832	118.807	20.35	28.6	A
75	7.861	53.676		115.706	45.75	20.03	L,D
53	9.461	55.442	53.682	120.152	42.5	45.7	L,D

Figure 3.3 Assigning possible amino acids to a connected fragment. Note that there are several assignment possibilities for some of the amino acids.

D108:Q109:A110:D111:L112. This laborious process of manual assignment has been aided recently by efforts to automate the assignment process, described in the following section.

Due to the broad range of $\text{C}\alpha$ shifts, and extensive overlap seen for many amino acids, there is a need to improve the resolution of amino acids using additional information. Recently it has been demonstrated that specificity of amino acid identification can be greatly improved by using local fragment geometry (ϕ, ψ) to restrict $\text{C}\alpha$ shift ranges¹⁰. In our case, fragment geometry from known crystal structures greatly facilitates the assignment process.

Computational aids to sequential assignments of proteins

In small proteins (<15 kDa) full assignment of backbone resonances can be obtained with the suite of 3D NMR spectra containing connectivity and additional amino acid type information. In this case, manual assembly of fragments and determination of amino acid type can be rapidly achieved. Collection of complete data sets for larger proteins is often more challenging, due to enhanced spin relaxation and resultant line broadening. This in turn complicates the assignment process which again depends on the assembly of fragments of sufficient length to assign unambiguously. The relaxation

properties of larger proteins can be improved with deuteration of all proton sites except amide protons¹¹. This does, however, reduce protein production.

A number of computational methods which utilize data from a suite of triple resonance experiments have been employed¹. Autoassign¹² utilizes six to eight different 3D spectra, and has been used extensively on small to medium sized proteins (<20 kDa). The large amount of data required makes this program of somewhat limited use, especially for larger proteins, or protein complexes. PACES¹³ is another program which uses exhaustive searching of all possible assignments to assign fragments to the amino acid sequence. Our laboratory has also developed a method which uses $C\alpha$ chemical shifts ($\delta C\alpha$) and torsion angles to determine the most likely assignment for a given fragment of amino acids.

SEASCAPE

A procedure using only connectivity data and chemical shifts from the most robust triple resonance experiments (HNCA and HN(CO)CA) has recently been developed in our laboratory. This probability based method of assigning sequentially linked fragments utilizes $C\alpha$ chemical shifts as well as ϕ and ψ to analyze the likelihood of potential assignment. The program, termed SEASCAPE (SEquential Assignment by Structure and Chemical shift Assisted Probability Estimation)¹⁰ is based upon a series of probability density functions (PDFs) which represent the three dimensional $C\alpha$, ϕ and ψ space for each of the amino acids. A given fragment is then fit along each possible position in the amino acid sequence, and evaluated by using the empirically determined PDF and ϕ , ψ , and $\delta C\alpha$ data. The segment with the largest score (highest probability) can be selected as the most likely location of assignment. The success of this

approach depends on the fragment length, and also having fragments with the correct sequential connectivity. For ten proteins represented in the PDB and BMRB (BioMagnetic Resonance Bank) SEASCAPE was able to place six residue fragments within the sequence with greater than 70% accuracy.

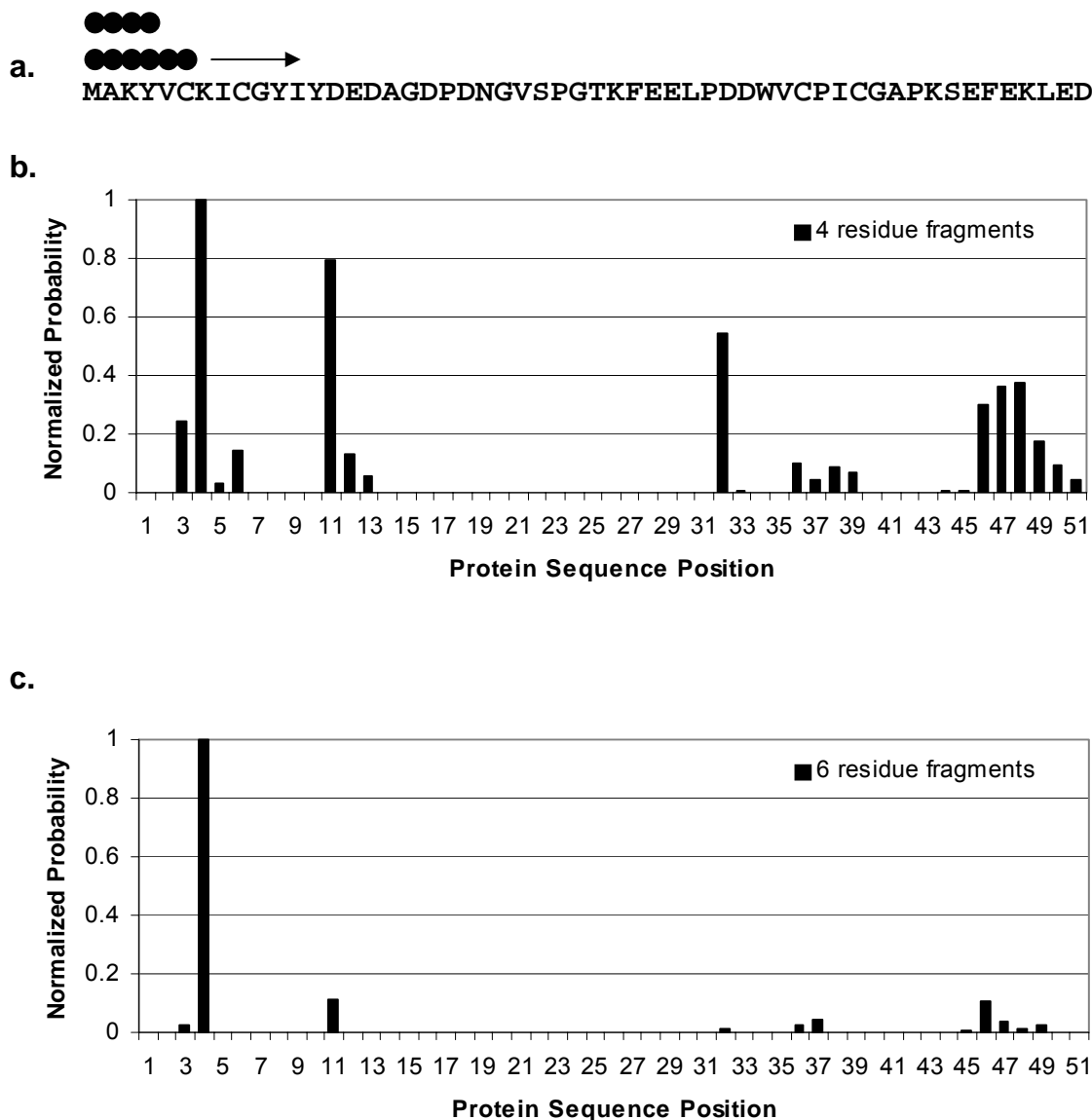


Figure 3.4 Example of assignment of a fragment to a position in a protein sequence. (a.) Sequence of a rubredoxin mutant (1M2Y) from *Pyrococcus furiosus*. Four and six residue fragments (circles) are moved along the sequence as probabilities are calculated for each possible position. Graphs of the normalized probabilities calculated for the 4 residue (b.) and 6 residue (c.) fragments, ICGY and ICGYIY respectively. Figure provided by L. Morris.

Utilization of the assignments of NMR spectra

NMR structure determination

The most common use of sequential protein assignments is in NMR structure determination. In this application assignment of the backbone resonances is combined with data from additional spectra (i.e. TOCSY, NOSY-HSQC) to obtain side chain assignments. Then structural constraints in the form of NOEs and J_3 couplings are obtained in order to iteratively refine the protein structure.

Probing protein interactions with resonance assignments

While assignments are usually obtained as part of the structure determination process, the assignments themselves have considerable value, especially when paired with the structural data from other sources such as X-ray crystallography. This is due to the fact that the chemical shift of specific resonances reports on the chemical environment of that specific residue, and any intermolecular interaction (i.e. ligand binding) which perturbs this environment will report a corresponding change in chemical shift. This fact has been exploited in a number of techniques which use simple 2D spectra, together with resonance assignments to aid in characterizing protein interactions¹⁴⁻¹⁶.

Ligand binding site identification

The most basic experiment utilizing assignment data is the titration of ligand into a protein sample which is monitored by 2D NMR spectra. This type of data is usually acquired as N-H HSQC spectra of protein with increasing concentrations of ligand. For a system with both assignment data and a structural model, these data can be used to

map the binding sites of a protein – ligand complex. An example of this type of experiment is shown in Figure 3.5. The set of peaks, highlighted in orange, at approximately 10.1 ppm (^1H) and 114 ppm (^{15}N) show a clear progressive change during LacNAc titration. These peaks belong to serine 77, a residue known to be in the binding site.

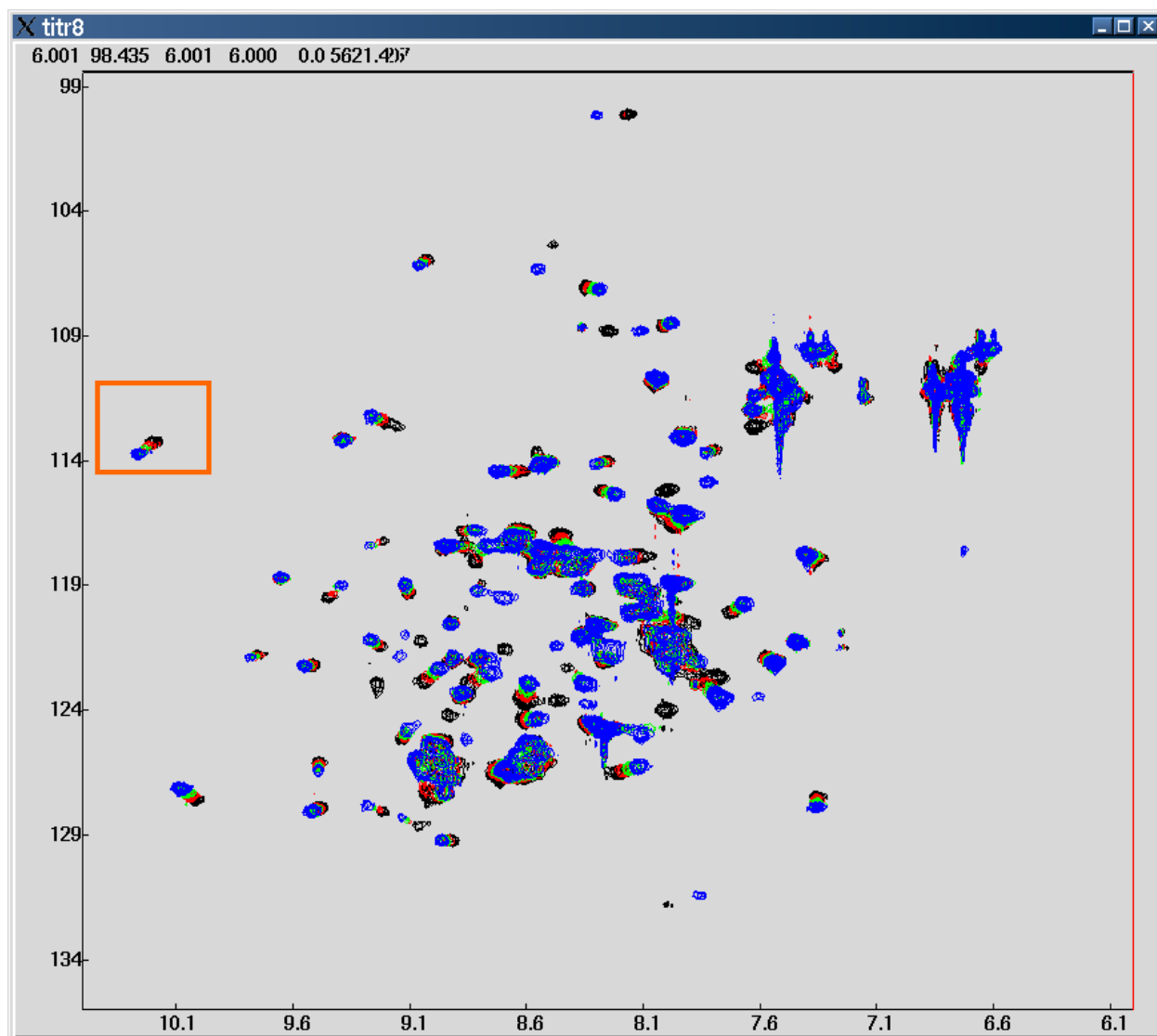


Figure 3.5 HSQC spectra of galectin-1 (C2S) titration with LacNAc. Overlaid spectra correspond to unliganded protein, **0.5**, **1.0**, and **2.0** mM LacNAc.

Protein – protein interaction

The same HSQC experiment used in ligand screening can also serve as a powerful tool for studying protein – protein interactions. The titration of unlabeled protein into a solution of ^{15}N labeled protein can indicate sites of protein – protein interaction¹⁵, which can be utilized as constraints in the building of a model of the structure of the protein – protein complex¹⁴. In principle, this type of experiment could be used to identify the dimer interface in the galectin system.

The residual dipolar coupling

The origin of the dipolar coupling

The residual dipolar coupling (RDC) of covalently bound nuclei is an important piece of data which reports on structural properties of a pair of covalently bound nuclei. The dipolar coupling arises from two magnetic dipoles close to one another in space. The RDC is usually associated with a pair of spin $\frac{1}{2}$ nuclei such as ^{15}N or ^1H . The mathematical description of the RDC is shown in Equation 3.1. The terms $\gamma_i\gamma_j$ indicate the gyromagnetic ratios of the respective nuclei, μ_0 is the permittivity of free space, h is Plank's constant, r_{ij} is the internuclear distance, and θ_{ij} is the angle between the internuclear vector and the external magnetic field. The RDC is measured from splittings of peaks within NMR spectra and is measured in Hz.

$$D_{ij} = -\left(\frac{\mu_0}{4\pi}\right) \frac{\gamma_i\gamma_j h}{2\pi^2 r_{ij}^3} \frac{\langle 3\cos^2 \theta_{ij} - 1 \rangle}{2} \quad \text{Equation 3.1}$$

In the case of covalently bound nuclei, the distance between the attached nuclei r_{ij} , can be treated as constant due to its vibrational time scale. Therefore, the resulting

RDC values can be assumed to be indicators of angular orientation alone. The angular dependence is seen in the term $\langle 3 \cos^2 \theta_{ij} - 1 \rangle$, in which the brackets denote a time averaged quantity. In the normal liquid state, in which molecules are sampling all possible orientations, however, there is no preference for a given orientation, and therefore the RDC averages out to zero.

