NMR METHODOLOGIES TO DETERMINE THE STRUCTURE OF FAST EXCHANGING CARBOHYDRATE PROTEIN COMPLEX

by

TIANDI ZHUANG

(Under the Direction of James H. Prestegard)

ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy has proven to be one of the most important techniques for determining the structure and dynamics of protein-carbohydrate complexes. However, the study of ligand-protein complexes, particularly when the ligands are carbohydrates, is not without its difficulties. Historically, structure determination by NMR relies heavily on the availability of inter- or intramolecular distance constraints from Nuclear Overhauser Effects (NOEs). However, the dominance of hydrogen bonding networks in carbohydrate recognition, and an inability to observe the rapidly exchanging hydrogen-bonding protons makes observation of intermolecular NOEs rare. As a result, this thesis looks to distance-independent residual dipolar couplings (RDC) and long range pseudo contact shifts (PCSs) to constrain the global structure of carbohydrate-protein complexes. It makes application of methods based on these observables to characterize carbohydrate interactions with galectin-3, a protein of considerable interest because of its cell-surface recognition roles and correlation of these roles with a number of human diseases.

The affinities of most carbohydrates to galectin-3 are very low. The dissociation constant for lactose, the model carbohydrate used in most studies presented, from galectin-3 is 0.2 mM. In

the case of weakly-binding fast-exchanging systems, the application of RDCs is inhibited by the dominant contribution from free-state ligands. In order to accurately extract bound-state RDCs from the observed average, significant enhancement of bound-state RDCs is needed.

Novel methods for the enhancement of bound-state RDCs of lactose are presented in chapters 2 and 3 of this thesis. These include increasing the association of galectin-3 with a surrounding liquid crystal medium through a hydrophobic propyl chain and electrostatic interactions between His-tagged galectin-3 and a nickel doped alignment medium. Unfortunately, these methods do not allow the study of hydrophobic ligands. A more universal method (described in chapter 4) relies on the application of paramagnetism-based constraints, including self-oriented RDCs and pseudo-contact shifts. The results clearly show that accurately-measured RDCs and PCSs can position lactose in the galectin-3 carbohydrate binding pocket well enough to provide useful data for the rational design of competitive inhibitors of natural ligands.

INDEX WORDS: NMR, Weak Binding, Fast Exchange, Carbohydrate Protein Interaction, Galectin-3, Carbohydrate Recognition Domain (CRD), Lactose, Residual Dipolar Couplings (RDC), Alignment Media, Paramagnetism, Pseudo Contact Shifts (PCS)

NMR METHODOLOGIES TO DETERMINE THE STRUCTURE OF FAST EXCHANGING CARBOHYDRATE PROTEIN COMPLEX

by

TIANDI ZHUANG

B.S., University of Science & Technology of China, China, 2000

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

2008

NMR METHOLOGIES TO DETERMINE THE STRUCTURE OF FAST EXCHANGING CARBOHYDRATE PROTEIN COMPLEX

by

TIANDI ZHUANG

Major Professor: James H. Prestegard

Committee: Jeffrey L. Urbauer
Robert S. Phillips

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008

ACKNOWLEDGEMENTS

I would like to thank a number of people who were essential to the completion of my thesis. First, my major professor, James Prestegard, for his support and guidance throughout this project. Without him this dissertation would never have been possible. Next, my committee members, Jeffrey Urbauer and Robert Phillips, for their encouragement and suggestions. I must recognize the contributions of a number of Dr. Prestegard's lab 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 friendship that has made my experience at the University of Georgia so special:. Ron Seidel, Greg Wylie, Xiaobing Yi, Anita Kishore, Lianmei Feng, John Glushka, Laura Morris, Fang Tian, Thomas Weldeghiorghis, Megan Macnaughtan, Xu Wang, Sonal Bansal, Yizhou Liu, Hsiau-wei Lee, Han-seung Lee, Ting Li, Boaz Shapira, Shan Liu, Fei Yu, Wendy Nkari. I would like to thank Iris Tropp for her assistance in proofreading the manuscript. Finally I would like to thank my many friends here in Athens and especially my parents in China, my wife Hong Jiang, and my son Steven J. Zhuang for reminding me of what is truly important in life.

TABLE OF CONTENTS

Chapter 1

Introduction

Protein-carbohydrate interactions are essential for many biological processes and are widely seen in the enzymatic synthesis and degradation of oligo- and polysaccharides, intracellular sorting of glycoconjugates, transport of carbohydrates into living cells and of their derivatives into subcellular organelles, the immunological response to carbohydrate antigens, and the migration of leukocytes to sites of inflammation [1]. The complexity of these carbohydrate-protein interactions continues to present research challenges. Knowledge of the three-dimensional structure and dynamic properties of carbohydrate and protein moieties, either free in solution or in a bound state is essential to understanding biological processes at the molecular level. Proper modeling of carbohydrates and their interactions with biomolecular receptor proteins requires substantial relevant experimental data that is currently lacking in the literature. The application of traditional NMR methods, such as NOE (nuclear Overhauser effects) in carbohydrate-protein systems usually fails to provide adequate data due to the prevalence of hydrogen bonding networks and the lack of intermolecular NOEs. The incorporation of long-range residual dipolar couplings (RDCs) or paramagentism-based constraints (pseudo-contact shifts (PCS) and self-oriented RDCs) has been successful. The availability of these alternative structure constraints greatly compensates for the lack of inter- and intramolecular NOEs frequently seen in the carbohydrate-protein interactions. However, the accuracy of structure determination is strongly dependent upon the accuracy of extracted bound-state data. In weakly-binding, fast-exchanging systems, a large excess of ligand is usually used for the acquisition of

NMR spectra with reasonable signal-to-noise ratio. The fraction of bound ligand is very small and the RDC average is dominated by the free-state RDC. This leads to a difference between averaged RDCs and free-state RDCs within a margin of experimental error and the inaccurate extraction of bound-state RDCs. This difference can be augmented by increasing the bound-state RDCs or zeroing the free-state RDCs. In the following chapters, we present three novel nuclear magnetic resonance (NMR) methods to allow the enhancement of bound-state RDCs or zeroing free-state data (RDCs and PCSs). As an illustration, we validated our methods using a lactose-galectin-3 system. Galectin-3, a lectin that specifically recognizes the β-galactose moiety, weakly binds with lactose (Gal-(1, 4)-Glc).

In this chapter, the basic backgrounds of carbohydrate-protein interactions are reviewed and the significance of new methods of structure determination of fast exchanging systems will be emphasized.

1.1 Biochemistry of Carbohydrate-Protein Interactions

There are only ten fundamental monosaccharides (D-glucose, D-galactose, D-mannose, D-sialic acid, N-acetyl-D-glucosamine, N-acetyl-D-glactosamine, L-fucose, D-xylose, D-glucuronic acid and L-iduronic acid) expressed in mammals [2]. Variations in branching and linkages between these molecules produce an extraordinary diversity of oligosaccharides. The resulting physiological interactions of these biomolecules with different substrates are critical to many intra-and extracellular functions [3]. Complex carbohydrates are commonly found on the cell surface [1, 4]. Many protein receptors, such as lectins, that interact with cell surface carbohydrates have been identified [1]. These carbohydrate moieties can take many forms: glycopeptides, glycolipids, glycosaminoglycans, and proteoglycans.

Most carbohydrate-protein interactions are highly selective and rely on a specific stereochemical fit between complementary surfaces on the interacting molecules. Two major groups of carbohydrate-binding proteins have been identified [5]. Proteins in Group I include the bacterial periplasmic transport proteins and enzymes, which have a deep binding site and make a large number of contacts with sugars. In contrast, proteins belonging to Group II have an open-ended, groove-type binding site. Examples of this group include the lectins, viral proteins (eg. influenza hemaglutinin), toxins (eg. cholera toxin), anti-carbohydrate antibodies and pentraxins [6].

Polysaccharides are flexible, polar, have many hydroxyl groups and are hydrophilic. The binding sites of lectins are typically very shallow. Ligands lie on the binding surface and do not protrude deeply into a binding pocket, with little steric hindrance. Clearly, factors other than steric constraints such as hydrogen bonds, hydrophobic interactions are responsible for the binding geometry of polysaccharides [7].

1.1.1 Hydrogen Bonds

Hydroxyl groups, which act as both electron pair donors and receptors, are abundant in carbohydrates. This structural feature allows for the formation of numerous hydrogen bonds between the carbohydrate and its receptor protein. Although hydrogen bonding is much weaker than covalent bonding, the energy associated with this interaction can reach up to 20 KJ/mol [8]. The resulting hydrogen bonding networks are often larger and stronger than other intermolecular interactions.

The relatively free rotation about the C–OH torsional angle of the carbohydrate enables the hydroxyl groups to make optimal linear bonds with their corresponding acceptor/donor groups. The specific recognition required for protein binding (e.g. anomeric configuration,

conformation, linkage-type) and binding affinity are usually based on direct hydrogen bonding patterns (due to positioning of hydroxyl groups) [9] or indirect hydrogen bonding patterns mediated by water [10, 11]. It has been shown experimentally that the suppression or derivatization of any of these key polar hydroxyl groups or mutation of these highly-conserved residues leads to a complete inhibition of binding [12].