To restore an observable RDC, an alignment media that spontaneously aligns in the magnetic field, and transfers a small degree of this alignment to the molecule is used. There are many types of alignment media, with the most common being phospholipid bilayers and filamentous phage pF1¹⁷. The most commonly measured dipolar couplings utilized for the analysis of protein structure are the amide NH dipolar couplings. These couplings are easily extracted from heteronuclear spectra, and report on the structure and dynamics of the peptide backbone¹⁸.

Measurement of NH dipolar couplings in proteins

Dipolar couplings are measured using pulse sequences which allow evolution due to the residual dipolar coupling, scalar one bond coupling and the chemical shift. Dipolar couplings add to a splitting normally associated with scalar coupling and must be separated by taking measurements from both aligned and isotropic spectra. In the aligned case, the total coupling between heteronuclei (i.e. N-H) contains both the scalar coupling (J) and dipolar coupling (D). The isotropic spectra contains only the scalar coupling, and by simple subtraction, the residual dipolar coupling is obtained. There are several ways in which these measurements can be obtained, each of which has certain limitations. Most are based upon the Heteronuclear Single Quantum Coherence (HSQC) experiment¹⁹, which correlates a proton with its covalently attached

heteronucleus (usually ^{15}N or ^{13}C). The simplest way to measure the RDC is to use a modified HSQC sequence in which the pulse that normally decouples the scalar coupling in the indirect dimension is removed, resulting in a pair of peaks for each HSQC resonance.

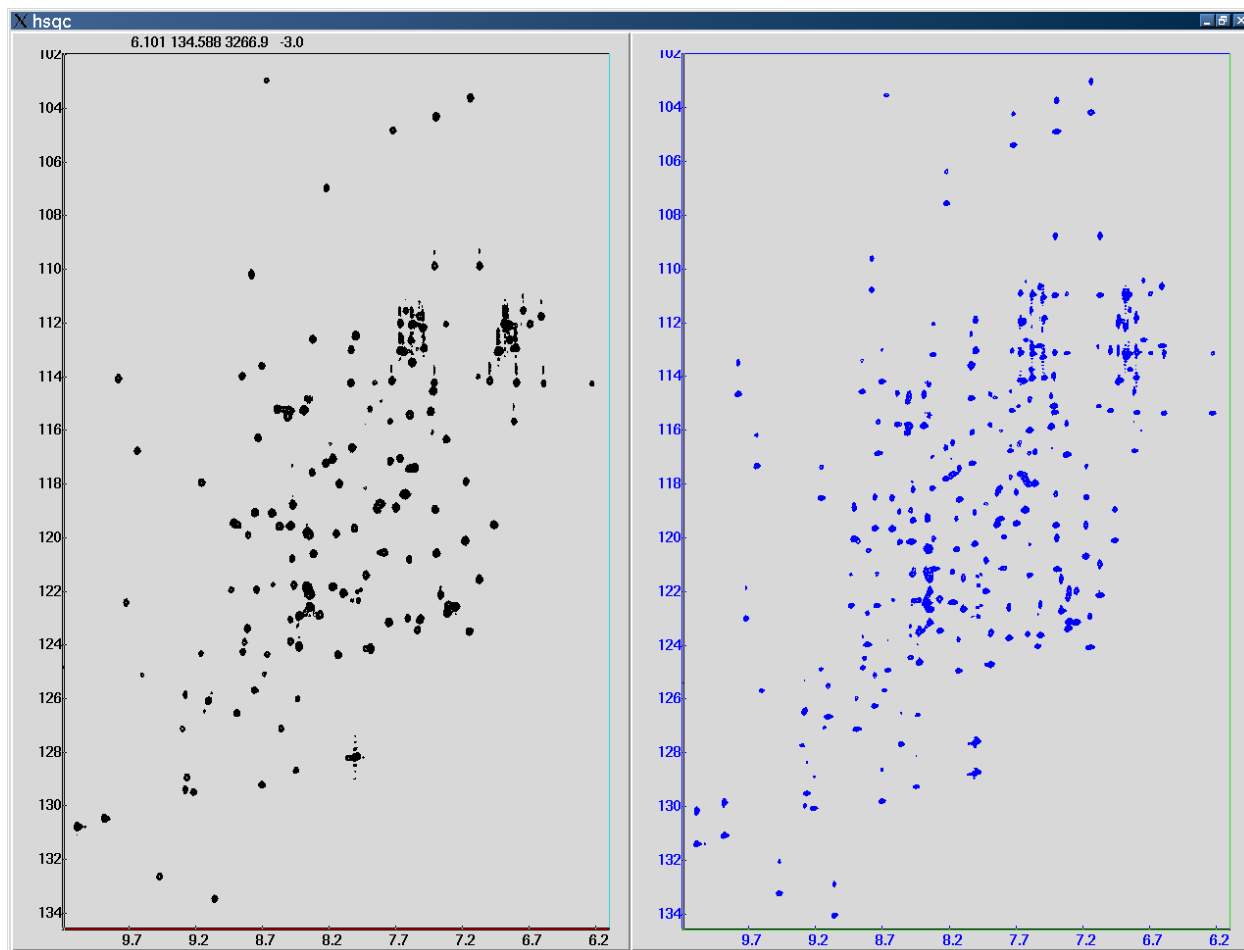
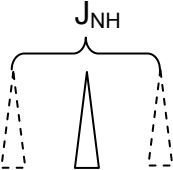


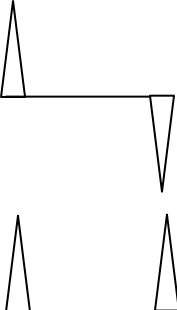
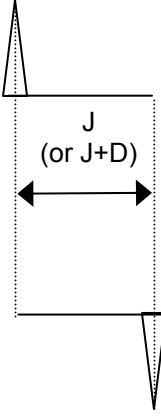


Figure 3.6 A comparison of the HSQC spectra and the corresponding coupled HSQC. Data collected on .5mM ARF1 with 1mM GDP. Data kindly provided by R. Seidel.

For well resolved spectra of small proteins, this coupled HSQC experiment is adequate for measuring NH RDCs to a high level of accuracy. Difficulties arise however, when working with larger proteins which often suffer from a large amount of spectral overlap. In this case, it becomes impossible to distinguish pairs of peaks in

order to make the RDC measurements. There are a number of pulse sequences which overcome this problem by relegating the doublet components into separate spectra. An example of this is seen in the inphase/antiphase (IPAP) experiment²⁰ in which an inphase doublet and an antiphase doublet are added and subtracted to produce two spectra, each containing one of the doublet components. An example of an intensity encoded experiment is the phase-encoded HSQC, which is based on the dependence of the efficiency of the INEPT transfer in the HSQC and the size of the coupling²¹.

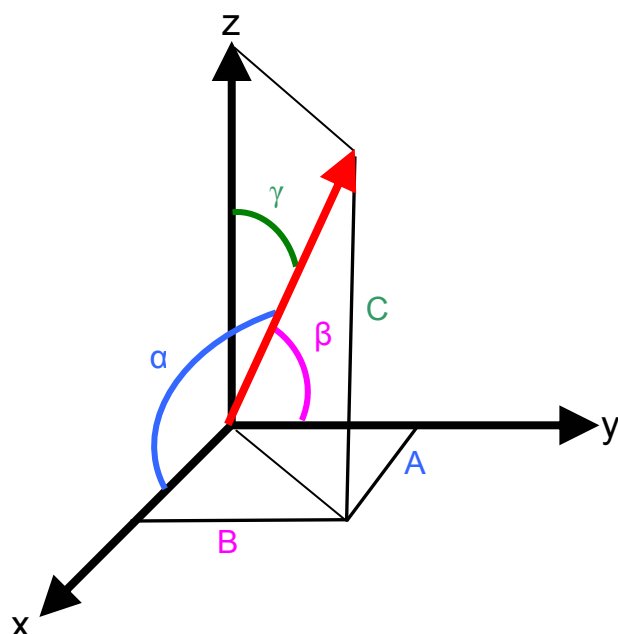
Table 3.2 Comparison of experimental approaches to the measurement of RDCs.

HSQC	
Peaks are refocused in the indirect dimension, giving 1 peak/NH pair 	
Coupled HSQC	IPAP
<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>Isotropic</p> <p>J</p>  </div> <div style="text-align: center;"> <p>Aligned</p> <p>J+D</p>  </div> </div> <p>Subtraction of aligned and isotropic splittings gives the dipolar coupling (D)</p>	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>anti-phase</p>  <p>in phase</p> </div> <div style="text-align: center;"> <p>sum</p>  <p>difference</p> </div> </div>
Simple spectra, not usable in areas of spectral overlap	Usable in overlapped spectra, as corresponding peaks differ in sign

Solution of the order tensor

The utility of RDCs derives from their ability to provide information on the orientation of a molecule or molecular fragment within the magnetic field. The mathematical analysis which extracts this orientational information consists of an algebraic solution to a system of linear equations. While there are several RDCs available for analysis of proteins, the NH RDCs are the focus of this work, due to their ease of measurement, and the fact that they can be measured for larger proteins. Using molecular coordinates of NH pairs, as well as RDC measurements, properties of the molecular alignment with respect to the magnetic field can be calculated. The resulting information consists of parameters which describe both the direction and strength of the alignment. The degree of alignment along three principle alignment axes is described by the variables S_{zz} , S_{xx} , and S_{yy} , with S_{zz} being taken as the direction of highest order. Each represents the average of the term $\left\langle \frac{3\cos^2 \rho_i - 1}{2} \right\rangle$, where ρ_i is the angle from the axis of interest to a single director. This analysis also gives a set of Euler angles (α, β, γ) needed to rotate the molecular frame into the principal alignment frame.

Whereas Equation 3.1 describes the RDC as a function of the angle between the vector and the magnetic field, the RDC can also be written in terms of direction cosines between an internuclear vector and a molecular fragment frame, as well as elements of the order matrix, described above. A general description of the direction cosines can be seen in Figure 3.7, while the mathematical description of the elements of the order matrix is shown in Equation 3.2.



For a vector with length r ,
The direction cosines are:

$$\cos(\theta_x) = A/r$$

$$\cos(\theta_y) = B/r$$

$$\cos(\theta_z) = C/r$$

Figure 3.7 Direction cosines of a vector in a cartesian system.

A set of solutions of Equation 3.2 for a series of dipolar couplings can be used to obtain the elements of the order matrix.

$$D_{ij} = -\left(\frac{D_{\max}}{r_{ij}^3}\right) \sum_{k,l} s_{kl} \cos(\theta_k) \cos(\theta_l) \quad \text{Equation 3.2}$$

The resulting order matrix (Equation 3.3) consists of elements which describe the orientation of the molecular frame with respect to the alignment frame. The matrix is symmetrical, such that $S_{xy} = S_{yx}$. In addition the matrix is traceless, and the combination of these properties results in a complete description of the order matrix with 5 independent variables. Diagonalization of the order matrix results in the principal order parameters and information about Euler angles, as described above.

$$\begin{array}{ccc}
 S_{xx} & & \\
 S_{xy} & S_{yy} & \\
 S_{xz} & S_{yz} & S_{zz}
 \end{array}
 \quad \text{Traceless matrix : } S_{xx} + S_{yy} + S_{zz} = 0$$

Equation 3.3

REDCAT

For this work the program Residual Dipolar Coupling Analysis Tool (REDCAT)²² was used to perform the order tensor analysis. This program utilizes singular value decomposition (SVD) to provide a solution to the system of linear equations described previously (Equation 3.2). SVD has previously been used in this context to analyze RDCs measured for barley lectin²³. REDCAT uses Monte Carlo sampling to sample the solution space that corresponds to estimated error. This combination of Monte Carlo sampling and SVD enables REDCAT to derive error estimates for order parameters and Euler angles returned by the program. The input needed for this analysis consists of the molecular coordinates of the pair of atoms (i.e. N and HN), the measured dipolar coupling, the maximum theoretical dipolar coupling and error estimates for the RDC. A sample set of input for REDCAT is shown in Figure 3.8.

The screenshot shows the 'Main window' of the REDCAT program. It features a menu bar with 'File', 'Edit', 'Tools', and 'Help'. Below the menu is a table with 9 rows of input data. Each row is labeled 'Eq.#' and contains values for X1, Y1, Z1, X2, Y2, Z2, Dipol, and Error. At the bottom of the window, there are input fields for 'Output file name:', 'Number of error space samplings:', 'Number of NULL space sampling:', and 'Search range (in +/- units of error):'. To the right of these fields are 'Run' and 'Quit' buttons.

Eq.#	X1	Y1	Z1	X2	Y2	Z2	Dipol	Error
1)	1.871	1.15	0.39	1.167	1.075	1.11	2.27	2.0
2)	-2.076	10.466	-11.393	-2.692	9.669	-11.477	-2.02	2.0
3)	1.723	17.218	-7.813	1.668	17.896	-8.56	-2.93	2.0
4)	4.333	17.354	-7.009	3.887	18.14	-7.46	-2.59	2.0
5)	-1.998	6.164	-10.717	-1.774	6.938	-10.108	-2.59	2.0
6)	-5.28	6.594	-12.28	-5.756	5.773	-11.935	-5.277	2.0
7)	-4.878	6.241	-15.0	-5.094	5.415	-14.46	-5.02	2.0
8)	14.09	6.899	-8.273	13.713	7.648	-8.836	-0.7	2.0
9)	16.007	6.369	-4.549	15.399	6.243	-5.345	3.48	2.0

Output file name: Results.dat

Number of error space samplings: 10000

Number of NULL space sampling: 10

Search range (in +/- units of error): 1

Run

Quit

Figure 3.8 Input for the REDCAT program. This frame shows some sample input used in the order tensor analysis. Note that all or some of the RDC values can be used for the solution.

Calculation of RDCs from structure and alignment data

The output from the REDCAT program is a series of solutions for the order tensor which are consistent with the RDC data. The angular component of these solutions can be represented in a Sausson-Flaumsted plot, which shows the relationship between the principal alignment frame, and a set of orthogonal axes, in the molecular frame. Figure 3.9 illustrates the orientational relationship between the alignment and molecular frames as described.

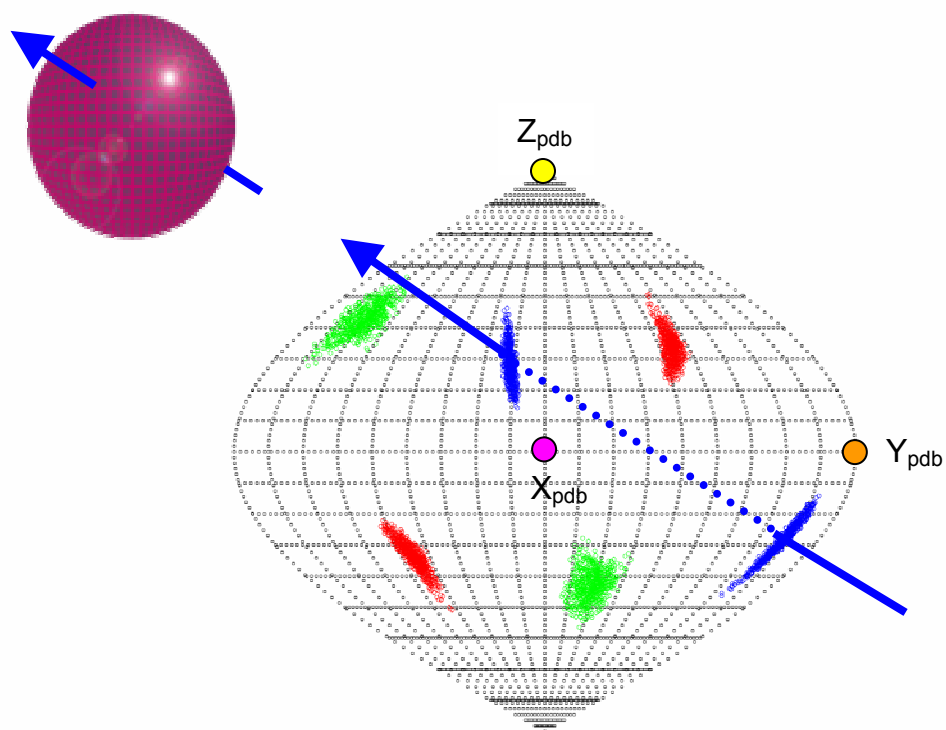


Figure 3.9 Sausson-Flamsteed plot of a set of order tensor solutions derived using REDCAT. Points represent the projection of the order tensor solutions onto a sphere. This plot represents a solution taken from 12 RDCs run with 10000 iterations using REDC

This plot shows the relationship between the molecular frame (PDB frame), and the alignment frame of the molecule. This relationship can be defined as a set of Euler rotations to rotate the molecule into the alignment frame. With the molecule in the alignment frame, it is possible to calculate the theoretical RDC for all of the NH vectors in the protein. These calculated RDCs will be used to validate the similarity of solution

and X-ray structures as well as part of the protocol for resonance assignment. A schematic representation of this is shown in Figure 3.10.

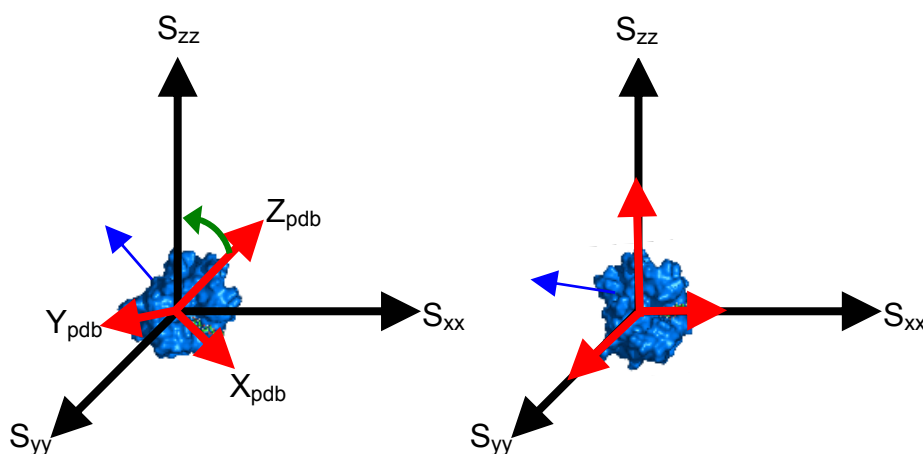


Figure 3.10 Rotation of the PDB into the alignment frame. With the protein in the alignment frame, it is possible to calculate the RDC of NH pair represented by the blue vector.

RDCs as an aid to protein resonance assignment

With the molecular coordinates of the PDB structure in the alignment frame, it is possible to calculate the RDC value for each NH vector in the protein. Even without specific resonance assignment, this set of calculated RDCs can then be compared with a distribution of experimental values, as seen in Figure 3.11.

Assignment of individual amino acids could theoretically be achieved by direct matching of an experimental RDC of an unassigned residue and a corresponding RDC value taken from the REDCAT calculations. The large degree of degeneracy seen in Figure 3.11 would obviously complicate this type of assignment strategy, to an extent which would prohibit most assignments. What is needed is the incorporation of the sequential connectivity information obtained for fragments of protein sequence, as well as the chemical shift information discussed previously.

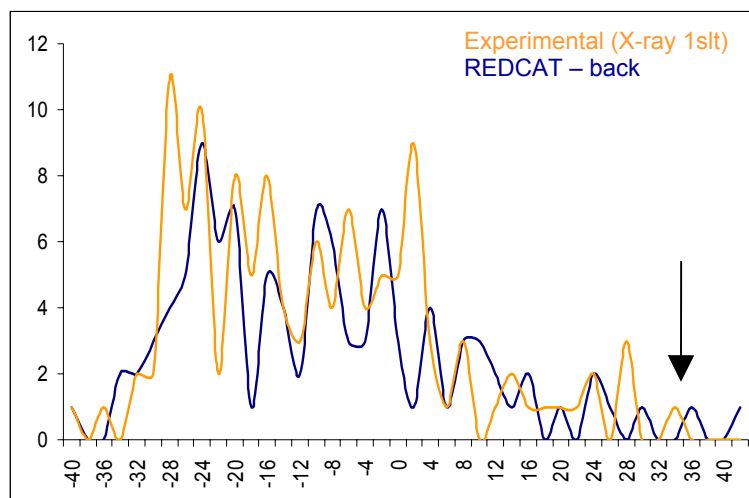


Figure 3.11 Histogram comparing experimental and calculated RDCs for galectin-1. Values on the X axis are NH RDCs measured in Hz, and the Y axis is the number of occurrences of a given RDC. The arrow indicates a site of possible assignment, as indicated by a close match between an experimental RDC, and a back calculated value taken from a known residue in the PDB.

The following chapter describes attempts to integrate the RDC measurements into the assignment process. In this effort, orientational information from the RDC is combined with both traditional NMR resonance assignment strategies based upon 2D and 3D data, as well as computational strategies based upon a probability-based estimation of the assignment of connected fragments.

References

1. Moseley, H.N. and Montelione, G.T., *Automated Analysis of NMR Assignments and Structures for Proteins*. Current Opinion in Structural Biology, 1999. 9(5): p. 635-642.
2. Hus, J., Prompers, J., and Bruschweiler, R., *Assignment Strategy for Proteins with Known Structure*. Journal of Magnetic Resonance, 2002. 157: p. 119 - 123.
3. Edison, A., Abildgaard, F., Westler, W., Mooberry, E., and Markley, J., *Practical Introduction to the Theory and Implementation of Multinuclear, Multidimensional NMR Experiments*. Methods in Enzymology, 1994. 239: p. 3 - 79.
4. Kay, L., Ikura, M., Tschudin, R., and Bax, A., *Three-Dimensional Triple-Resonance Spectroscopy of Isotopically Enriched Proteins*. Journal of Magnetic Resonance, 1990. 89: p. 496 - 514.
5. Bax, A. and Ikura, M., *An Efficient 3D NMR Technique for Correlating the Proton and ^{15}N Backbone Amide Resonances With the alpha-Carbon of the Preceding Residue in Uniformly $^{15}\text{N}/^{13}\text{C}$ Enriched Proteins*. Journal of Biomolecular NMR, 1994. 1: p. 99 - 104.
6. Grzesiek, S. and Bax, A., *Correlating Backbone Amide and Sidechain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR*. Journal of Magnetic Resonance, 1992. 96: p. 432 - 440.
7. Grzesiek, S. and Bax, A., *An Efficient Experiment for Sequential Backbone Assignment of Medium Sized Isotopically Enriched Proteins*. Journal of Magnetic Resonance, 1992. 99: p. 201 - 207.
8. Wishart, D. and Case, D., *Use of Chemical Shifts in Macromolecular Structure Determination*. Methods in Enzymology, 2001. 338: p. 3 - 34.
9. Cavanagh, J., Fairbrother, W., Palmer, A., and Skelton, N., *Protein NMR Spectroscopy: Principles and Practice*. 1996: Academic Press, Inc.
10. Morris, L., Valafar, H., and Prestegard, J., *Assignment of Protein Backbone Resonances Using Connectivity, Torsion Angles and ^{13}C Chemical Shifts*. Journal of Molecular Biology. In press.

11. LeMaster, D. and Richards, F., *NMR Sequential Assignment of Escherichia coli Thioredoxin Utilising Random Fractional Deuteration*. Biochemistry, 1988. 27: p. 142 - 150.
12. Zimmerman, D.E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C.-y., Powers, R., and Montelione, G.T., *Automated Analysis of Protein NMR Assignments Using Methods from Artificial Intelligence*. Journal of Molecular Biology, 1997. 269(4): p. 592-610.
13. Coggins, B. and Zhou, P., *PACES: Protein Sequential Assignment by Computer-Assisted Exhaustive Search*. Journal of Biomolecular NMR, 2003. 26(2): p. 93 - 111.
14. Dominguez, C., Boelens, R., and Bonvin, A., *HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information*. Journal of the American Chemical Society, 2003. 125(7): p. 1731 - 1737.
15. Zuiderweg, E., *Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy*. Biochemistry, 2002. 41: p. 1 - 7.
16. Stockman, B. and Dalvit, C., *NMR Screening Techniques in Drug Discovery and Drug Design*. Progress in Nuclear Magnetic Resonance Spectroscopy, 2002. 41: p. 187 - 231.
17. Prestegard, J.H. and Kishore, A.I., *Partial Alignment of Biomolecules: An Aid to NMR Characterization*. Current Opinion in Chemical Biology, 2001. 5(5): p. 584-590.
18. Prestegard, J., Tolman, J., Al-Hashimi, H., and Andrec, M., *Protein Structure and Dynamics from Field-Induced Residual Dipolar Couplings*, in *Biological Magnetic Resonance: Structure Computation and Dynamics in Protein NMR*, N. Krishna and L. Berliner, Editors. 1999, Kluwer Academic/Plenum Publishers: New York. p. 311 - 355.
19. Bax, A., Ikura, M., Kay, L., Torchia, D., and Tschudin, R., *Comparison of Different Modes of Two-Dimensional Reverse-Correlation NMR for the Study of Proteins*. Journal of Magnetic Resonance, 1990. 86: p. 304 - 318.
20. Ottiger, M., Delaglio, F., and Bax, A., *Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra*. Journal of Magnetic Resonance, 1998. 131: p. 373 - 378.
21. J. R. Tolman, J.H.P., *A Quantitative J-Correlation Experiment for the Accurate Measurement of One-Bond Amide ^{15}N , ^1H Couplings in Proteins*. Journal of Magnetic Resonance - Series B, 1996. 112(3): p. 245-252.

22. Valafar, H. and Prestegard, J., *REDCAT: A Residual Dipolar Coupling Analysis Tool*. Journal of Biomolecular NMR. in press.
23. Losonczi, J.A., Andrec, M., Fischer, M., and Prestegard, J.H., *Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition*. Journal of Magnetic Resonance, 1999. 138(2): p. 334-342.

Chapter 4

A Combined Approach for Protein Resonance Assignment Using Residual Dipolar Couplings, C α Chemical Shifts and ϕ, ψ Values

Abstract

We present here a tool for utilizing residual dipolar couplings (RDCs), in combination with C α chemical shifts and protein structure, to assign resonances in large protein complexes. 2D and 3D data on the 34 kDa dimer of galectin-1, a carbohydrate binding protein, are paired with structural information present in the angular constraints extracted from RDCs. Using a crystal structure of the dimeric protein as a model, orientational information was obtained from analysis of dipolar couplings for a subset of resonances that could be independently assigned to regions of backbone structure. Couplings for remaining resonances were then backcalculated from the model. The resulting RMSD fit between experimentally determined RDCs of unassigned residues and the corresponding back calculated values obtained from order matrix analysis is used as a filter for the assignment of sequentially connected amino acid fragments. This approach is implemented using a variant of the SEASCAPE program, which uses probability density estimation to evaluate the fit of fragments based on C α chemical shift and backbone torsion angles. The resulting assignment of the active site of galectin-1 shows that this approach is a method of general use in the assignment process.

Introduction

The process of assigning protein backbone resonances to specific residues in a protein primary sequence is important to both the determination of protein structure by NMR, and the study of protein – ligand and protein – protein interactions. The suite of double and triple resonance (^{15}N , ^{13}C , ^1H) experiments normally used for the assignment process yields both the sequential connectivities of backbone resonances, and amino acid types. While this process is relatively simple for small proteins (< 100 residues), it remains a major obstacle to high-throughput structure determination for larger proteins. Due to the difficulties inherent in obtaining this assignment information for larger proteins, there is considerable interest in the automation and improvement of

assignment methods¹⁻³. One source of information, which has been used extensively, is the characteristic chemical shifts of alpha and beta carbons of the polypeptide chain. More recently, this information has been coupled with the determination of backbone torsion angles (ϕ, ψ) to assign protein fragments in small proteins⁴.

Additionally, the study of proteins which have been initially characterized structurally, either by structure prediction methods or by crystallography, continues to grow. NMR studies on such systems will be greatly facilitated by a targeted assignment strategy in which an initial structural model can be used to speed up the assignment process. Such tools will aid in drug design, drug screening, protein engineering and studies of protein – ligand interaction⁵.

Partial assignment of the galectin-1 resonances obtained from traditional 2D and 3D experiments was combined with RDC measurements and structural information from the X-ray structure of the galectin-1 dimer in complex with a disaccharide ligand, Gal-B-(1,4)GlcNAc or LacNAc⁶, in order to calculate the order tensor. This information then allowed the calculation of the RDCs for the unassigned residues. Comparison of calculated and experimental values then serves as a filter for assignment of the remainder of the protein. This assignment strategy was implemented using a variant of the SEASCAPE program⁴, developed in our lab. RMSD fit of measured and calculated dipolar coupling values was used in addition to the alpha carbon chemical shift and backbone torsion values to evaluate the placement of connected fragments in the protein primary sequence. Our results show that, given an initial structural model, assignments can be leveraged from an incomplete data set.

Materials and Methods

The protein samples of galectin-1 were obtained from a pQE-11 construct of the C2S mutant of chinese hamster galectin-1⁷ expressed in *E. coli* BL21* (Stratagene). The C2S mutant sequence of hamster galectin-1 was used for NMR studies due to its improved solubility and stability⁸. The expression of ¹³C, ¹⁵N labeled galectin-1 was performed using M9 minimal media supplemented with vitamins⁹ and micronutrients¹⁰ and containing 80% D₂O. HNCACB, HNCA, HNCOCA and HSQC experiments were collected on a sample of 1mM ²H (80%), ¹³C, ¹⁵N labeled galectin-1 with ligand (10mM Gal-β-(1,4)-GlcNAc) added for increased stability. All spectra were collected on an Inova600 equipped with a cryoprobe (Varian Inc.), and the acquisition parameters are listed in Table 4.1. The resulting spectra were processed with NMRPipe¹¹, and analyzed with NMRView¹².

Table 4.1 Acquisition parameters for NMR spectra

Expt.	¹ H points	¹³ C points	¹⁵ N points	No. Scans	Acq. Time
HNCA	4096	96	32	8	36h
HNCACB	1024	70	32	16	44h
HN(CO)CA	1024	64	24	8	16h
HSQC	2048		128	16	1.25h
IPAP	2048		384	32	16h

Alignment of the protein sample was achieved by the addition of filamentous phage (pF1) to a final concentration of 10mg/ml for a sample of 1mM galectin-1. The measurement of N-H dipolar couplings was performed with the IPAP experiment¹³. The resulting spectra were peak picked using pipp¹⁴. Order tensor solutions were obtained with the REDCAT program¹⁵. The input consisted of 25 dipolar couplings values of assigned amino acids of galectin-1, along with

the Cartesian coordinates of the corresponding backbone amide NH groups. The output from this program is a set of solutions to the order tensor consistent with a set of RDC measurement. In addition, REDCAT can provide a 'best' solution, which can be used for back calculations of RDCs for NH vectors for a given protein structure.

Results

Triple resonance data on galectin-1 allows for both the connection of sequential residues, and the identification of amino acid type. An example of portion of the HNCACB spectra of a connected fragment is shown in Figure 4.1.

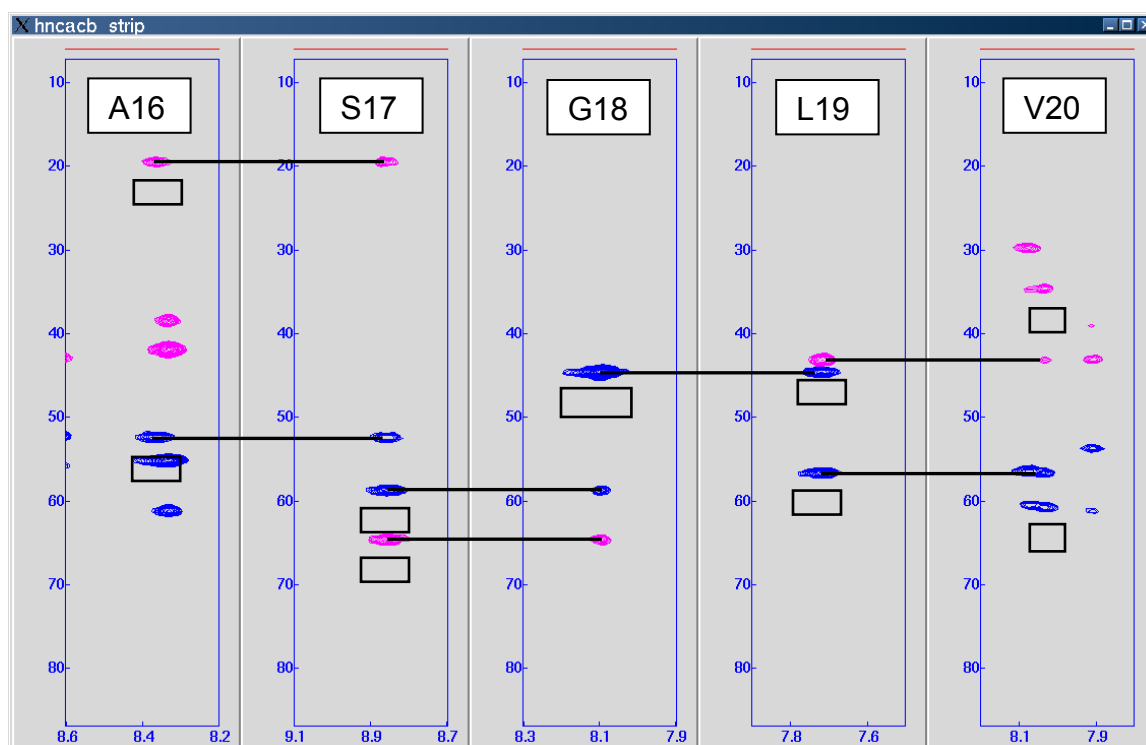


Figure 4.1 Slices of HNCACB spectra for galectin-1 (C2S). Each vertical strip contains the C α and C β residues for an amino acid (residue I) as well as resonances from the preceding residue (I-1). C α residues are blue, and C β residues are magenta. The I residues are boxed.

Partial assignment of galectin-1 by traditional 3D NMR experiments

Partial assignment of galectin-1 was achieved by traditional assignment strategies utilizing the HSQC, HNCA, HN(CO)CA and HNCACB experiments. Given the size of the system it was necessary to use a deuterated (80%) protein sample. It is apparent, however, from a simple representation of the protein assignments, that much of the molecule remains to be assigned, including residues in the binding site. This view of the galectin-1 assignments is shown in Figure 4, with residues highlighted in green representing connected and assigned residues. The 3D spectra of galectin-1 made possible the assignment of 63 out of 146 residues (43% of the protein backbone resonances) to specific positions in the sequence. A table including chemical shift assignments based on these experiments is found later in this chapter.

(His tag) *MRGSHHHHHH***GI****LQ**
MASGLVASNLNLKPGECLKV
 RGEVAPDAKSFV**LN**LKGDSN
 NLCL**H****F****N****P****R****F****N****A****H****G****D****A****N****T****I****V**
CNS**K**DNGT**W**GTE**H**REPAFPF
 QPG**STVEVCITFDQADLT****IK**
LPDGHEFKFPNRLNMEAINY
 M**AADGDFKIKC**VAFE

Figure 4.2 Partial assignment of galectin-1 from 3D NMR data. Highlighted residues show assigned segments of galectin-1. Residues interacting with LacNAc are underlined, and residues comprising a proposed extended binding site are in bold, and the 6x His tag and linker are italicized.

Despite the connectivities and chemical shift information obtained from the NMR spectra, numerous sets of connected resonances could not be assigned. This was due primarily to lack of unique $C\beta$ chemical shifts which aid in the amino acid type assignment. An example of such a fragment is shown in Figure 4.3.

Peak ID	H	$C\alpha$	$C\alpha$ (I-1)	N	$C\beta$	$C\beta$ (I-1)
44	9.2	55.7	57.6	121.9	35.1	30.5
19b	8.2	54.3	55.7	126.1	31.3	
61b	8.4	51.6	54.4	125.2	32.2	

Figure 4.3 Connected fragment of galectin-1 resonances. This fragment could not be assigned from $C\alpha$ and $C\beta$ data alone, due to degeneracy in the chemical shift values. This fragment was assigned to H88, R89, E90 by subsequent analysis.

SEASCAPE

A procedure using only connectivity data and chemical shifts from the most robust triple resonance experiments (HNCA and HN(CO)CA) has recently been developed in our laboratory. This probability based method of assigning sequentially linked fragments utilizes $C\alpha$ chemical shifts as well as ϕ and ψ to analyze the likelihood of potential assignment. The program, termed SEASCAPE (SEquential Assignment by Structure and Chemical shift Assisted Probability Estimation)⁴ is based upon a series of probability density functions (PDFs) which represent the three dimensional $C\alpha$, ϕ and ψ space for each of the amino acids. A given connected, but as yet unassigned fragment is then fit along each possible position in the amino acid sequence, and evaluated by using the empirically determined PDF and ϕ , ψ , and $\delta C\alpha$ data. The segment with the largest score (highest probability) can be selected as the most likely location of assignment. The success of this approach depends on the fragment length, and also having fragments with the correct sequential connectivity. For ten proteins

represented in the PDB and BMRB (BioMagnetic Resonance Bank) SEASCAPE was able to place six residue fragments within the sequence with greater than 70% accuracy. It was also able to suggest several putative assignments for connected resonances with no previous sequential assignment, as seen in Figure 4.4.

MRGSHHHHHHGIQMASGLVASNLNLKPGECKVRGEVAPDAKSFVLNLGKDSNNLCLHF
NPRFNAHGDANTIVCNSK **DNGT**WGTEHREPAFPFQPGSTVEVCITFDQADLTIKLPDGHEFK
FPNRLNME **AINYM**AADGDFKIKCVAFE

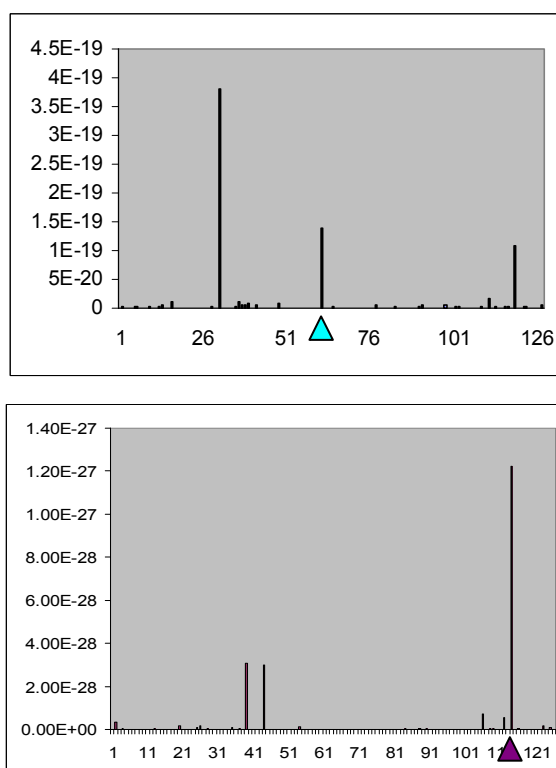


Figure 4.4 Assignments of galectin-1 achieved with SEASCAPE analysis of connected fragments. The correct assignments are highlighted in the amino acid sequence, and correspond to SEASCAPE scores indicated by colored triangles. These segments were assigned solely based upon SEASCAPE analysis and subsequent evaluation of $C\alpha$ and $C\beta$ chemical shifts, due to incomplete or missing RDC data.

As shown here, SEASCAPE does not always give the highest score to the correct position for a given fragment. While $C\alpha$ and $C\beta$ data may be able to

resolve such ambiguities in some cases, it would be valuable to incorporate the RDC data into this assignments strategy.

RMSD fitting of assigned fragments

Chemical shifts are not the only information we have for connected sets of residues. We also have ^{15}N - ^1H RDCs for these resonances, which allow for the solution of the order tensor and subsequent back calculation of RDCs for unassigned NH pairs in the galectin-1 structure. A strategy involving a sequential scanning of RDCs of a connected fragment of residues against the back calculated RDC values of galectin-1 has been devised to utilize the structural information contained in the RDCs for the assignment process. In a method directly analogous to the scanning used in the SEASCAPE program, the RMSD fit of experimental and calculated RDCs to a given position within the protein is evaluated. Regions of low RMSD are candidate positions for assignment, which can be subsequently analyzed for fitness based on $\text{C}\alpha$ and $\text{C}\beta$ chemical shift values. This strategy was first tested on fragments that were previously assigned, using the 3D data on ^{15}N , ^{13}C labeled galectin-1 (HSQC, HNCA, HN(CO)CA and HNCACB). The results of this analysis are shown in Figure 4.5.

MRGSHHHHHHGHILQMASGLVASNLNLKPGECKVKVRGEVAPDAKSFVLNLGKDSNNLCLHFNPR
 FNAHGDANTIVCNSKDNGTWGTEHREPAFPFQPGSTVEVCITFDQADLTIKLPDGHEFKFPNRL
 NMEAINYMAADGDFKIKCVAFE

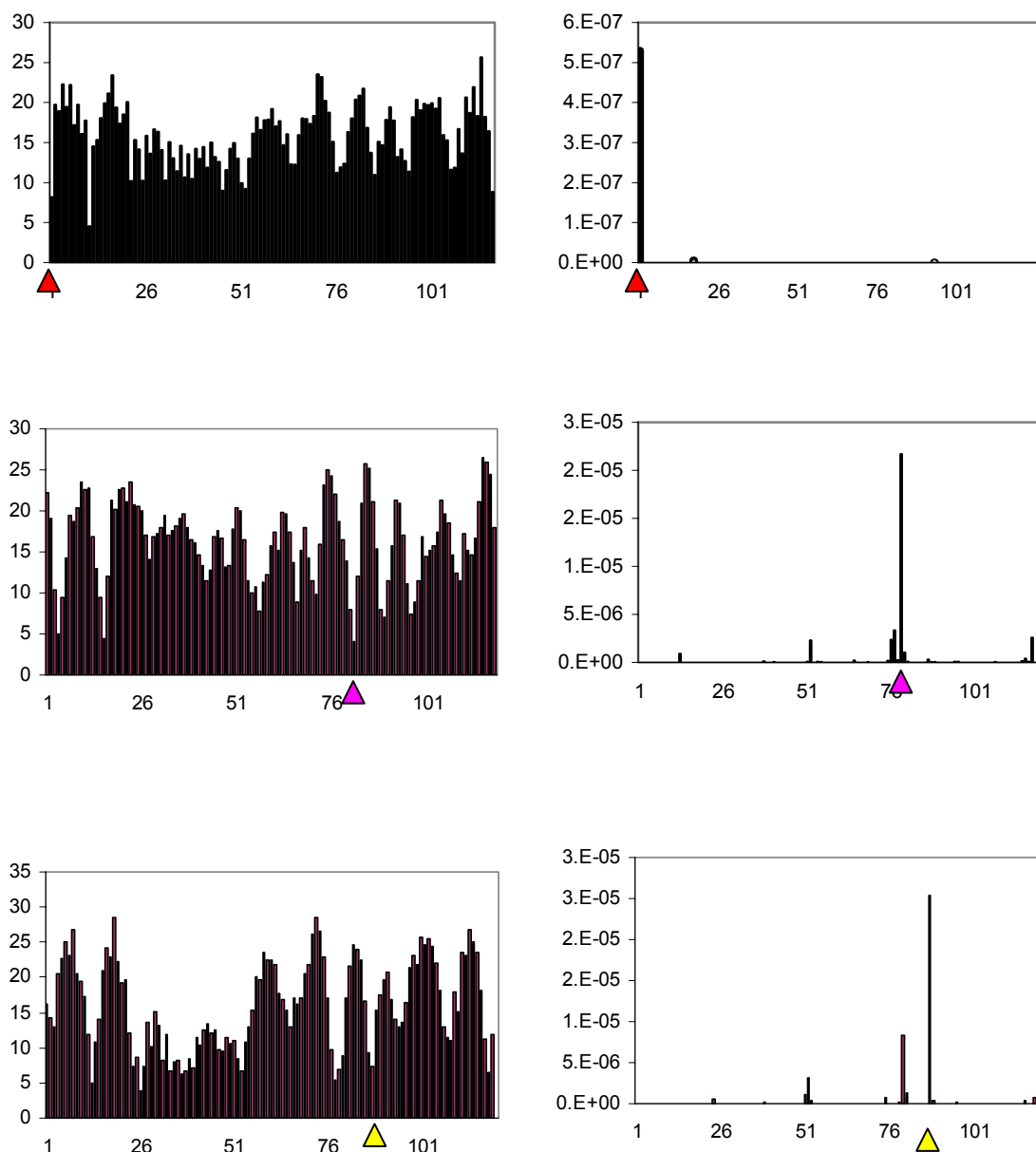


Figure 4.5 RMSD fitting of RDCs for connected fragments of galectin-1 previously assigned from 3D NMR spectra. For fragments of assigned residues, the RMSD fit between experimental and calculated RDCs was evaluated (left), and compared to SEASCAPE analysis (right). The correct assignments are highlighted in the amino acid sequence, and correspond to regions of low RMSD indicated by colored triangles. The PDB starts with G18, and there is no structural model for the His-tag.

It is important to note that the correct assignment was not always the region of lowest RMSD. It is possible however, to evaluate potential assignments by analyzing their fit based upon the chemical shift ranges of C α and C β values for residues in the fragment of interest. Use of RMSD fitting alone is also limited by the availability of fragments for which a complete set of RDCs can be obtained. In order to expand the utility of this assignment filter, RMSD fitting of RDCs can be used in conjunction with SEASCAPE analysis of connected fragments.

REDSEASCAPE, a combined RDC RMSD matching and SEASCAPE approach

Neither RMSD fitting of RDCs nor SEASCAPE analysis gives a definitive result for all connected fragments; it is therefore advantageous to combine the probabilities obtained from the two methods. This integrated approach, termed REDSEASCAPE SEASCAPE (RESidual Dipolar coupling assisted SEquential Assignment by Structure, Chemical shift And Probability Estimation) is described in Figure 4.5, with results shown in Figure 4.6.

$P(1|M)$ - probability evaluation of δC_α , ϕ and ψ for residue 1, given that the residue is methionine

$$P_{\text{seascape}} = P_{(1|M)} * P_{(2|A)} * P_{(3|C)}$$

1 2 3
MACGLVAS.....

$$P_{\text{tot}} = P_{\text{seascape}} * P_{\text{RMSD}}$$

Pseudo probability (P_{RMSD})

$$P_{\text{RMSD}} = e^{(-\text{RMSD}_{\text{frag}})}$$

$$\text{RMSD}_{\text{frag}} = \sqrt{\frac{(\text{RDC}_{\text{Exp}}^{\text{M1}} - \text{RDC}_{\text{Calc}}^{\text{M1}})^2 + (\text{RDC}_{\text{Exp}}^{\text{A2}} - \text{RDC}_{\text{Calc}}^{\text{A2}})^2 + (\text{RDC}_{\text{Exp}}^{\text{C3}} - \text{RDC}_{\text{Calc}}^{\text{C3}})^2}{N_{\text{res}}}}$$

Figure 4.5 Schematic representation of the procedure for assignment using an approach combining SEASCAPE methodology with RDC fitting (REDSEASCAPE).

MRGSHHHHHHGLQMASGLVASNLNLKPGECLKVRGEVAPDAKSFVLNLGKDSNNLCLHF
 NPRFNAHGDANTIVCNSKDNGTWGTEHREPAPFPQPGSTVEVCITFDQADLTIKLPDGHEFK
 FPNRLNMEAINYMAADGDFKIKCVAFE

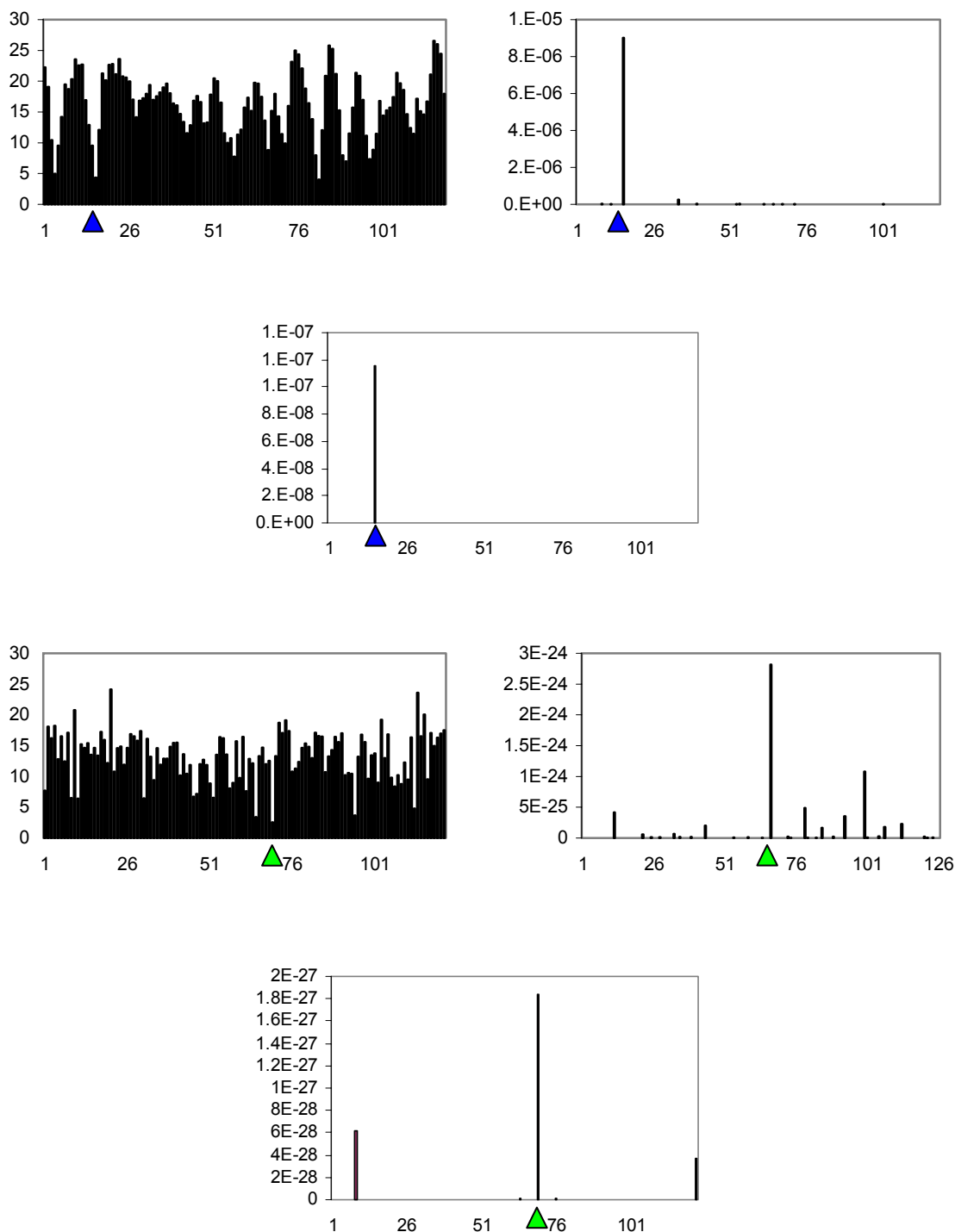


Figure 4.6 REDSEASCAPE results for unassigned connected fragments of galectin-1. The RMSD fitting (left) and SEASCAPE (right) results are shown with REDSEASCAPE pseudo probabilities (center).

The combinatorial approach to the assignment of galectin-1 backbone resonances allowed for the assignment of several additional portions of galectin-1, including several critical residues of the LacNAc binding site. A schematic representation of additional assignments obtained are shown in Figure 4.6, while a full listing of all chemical shift assignments are included in Table 4.2.

(His tag) MRGSHHHHHHGILQ
 MASGLVASNLNLK**PGECLK****V**
 RGEVAPDAKSFVLNLGKDSN
 NLCLHFNPRFNAHGDANTIV
CNS**K****D****N****G****T****W****G****T****E****H****R**EPAPPF
 QPG**S****T****V****E****V****C****I****T****F****D****Q****A****D****L****T****I****K**
L**P****D****G****H**EFKFPNRLNME**A****I****N****Y**
M**A****A****D****G****D****F****K****I****K****C****V****A****F****E**

Figure 4.6. Additional assignments of galectin-1 residues obtained from REDSEASCAPE analysis. Residues assigned from triple resonance data are highlighted in green. Residues assigned from REDSEASCAPE are shown in purple. The residues comprising the primary binding site are underlined, while the residues of the proposed extended binding site are in bold.

The REDSEASCAPE analysis of galectin-1 allowed for the assignment of 23 additional residues, including the binding site residues T83, E87 and R89. The assignments obtained for galectin-1 are shown in Table 4.2.

Table 4.2 Backbone resonances assignments for galectin-1. The sequence of the N-terminally 6x His tagged chinese hamster C2S galectin-1 (CHO) protein used in NMR analysis is shown together with the bovine sequence used to obtain the X-ray structure (pdbid = 1SLT) used for the structural model of the galectin-1 dimer. The PDB of the X-ray structure of bovine galectin-1 begins with glycine 3. The resonances assigned by SEASCAPE analysis and REDSEASCAPE methodology are indicated in the right hand column.

Peak ID	H	Ca	Ca (I-1)	N	Cβ	Cβ (I-1)	CHO	bovine	
							M1		
							R2		
							G3		
							S4		
							H5		
							H6		
							H7		
							H8		
							H9		
							H10		
							G11		
54	8.0	61.0	45.1	120.1	38.7		I12		
20b	8.3	55.0	61.0	125.8	42.2	38.8	L13		
							Q14		
							M15		
20a	8.4	52.4	55.3	125.7	19.8	33.1	A16	A1	
74	8.9	58.8	52.4	115.6	65.0	19.7	S17	C2	
91	8.1	44.5	58.7	111.9		64.9	G18	G3	
48	7.7	56.5	44.5	120.7	43.4		L19	L4	
39a	8.0	60.6	56.5	122.7	34.8	43.4	V20	V5	
4	9.6	50.2	60.6	129.0	21.3	34.4	A21	A6	
85	9.5	56.8	50.2	114.2	66.3	21.6	S22	S7	
78	8.4	53.9	56.8	115.2	37.3	66.3	N23	N8	
80	7.9	57.3	53.9	114.8	42.3	37.3	L24	L9	
77b	8.6	53.3	57.3	115.2	38.5		N25	N10	
50	8.2	55.5	53.3	120.5	43.5	38.5	L26	L11	
63b	8.6	54.3	55.5	118.6	36.2	41.8	K27	K12	
							P28	P13	
87	9.4	44.5	64.0	113.8		31.2	G29	G14	
46	8.2	56.3	44.5	121.3	31.1		E30	E15	
11b	9.0	53.3	56.0	126.9	45.4	31.3	C31	C16	
							L32	L17	
							K33	R18	
							V34	V19	Ca/rdc
14c	9.2	53.3	60.4	127.0	33.9		R35	R20	Ca/rdc
88	9.3	45.8	53.3	113.2		33.4	G36	G21	Ca/rdc
62	8.4	54.1	45.8	118.9	32.1		E37	E22	Ca/rdc
10	9.4	61.8	54.1	128.9	32.0		V38	V23	Ca/rdc
0	7.9	52.0	61.8	132.5	17.9	32.1	A39	A24	

							F95	F79	
							Q96	Q80	
							P97	P81	
90	7.2	46.2	63.2	111.9		31.6	G98	G82	
							S99	S83	
65b	8.7	62.3	57.5	118.3	69.4	66.5	T100	V84	
55	9.2	59.0	62.3	120.1	36.8	69.4	V101	V85	
39c	8.1	54.4	59.0	122.6	33.8	31.9	E102	E86	
32a	9.0	60.2	54.4	123.1	35.3	33.8	V103	V87	
39b	8.1	56.3	60.2	122.8	30.0	35.1	C _r 104	C _r 88	
8	9.6	60.2	56.3	127.6	41.1	29.7	I105	I89	
34	9.2	61.6	60.2	123.1	70.3	41.2	T106	S90	
17	8.9	56.0	61.6	126.3	41.6		F107	F91	
24b	7.6	52.8	56.0	124.6	43.6	41.2	D108	N92	
77a	8.6	58.8	52.7	115.2	28.9	43.9	Q109	Q93	
63c	8.6	53.6	58.8	118.8	20.4	28.6	A110	T94	
75	7.9	53.7		115.7	45.8	20.0	D111	D95	
53	9.5	55.4	53.7	120.2	42.5	45.7	L112	L96	
70b	8.9	62.4	55.5	117.9	69.6	42.8	T113	T97	
1	9.1	60.7	62.4	130.3	38.2	69.3	I114	I98	
15	9.1	54.8	60.7	126.4	34.9	38.6	K115	K99	
21a	9.2	53.2	54.8	125.8	45.3	34.7	L116	L100	
							P117	P101	
84	8.1	53.6	64.6	114.2	40.4	31.9	D118	D102	
99	8.3	44.5	53.7	108.4		40.2	G119	G103	
							H120	Y104	
							E121	E105	
							F122	F106	
							K123	K107	
							F124	F108	
							P125	P109	
73	8.3	52.4	64.1	116.5	39.7	31.2	N126	N110	
							R127	R111	
24a	7.7	53.6	57.8	124.7	33.8	31.7	L128	L112	
65a	8.7	55.5	53.6	118.3	41.8	33.9	N129	N113	
							M130	L114	
							E131	E115	
62	7.5	51.4	56.8	118.8	22.0	30.6	A132	A116	Ca
59	8.7	61.0	51.4	119.4	39.3	22.0	I133	I117	Ca
36	7.9	53.7	61.0	123.1	43.4	39.4	N134	N118	Ca
35a	8.3	56.8	53.7	123.0	42.9		Y135	Y119	Ca
71	8.9	53.4		124.3			M136	M120	Ca
27	8.9	51.1	53.4	124.3	24.4	37.5	A137	A121	
29	8.7	50.0	51.1	123.9	22.9	24.4	A138	XXXX	
69a	9.0	52.8	50.0	118.4	46.0	22.8	D139	G122	
95	8.6	45.5	52.8	107.6			G140	G123	
30b	8.8	55.4	45.4	123.2	45.6		D141	D124	
							F142	F125	

							K143	K126	
100	7.6	62.5	56.3	126.1	39.4	30.7	I144	I127	
26	9.3	55.0	62.7	124.4	33.2	39.5	K145	K128	
60a	8.3	57.9	55.0	119.1	29.7	33.2	C146	C129	
11a	9.0	60.8	57.9	127.2	61.2	29.7	V147	V130	Ca/rdc
7	9.0	50.2	60.8	127.8	23.3	34.6	A148	A131	Ca/rdc
52	8.4	55.6	50.2	120.2	40.1	23.3	F149	K132	Ca/rdc
14a	8.8	56.8	55.6	127.5	32.6	40.2	E150	E133	Ca/rdc

Discussion

A novel application of residual dipolar couplings is illustrated in which the comparison of experimental and calculated RDC values is used as an assignment filter. The results on fragments assigned from 3D spectra show that correct assignments correlate with regions of low RMSD deviation for connected fragments. Neither RMSD analysis of RDC values, nor the SEASCAPE approach based upon C α chemical shifts and structural (ϕ, ψ) data, are able to unambiguously assign all connected fragments. It is therefore, advantageous to combine the results from these two approaches to obtain assignments with the highest level of confidence.

The use of RDCs in the assignment process is subject to several caveats which currently limit their applicability. The use of RMSD as an indicator of possible assignment contains a dependence on fragment length which is not seen in SEASCAPE analysis. In addition, the periodicity seen in the RMSD plots reflects the repeating nature of the beta sheets seen in the galectin-1 X-ray structure. The beta sheet structure of this protein also complicates the RMSD analysis due to the parallel orientation of many of the NH vectors. It is possible that further testing on a larger set of proteins will yield improved results with proteins composed of different secondary structure elements. While RDCs have

previously been proposed as an aid to resonance assignment, it has utilized five or six dipolar couplings per amino acid on the small protein ubiquitin⁵. This approach is of limited applicability, due to the difficulty in measuring these additional couplings (N-C α , H α -HN etc).

It is likely that the probabilities derived from SEASCAPE analysis and RMSD fitting of RDC values differ in their ability to correctly assign a fragment, and it may become necessary to weight the relative probabilities to obtain a set of scores which has the highest possible confidence level. SEASCAPE has been tested on ten different proteins from the BioMag Resonance Bank (BMRB), and criteria for accuracy of assignment have been determined. In contrast, the RMSD fitting on galectin-1 is much more preliminary. Expansion of the scope of data used for RMSD fitting to include BMRB data would allow for the relative comparison of confidence levels of the two methods. This would facilitate better integration of the methods into a single score which better reflects the match of a given fragment to a proposed site of assignment.

Another source of improvement in this assignment strategy would be the incorporation of C β data into the SEASCAPE scoring. C β chemical shifts have a higher correlation with amino acid type than C α values, and the combination of C α and C β data would improve the accuracy of SEASCAPE assignments. Additionally, it is often the case that a given fragment can be evaluated by SEASCAPE analysis, but not by RMSD fitting of RDCs due to gaps in the RDC data. The matching script does not currently have the capability to evaluate fragments which have gaps in the RDC values. While this has limited the scope

of its application, in this case, the implementation of RDC matching for gapped fragments should be relatively easy to implement.

An added difficulty in using RDCs as an assignment tool is the fact that the back calculated RDC values may, for certain residues, be larger than the experimentally derived values. This is due to the fact that motional averaging of NH vectors in solution can lead to a decrease in the apparent RDC for NH vectors *in vitro*. An independent method for detecting motion would allow us to exclude such data.

Another complication arises from the utilization of a solid-state structure as a model for the back calculation of galectin-1 RDCs. While the two monomers seen in the X-ray structure overlay closely for most of the protein backbone, the dimer interface contains residues which differ significantly. These differences do not show up in the NMR spectra, and indicate a difference in the X-ray and solution structures for this region. This makes necessary the elimination of these regions from both the order tensor calculations and the RMSD fitting.

A combinatorial strategy, termed REDSEASCAPE, integrates both the RMSD fitting of RDC data and probability density estimation of assignment by chemical shift ($C\alpha$) and structural data (ψ, ϕ). This approach resolves ambiguities seen in the results of the individual RDC RMSD fitting and SEASCAPE methods. This approach has been applied to assign the residues of the carbohydrate binding protein, galectin-1, including much of the binding site. This approach should serve as a general tool for the assignment of protein backbone resonances with a minimum of NMR spectra.

References

1. Atreya, H.S., Chary, K., and Govil, G., *Automated NMR Assignments of Proteins for High Throughput Structure Determination: TATAPRO II*. Current Science India, 2002. **83**(11): p. 1372 - 1376.
2. Kanelis, V., Forman-Kay, J., and Kay, L., *Multidimensional NMR Methods for Protein Structure Determination*. IUBMB Life, 2001. **52**(6): p. 291 - 302.
3. Moseley, H.N.B., Monleon, D., and Montelione, G.T., *Automatic Determination of Protein Backbone Resonance Assignments from Triple Resonance Nuclear Magnetic Resonance Data*, in *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B*. 2001. p. 91-108.
4. Morris, L., Valafar, H., and Prestegard, J., *Assignment of Protein Backbone Resonances Using Connectivity, Torsion Angles and ¹³C Chemical Shifts*. Journal of Molecular Biology, 2003. In press.
5. Hus, J., Prompers, J., and Bruschweiler, R., *Assignment Strategy for Proteins with Known Structure*. Journal of Magnetic Resonance, 2002. **157**: p. 119 - 123.
6. Liao, D., Kapadia, G., Ahmed, H., Vasta, G., and Herzberg, O., *Structure of S-lectin, a Developmentally Regulated Vertebrate β -Galactoside-Binding Protein*. Proceedings of the National Academy of Sciences USA, 1994. **91**(Feb.): p. 1428-1432.
7. Stang, H., *Expression, Purification, Isotope Enrichment, and Mutagenesis of Galectin-1 (Masters Thesis)*, in *Department of Biochemistry and Molecular Biology*. 1997, University of Georgia: Athens. p. 67.
8. Cho, M. and Cummings, R., *Galectin-1, a β -Galactoside-Binding Lectin in Chinese Hamster Ovary Cells I. Physical and Chemical Characterization*. The Journal of Biological Chemistry, 1995. **270**(10): p. 5198 - 5206.
9. Markley, J., Lu, M., and Bracken, C., *A Method for Efficient Isotopic Labeling of Recombinant Proteins*. Journal of Biomolecular NMR, 2001. **20**: p. 71 - 75.
10. Neidhardt, F.C., Bloch, P.L., and Smith, D.F., *Culture Medium for Enterobacteria*. Journal of Bacteriology, 1974. **119**(3): p. 736 - 747.
11. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A., *NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes*. Journal of Biomolecular NMR, 1995. **6**: p. 277 - 293.
12. Johnson, B. and Blevins, R., *NMRView: A Computer Program for the Visualization and Analysis of NMR Data*. Journal of Biomolecular NMR, 1994. **4**: p. 603 - 614.
13. Ottiger, M., Delaglio, F., and Bax, A., *Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra*. Journal of Magnetic Resonance, 1998. **131**: p. 373 - 378.
14. Garrett, D., Powers, R., Gronenborn, A., and Clore, M., *A Common Sense Approach to Peak Picking Two-, Three- and Four-Dimensional Spectra Using Automatic Computer Analysis of Contour Diagrams*. Journal of Magnetic Resonance, 1991. **95**: p. 214 - 220.
15. Valafar, H. and Prestegard, J., *REDCAT: A Residual Dipolar Coupling Analysis Tool*. Journal of Biomolecular NMR. In press.

Chapter 5

Order Tensor Solutions and Interpretations of Molecular Symmetry of the Galectin-1 Dimer

The RDC and molecular symmetry

The structural information contained within the RDCs has been incorporated into a number of structure refinement protocols. RDCs have been used both in simulated annealing of NMR structures¹ and as restraints in MD calculations². This information can also be used to obtain the relative orientation of the components of a protein complex³. We have attempted to use RDC measurements on galectin-1 to verify a structural model of the dimeric structure. This approach is applicable to the testing of structural models of multimeric protein complexes in general.

Measurement of NH RDCs for galectin-1

The measurement of N-H dipolar couplings was performed with aligned protein samples (10mg/ml pF1, 1mM ¹⁵N labeled galectin-1) that were analyzed using the IPAP experiment⁴, as described previously. The resulting spectra were peak picked using pipp⁵, and the resulting RDC measurements are shown in Table 5.1.

Table 5.1 NH RDCs of galectin-1. The measurements from aligned and isotropic spectra are shown. Asterisks indicate resonances for which no isotropic splittings could be obtained. In these cases a standard value of 93 Hz for the one-bond scalar NH coupling ($^1J_{\text{NH}}$) was used to obtain the RDC.

^1H (ppm)	^{15}N (ppm)	Align (Hz)	Iso (Hz)	RDC (Hz)	
7.825	132.568	77.542	93	-15.458	***
8.957	130.336	72.453	95.852	-23.399	
9.125	129.284	72.018	93.933	-21.915	
9.317	129.012	113.528	93.483	20.045	
7.347	129.012	76.466	92.418	-15.952	
9.495	128.941	73.749	92.311	-18.562	
8.899	127.688	73.678	92.912	-19.234	
9.479	127.428	72.784	93.211	-20.427	
8.706	127.485	86.388	91.992	-5.604	
8.055	127.321	107.158	92.985	14.173	
9.072	127.078	80.247	94.458	-14.211	
8.952	127.062	69.635	95.234	-25.599	
8.571	126.485	86.445	85.396	1.049	
8.988	126.34	79.815	91.652	-11.837	
8.269	125.997	93.001	100.528	-7.527	
9.093	125.813	78.206	100.265	-22.059	
7.766	124.82	97.986	83.534	14.452	
7.593	124.695	86.933	91.499	-4.566	
9.257	124.363	73.704	90.553	-16.849	
8.839	124.156	74.493	90.712	-16.219	
8.317	124.028	83.85	94.444	-10.594	
8.546	123.792	73.629	93.351	-19.722	
8.722	123.379	73.75	91.308	-17.558	
8.938	123.3	83.26	92.23	-8.97	
8.916	123.22	82.78	89.632	-6.852	
8.787	123.252	74.893	91.797	-16.904	
9.532	123.153	72.976	92.89	-19.914	
7.494	123.134	92.639	93.237	-0.598	
7.816	123.032	87.699	88.073	-0.374	
9.783	122.921	82.815	92.1	-9.285	
9.128	122.945	77.82	90.632	-12.812	
8.213	122.608	64.191	99.358	-35.167	
7.979	122.656	75.028	93	-17.972	***
9.265	122.031	71.623	90.557	-18.934	
8.317	121.811	90.298	92.935	-2.637	
8.916	121.657	89.504	92.213	-2.709	
7.646	120.699	106.841	91.881	14.96	
8.753	120.594	81.254	95.734	-14.48	
8.070	120.483	104.612	81.32	23.292	

¹ H (ppm)	¹⁵ N (ppm)	Align (Hz)	Iso (Hz)	RDC (Hz)	
8.800	120.388	71.056	91.494	-20.438	
9.383	120.131	87.286	91.86	-4.574	
8.356	120.113	73.32	90.551	-17.231	
7.962	120.189	92.66	92.376	0.284	
9.103	120	102.675	93.473	9.202	
8.172	119.954	77.695	69.111	8.584	
9.641	119.59	71.496	93	-21.504	***
8.398	119.425	113.436	90.627	22.809	
8.560	119.296	103.036	91.795	11.241	
8.161	119.082	76.746	58.081	18.665	
7.376	118.853	101.759	91.259	10.5	
8.314	118.821	97.127	92.959	4.168	
9.352	118.664	96.36	93.279	3.081	
8.739	118.577	83.038	92.219	-9.181	
8.524	118.624	76.219	90.931	-14.712	
8.933	118.4	84.941	93.588	-8.647	
8.667	118.35	83.94	93.935	-9.995	
7.901	117.233	89.416	93	-3.584	***
8.026	116.826	98.162	95.333	2.829	
8.200	116.512	88.26	90.667	-2.407	
7.780	115.836	91.495	93	-1.505	***
8.797	115.535	81.689	92.674	-10.985	
8.501	115.105	99.134	92.778	6.356	
8.282	115.074	88.936	93.528	-4.592	
10.238	114.895	81.577	89.88	-8.303	
7.829	114.767	93.977	83.592	10.385	
9.411	114.22	73.411	92.868	-19.457	
7.956	114.119	92.796	91.797	0.999	
7.894	114.096	66.302	89.083	-22.781	
9.249	113.106	89.899	97.456	-7.557	
8.000	111.957	84.203	93.671	-9.468	
8.077	109.96	90.382	93.833	-3.451	
8.018	109.81	79.095	93.929	-14.834	
8.257	108.28	96.539	93.578	2.961	
8.496	107.594	100.579	94.213	6.366	
9.032	107.155	76.489	94.078	-17.589	
8.295	101.251	104.527	91.925	12.602	

Solution of the order tensor

As discussed earlier, the molecular coordinates of covalently bonded spin $\frac{1}{2}$ nuclei (i.e. NH pairs), can be combined with corresponding RDC measurements to determine the molecular alignment with respect to the external

magnetic field. The resulting information consists of parameters which describe both the direction and strength of the alignment. The degree of alignment along three principle alignment axes is described by the variables S_{zz} , S_{xx} , and S_{yy} , with S_{zz} being taken as the direction of highest order. The solutions for the order tensors described here were obtained from the REDCAT program⁶.

Utilization of molecular symmetry

The initial structural model central to this work is the X-ray structure of the galectin-1 dimer⁷. This structure served as the source of coordinates for NH pairs needed to calculate the order tensor of the molecule. One important property of the galectin-1 dimer is the presence of a two-fold pseudo C_2 axis of symmetry, shown in Figure 5.1. As both chemical shifts and RDCs are insensitive to rotation by 180° about such an axis, a single set of resonances and a single RDC is observed for symmetry related pairs of atoms in this structure.

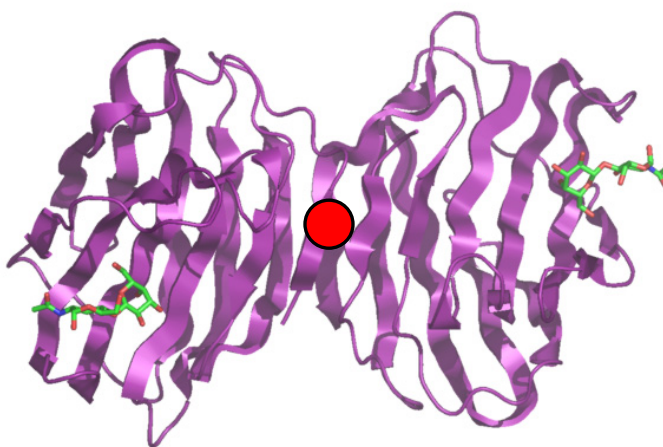


Figure 5.1 Structure of the galectin-1 dimer highlighting the C_2 axis of symmetry, shown in red.

A similar symmetry property has been exploited to aid in the orientational analysis of a multimeric protein complex of mannose binding protein (MBP)⁸. This work also showed that at least one of the alignment axes of the principle alignment frame must coincide with a three fold rotation axis. In the case of galectin-1, the coincidence of axes of the alignment frame with the pseudo 2-fold axis of symmetry seen in the X-ray structure can be used to assess the accuracy of the dimer model.

Order tensor analysis was performed using one monomer taken from the galectin-1 PDB (1SLT), with the axis of symmetry aligned with the X axis of the PDB frame. 25 RDCs were used to obtain 9832 order tensor solutions from 10000 iterations of the REDCAT algorithm, with a constant error of 4 Hz. These solutions are shown in Figure 5.2 as directions for alignment axes in the Sausson-Flamsteed plot.

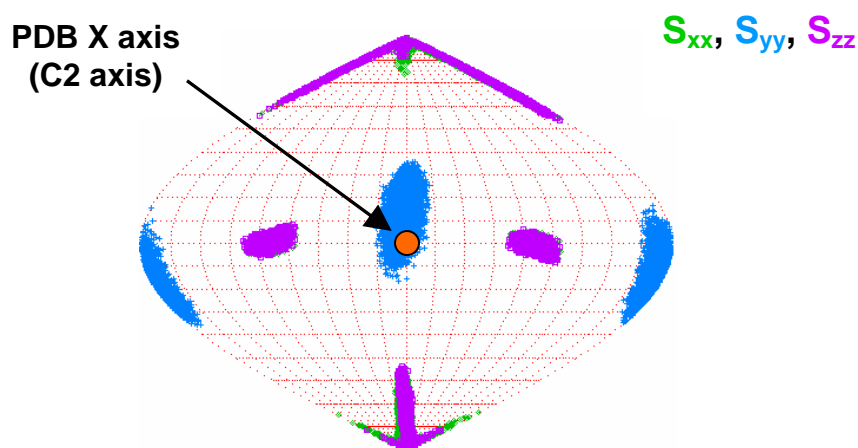


Figure 5.2 Sausson-Flamsteed plot of order tensor solutions for galectin-1 showing the coincidence of the C2 axis of pseudo symmetry and S_{yy} .

The coincidence of the C2 axis of pseudo symmetry with the S_{yy} element of the order tensor indicates that the symmetry observed in the crystal structure is maintained in solution. In addition to serving as an indicator of the validity of a set of order tensor solutions, the dimeric nature of galectin-1 also serves to multiply the information content of a given RDC. The angular information derived from each N-H RDC will be represented twice in the dimer, resulting in an effective doubling of the information contained in each measurement. The effect of utilizing the dimer for this analysis is seen in Sausson-Flaamsted plots of the resulting order tensor solutions (Figure 5.).

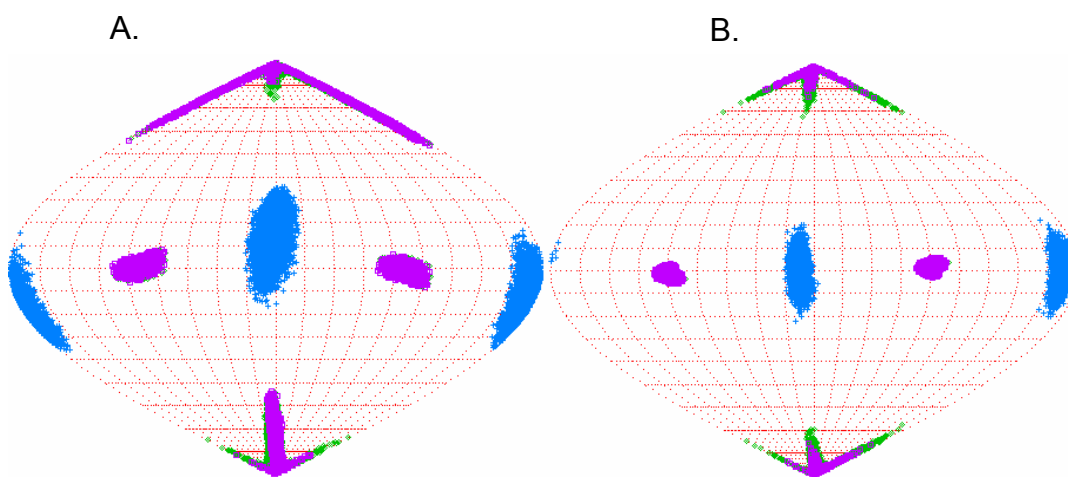


Figure 5.3 Sausson-Flaamsted plots of order tensor solutions from REDCAT showing the comparison of treating galectin-1 as a monomer (A) and as a dimer (B). In panel A, each RDC was included in the order tensor calculation once. For the solution shown in panel B, the order tensor solution was derived using the dimer structure.

It is clear that the use of the dimer results in an improved definition of order tensor axes in panel B. The order parameters obtained from the best solution, as calculated by REDCAT, were as follows: $S_{xx} = 7.36 \times 10^{-4}$, $S_{yy} =$

9.39×10^{-5} , $S_{zz} = -8.3 \times 10^{-4}$ $\alpha = 82.7$ $\beta = 88.7$, $\gamma = 180.5$ $\eta = -0.774$. The solutions were derived using 25 RDCs per monomer, and 10000 iterations of the REDCAT algorithm. These values were calculated with a constant 4 Hz error. The order tensor has an η value near 1, as indicated by the similar magnitude in the solutions for S_{zz} and S_{xx} . This similarity leads to some ambiguity in choice of S_{zz} or S_{xx} as the most ordered direction, as seen in the overlap of axes plotted in Figure 5.3.

This analysis has served to validate the model used for calculation of RDC values. The dimeric nature of the galectin-1 dimer in the solution state has important implications for order tensor calculations which serve an indispensable role in the back calculation of RDCs from the structure.

References

1. Tjandra, N., Omichinski, J., Gronenborn, A., Clore, M., and Bax, A., *Use of Dipolar ^1H - ^{15}N and ^1H - ^{13}C Couplings in the Structure Determination of Magnetically Oriented Macromolecules in Solution*. *Nature Structural Biology*, 1997. **4**: p. 732 - 738.
2. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Ferguson, D.M., Radmer, R.J., Seibel, G.L., Singh, U.C., Weiner, P.K., and Kollman, P.A., *AMBER 5.0*. 1997.
3. Fischer, M., Losonczi, J., Weaver, J., and Prestegard, J., *Domain Orientation and Dynamics in Multidomain Proteins from Residual Dipolar Couplings*. *Biochemistry*, 1999. **38**(July 13): p. 9013 - 9022.
4. Ottiger, M., Delaglio, F., and Bax, A., *Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra*. *Journal of Magnetic Resonance*, 1998. **131**: p. 373 - 378.
5. Garrett, D., Powers, R., Gronenborn, A., and Clore, M., *A Common Sense Approach to Peak Picking Two-, Three- and Four-Dimensional Spectra Using Automatic Computer Analysis of Contour Diagrams*. *Journal of Magnetic Resonance*, 1991. **95**: p. 214 - 220.
6. Valafar, H. and Prestegard, J., *REDCAT: A Residual Dipolar Coupling Analysis Tool*. *Journal of Biomolecular NMR*. In press.
7. Liao, D., Kapadia, G., Ahmed, H., Vasta, G., and Herzberg, O., *Structure of S-lectin, a Developmentally Regulated Vertebrate β -Galactoside-Binding Protein*. *Proceedings of the National Academy of Sciences USA*, 1994. **91**(Feb.): p. 1428-1432.
8. Al-Hashimi, H.M., Bolon, P., and Prestegard, J., *Molecular Symmetry as an Aid to Geometry Determination in Ligand Protein Complexes*. *Journal of Magnetic Resonance*, 2000. **142**(1): p. 153 - 158.

Chapter 6

Mapping of the Galectin-1 Binding Site and Conclusions

Mapping the galectin-1 binding site

With the set of partial assignments obtained from 3D NMR data as well as SEASCAPE and RDC RMS matching, it is possible to use ligand titrations to map the ligand binding site. This was achieved by collection of HSQC spectra of 0.8 mM galectin-1 (C2S) with increasing concentrations of LacNAc and Neu5Ac-LacNAc (0, 0.5, 1, 2, 5, 10mM).

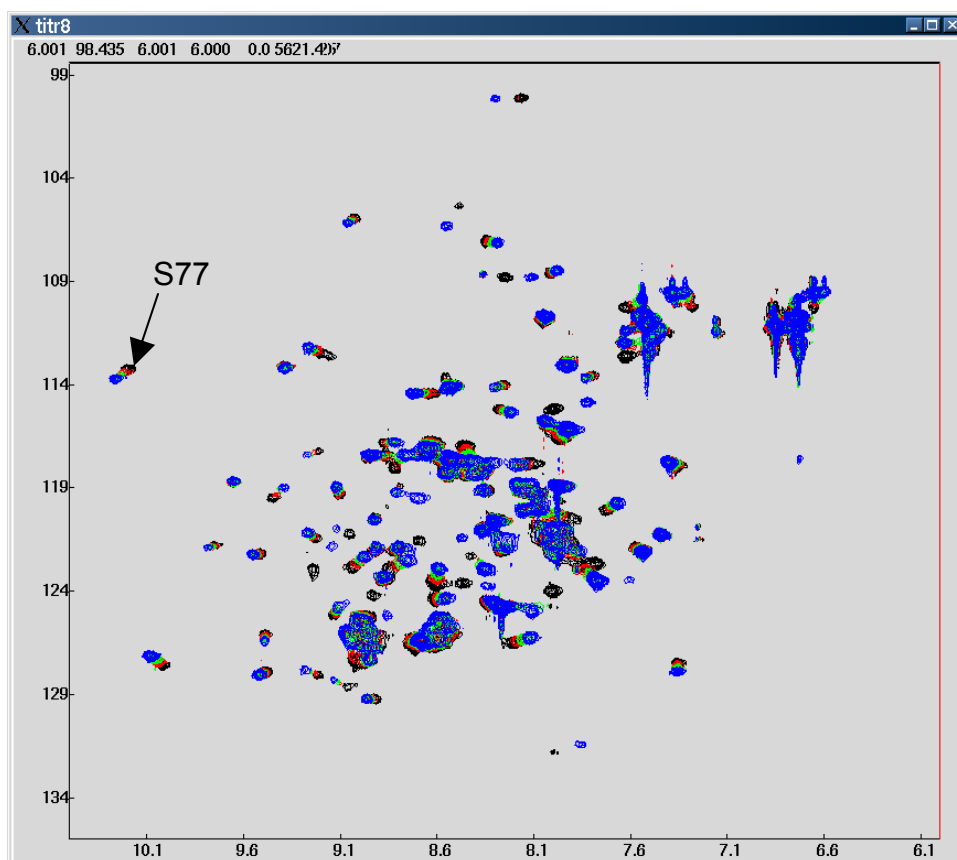


Figure 5.1 Ligand titration of LacNAc (0, 0.5, 2 and 5mM) into galectin-1.

While it is possible in many cases to illustrate the chemical shift changes by labeling of HSQC spectra, the resulting illustration does not aid in the structural visualization or quantification of the data. Therefore, it is more illustrative to quantify the chemical shift changes with and without ligand, and visualize the results using the X-ray structure. The chemical shift differences between the protein without sugar, and the titration end points were quantified using Equation 5.1. This equation takes the differences in chemical shifts and weights them according to the ratio of the gyromagnetic ratios.

$$\delta^1\text{H}^{15}\text{N} = \sqrt{(\delta^1\text{H})^2 + (0.1(\delta^{15}\text{N}))^2} \quad \text{Equation 5.1}$$

This analysis has been used to compare ligand – induced chemical shift changes in the binding site of galectin-1. It was predicted from MD trajectories that Neu5Ac- α -(2,3)-LacNAc should interact with several residues in a proposed extended binding site. This prediction was tested by collecting HSQC spectra for LacNAc and Neu5Ac- α -(2,3)-LacNAc titrations. The quantification of the resulting $\delta^1\text{H}^{15}\text{N}$ values are shown in Table 5.1.

Table 5.1 Quantification of ligand – induced chemical shift changes. $\delta^1\text{H}^{15}\text{N}$ values were measured as changes in peak position between 0 and 10mM ligand.

Pk. ID	^1H	^{15}N	LacNAc	Neu5Ac-LacNAc
1	10.35	115.10	0.11	0.05
2	10.12	114.14	0.10	0.04
3	9.93	126.99	0.02	0.01
4	9.90	121.88	0.04	0.03
5	9.86	123.14	0.05	0.01
6	9.74	119.77	0.02	0.01
7	9.65	123.93	0.04	0.02
8	9.65	118.66	0.07	0.08
9	9.64	121.10	0.10	0.09
10	9.62	125.71	0.01	0.01
11	9.60	129.25	0.10	0.08
12	9.58	123.85	0.01	0.00

Pk. ID	¹ H	¹⁵ N	LacNAc	Neu5Ac- LacNAc
13	9.52	114.43	0.01	0.01
14	9.47	126.50	0.03	0.02
15	9.46	124.08	0.11	0.10
16	9.45	122.59	0.04	0.04
17	9.46	123.06	0.07	0.06
18	9.43	113.29	0.07	0.05
19	9.43	124.99	0.04	0.01
20	9.43	118.69	0.10	0.08
21	9.39	128.09	0.07	0.04
22	9.38	122.43	0.04	0.04
23	9.37	129.01	0.07	0.07
24	9.35	128.76	0.05	0.04
25	9.27	121.61	0.04	0.03
26	9.26	126.65	0.02	0.00
27	9.27	125.34	0.03	0.03
28	9.20	120.29	0.04	0.02
29	9.19	128.85	0.06	0.05
30	9.17	127.35	0.04	0.00
31	9.16	130.00	0.01	0.02
32	9.15	118.18	0.03	0.03
33	9.13	126.66	0.01	0.01
34	9.12	107.40	0.05	0.02
35	9.12	127.81	0.04	0.03
36	9.12	123.08	0.06	0.07
37	9.09	124.55	0.04	0.02
38	9.07	118.87	0.04	0.05
39	9.07	123.68	0.08	0.06
40	9.04	125.89	0.02	0.01
41	9.03	121.81	0.06	0.04
42	9.02	127.93	0.04	0.03
43	9.01	128.41	0.01	0.01
44	8.99	130.11	0.06	0.04
45	8.99	118.43	0.06	0.07
46	8.98	120.87	0.04	0.01
47	8.96	124.41	0.03	0.02
48	8.92	121.83	0.01	0.01
49	8.84	119.06	0.01	0.01
50	8.84	124.29	0.05	0.01
51	8.82	120.57	0.35	0.02
52	8.77	127.77	0.07	0.04
53	8.76	121.83	0.04	0.04
54	8.74	118.19	0.02	0.02
55	8.71	120.80	0.02	0.01
56	8.68	124.23	0.07	0.04
57	8.63	119.70	0.04	0.03
58	8.62	107.60	0.03	0.03
59	8.61	125.58	0.06	0.08

Pk. ID	¹ H	¹⁵ N	LacNAc	Neu5Ac- LacNAc
60	8.61	118.69	0.02	0.02
61	8.58	120.75	0.05	0.00
62	8.58	115.03	0.03	0.03
63	8.54	119.56	0.38	0.13
64	8.52	119.20	0.09	0.05
65	8.43	114.80	0.04	0.01
66	8.42	101.51	0.19	0.22
67	8.39	121.92	0.08	0.08
68	8.36	126.13	0.02	0.01
69	8.36	115.24	0.05	0.04
70	8.36	123.15	0.02	0.02
71	8.34	119.86	0.03	0.02
72	8.33	116.58	0.08	0.02
73	8.32	108.06	0.08	0.06
74	8.27	120.30	0.03	0.02
75	8.24	117.87	0.05	0.03
76	8.21	120.05	0.04	0.02
77	8.21	110.17	0.14	0.02
78	8.18	126.35	0.08	0.10
79	8.15	121.31	0.07	0.09
80	8.14	114.60	0.05	0.05
81	8.13	117.27	0.05	0.04
82	8.12	112.06	0.04	0.02
83	8.08	109.84	0.04	0.02
84	8.08	106.57	0.06	0.06
85	8.06	120.23	0.01	0.01
86	8.06	117.37	0.05	0.04
87	8.05	123.17	0.02	0.01
88	8.01	121.15	0.02	0.02
89	8.01	114.37	0.02	0.01
90	8.00	117.45	0.07	0.07
91	7.98	124.29	0.01	0.01
92	7.93	122.02	0.03	0.01
93	7.93	115.19	0.04	0.01
94	7.92	123.56	0.07	0.04
95	7.80	123.17	0.01	0.01
96	7.79	125.23	0.16	0.15
97	7.69	124.48	0.13	0.13
98	7.69	120.98	0.01	0.01
99	7.62	123.44	0.03	0.02
100	7.53	122.61	0.01	0.01
101	7.50	119.19	0.05	0.03
102	7.44	129.37	0.03	0.12
103	7.43	116.22	0.04	0.02
104	7.36	123.05	0.04	0.02
105	7.34	128.76	0.07	---
106	6.85	118.89	0.10	0.07
107	6.21	115.27	0.14	0.11

In order to visualize these data in a structural context, the chemical shifts of assigned residues were binned according to their $\delta^1\text{H}^{15}\text{N}$ values and mapped onto the structure of the protein – ligand complex, as seen in Figure 5.2.

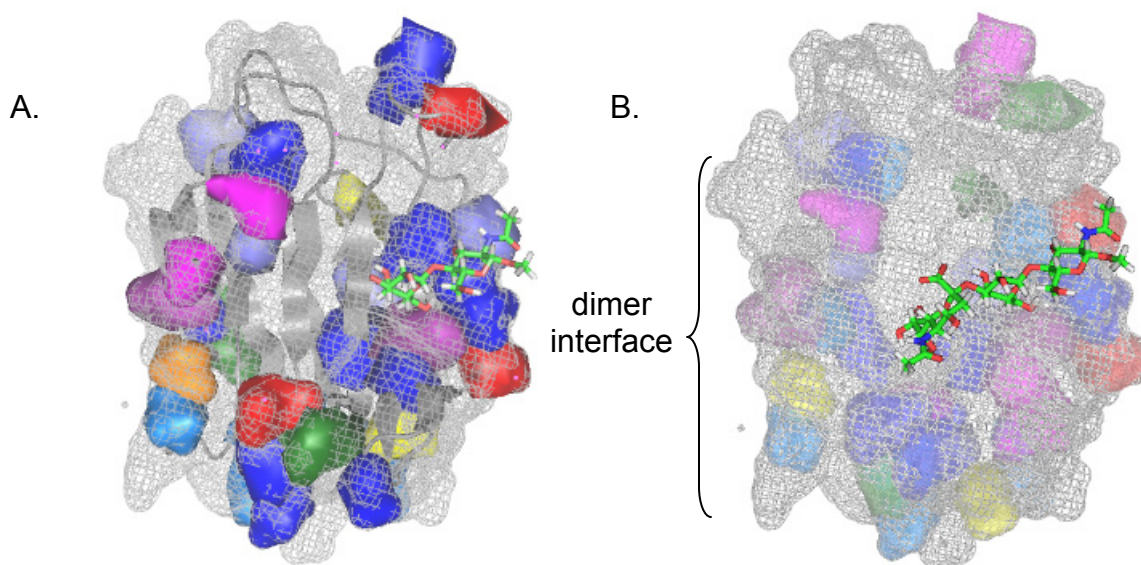


Figure 5.2 Graphical representation of chemical shift changes mapped onto structures of galectin-1 with bound LacNAc (A) from the X-ray structure by Liao et al, and Neu5Ac-LacNAc (B) as taken from the initial docked structure used in MD. The titration values (ppm) are color coded as follows: (.11<)-red, (.1-.11)-orange, (.08-.1)-yellow, (.06-.08)-green, (.04-.06)-blue, (.02-.04)-purple, (.01-.02)-cyan, (0-.01)-grey

There are several features which become apparent from the comparison of the titration results for the two ligands. The first feature is that there is no significant increase in chemical shifts in the extended binding site for the Neu5Ac-LacNAc titration. Obviously the lack of additional shifts in the proposed extended binding site is in contrast to the data obtained from the MD trajectories on these complexes. It is difficult to quantitatively interpret chemical shift perturbations of this type. However, it is possible that side chain interactions critical for binding minimally affect backbone structural properties.

It is also apparent that, on the whole, the chemical shifts are somewhat smaller for the Neu5Ac-LacNAc titration. This may be indicative of a lower binding affinity. In the future this could be quantified by a more complete chemical shift analysis of each titration point to obtain a precise K_D for each complex. Another feature evident from the titration data is the shifts seen in the dimer interface. It is possible that the binding of ligand alters the dimer equilibrium resulting in chemical shift changes due to an increased or decreased proportion of dimeric protein.

It is also interesting to view the titration data in comparison to data on the galectin-3 CRD. The NMR structure of the highly homologous galectin-3 CRD had been determined, both in the presence and absence of LacNAc¹. Regions of structural difference seen with and without ligand correspond to 'hot spots' of chemical shift difference in galectin-1, as seen in Figure 5.3. It is worth noting that the CRDs of galectin-3 do not dimerize like galectin-1.

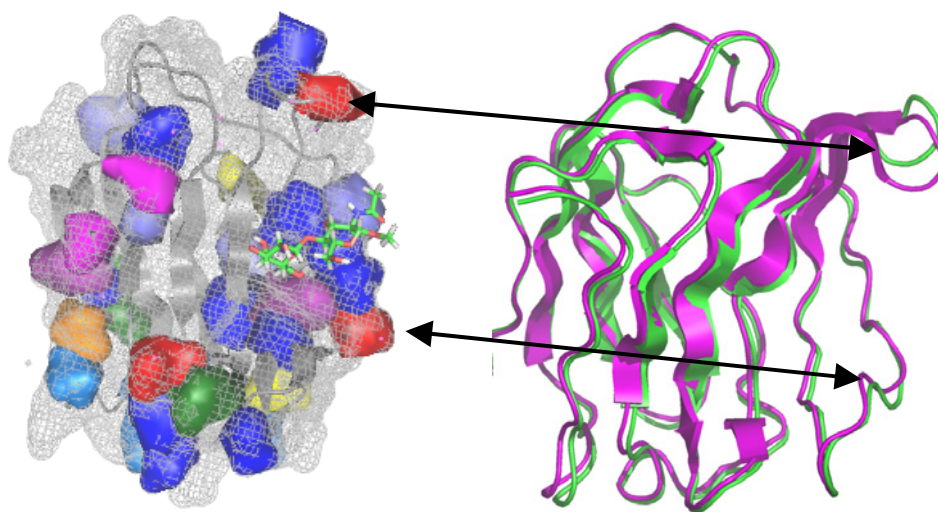


Figure 5.3 Comparison of galectin-1 chemical shift changes (A) with the NMR structure of galectin-3 with (purple) and without (green) LacNAc.

Changes in the protein structure with and without ligand would also result in chemical shift changes, adding to the complexity of interpreting this data. It is clear, nevertheless, that the NMR data do confirm a common binding site, as hoped.

Conclusions

Molecular dynamics of small oligosaccharide ligands of galectin-1

From the MD data obtained we were able to develop a series of predictions about the behavior galectin-1 ligands which gave a structural basis to the observed differences in binding affinity. It must be remembered that all of the oligosaccharides simulated in the MD studies are fragments of much larger glycans. It has recently been shown by frontal affinity chromatography that the CRDs of galectins have a higher affinity for the intact glycans, compared to di- and trisaccharides². Such glycans would make attractive targets for continuation of MD studies of galectin-1.

RDCs as an aid to assignment of protein NMR resonances

The incorporation of RMS analysis of RDC values appears to correlate well with both assignments obtained from 3D NMR data and the SEASCAPE approach. It was possible to utilize this combined approach to assign a significant portion of the galectin-1 resonances which aided in analysis of ligand titration data. The RMS analysis of RDCs remains to be tested on proteins of varying molecular weights and secondary structure composition. This work should allow for the assignment of protein backbone resonances with a minimal set of 3D experiments and NH RDCs.

Pinpointing functions for galectin-1

The functional diversity of galectin-1, together with the promiscuous binding seen for this lectin, makes defining a specific function for galectin-1 difficult. It is probably not just challenging, but perhaps misleading to ascribe a specific function to galectin-1. Instead the function of galectin-1 in a specific tissue or developmental stage will be dependent upon the context of a given biological system. This means that a complete picture of galectin-1 function must include characterization of the oligosaccharides of putative glycoprotein ligands and a detailed picture of relevant glycosyltransferases and carbohydrate modifying enzymes, such as sulfotransferases and sialyltransferases. This is best illustrated in the picture of galectin-1-induced apoptosis of CD45+ thymocytes. For example, the effects of sialyltransferases responsible for the addition of sialic acid to the N-glycans of CD45 can be fit into our model for galectin-1 ligand binding.

This structural definition of the predicted binding modes of galectin-1 ligands gives a concrete explanation for how sialylation would affect binding to this lectin. The titration data suggesting a common binding site for both LacNAc and Neu5Ac- α -(2,3)-LacNAc shows how galectin-1 could bind this modified LacNAc structure. The protein-carbohydrate interactions mapped by MD studies show a complex hydrogen bonding network which includes interactions with the galactose O6 group, which would be disrupted by the addition of an Neu5Ac- α -(2,6)-LacNAc linkage. Taken together, this shows how this important regulatory step in thymocyte development relates to the structural picture of galectin-1 binding.

References

1. Umemoto, K., Leffler, H., Venot, A., Valafar, H., and Prestegard, J., Conformational Differences in Liganded and Unliganded states of Galectin-3. 2000. **42**: p. 3688 - 3695
2. Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Muller, W., Yagi, F., and Kasai, K., *Oligosaccharide Specificity of Galectins: A Search by Frontal Affinity Chromatography*. Biochimica et Biophysica Acta, General Subjects, 2002. **1572**: p. 232 - 254.