The binding of carbohydrates with proteins is enthalpy-driven [13, 14]. However, the effective contribution of a hydrogen bond to the formation of a carbohydrate-protein complex depends not only on its intrinsic energy in the complex, but also on the energy associated with dehydrating and transferring the hydrophilic pair to an environment of lower polarity [12]. It is the net balance of these contributions that determines whether or not a hydrogen bond is stabilizing.

1.1.2 Hydrophobic Interactions

As discussed in the previous section, hydrogen bonding usually dominates selectivity of the carbohydrate recognition process and binding affinity. Carbohydrates are quite hydrophilic due to their many polar hydroxyl groups. However, the configuration of the hydroxyl groups often creates hydrophobic patches on the carbohydrate surfaces [15]. One example is the glucopyranose ring. The flat faces of glucopyranose sugar rings are rather nonpolar. Therefore, when these hydrophobic patches interact with the hydrophobic side chains of some amino acid residues in the binding site (e.g His, Trp), the energy-favored hydrophobic interaction always enhances affinity.

To take advantage of this interaction, carbohydrate binding proteins often contain one or more aromatic residues that are well conserved in the binding site. These residues play an important role in hydrophobic stacking interactions with polysaccharide substrates [16, 17]. It is

likely that these flat surfaces of the carbohydrate contribute to the hydrophobic interactions with the aromatic side chains of Trp, His, Tyr, and Phe residues [17]. This type of interaction is commonly known as "ring stacking" and is shown in Figure 1.1.

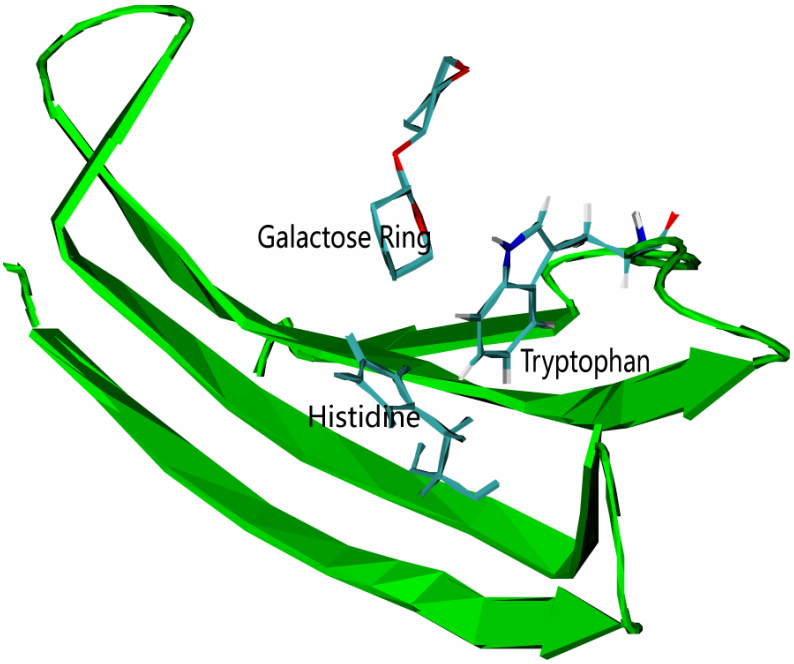

Figure 1.1 Ring stacking in a typical carbohydrate-lectin complex. The ring stacking usually involves a hydrophobic pyranose ring and the aromatic side chains of some highly conserved amino acids in the protein binding site.

The mechanisms underlying ring stacking have been well documented [17-19]. In addition to the hydrophobic interactions, ring stacking can be enhanced by either π-π or CH-π interactions. For carbohydrate-protein complexes, ring stacking is stabilized by CH-π interactions [20, 21]. In carbohydrates, this is due to the interactions between partial positive charges on the aliphatic protons on one face of a pyranose ring and a partial negative charge from the π-electrons of the aromatic ring. This mechanism also underlies the interaction between the methyl group of acetamido sugar residues and the aromatic rings of residues in the binding pocket. In fact, when modified sugars are synthesized as potential drug targets, additional ring

stacking by π-π interaction is often exploited. This is possible only when the modified sugar carries an aromatic ring. Based on this principle, many lactose derivatives carrying aromatic rings have been synthesized as potential galectin-3 inhibitors [22-25]. When an aromatic ring is attached to the galactose residue, the binding constant is increased at least 4-fold [25]. It is these ring stacking interactions in combination with other hydrophobic interactions that are considered to be the driving force for nonspecific sugar binding.

1.1.3 Electrostatic Interaction

Electrostatic interaction between two ion pairs is not commonly observed between proteins and carbohydrates due to the neutrality of the carbohydrates, but for carbohydrates that contain at least one charged group (e.g. sulfated carbohydrates, sialic acid, aminoglycans), carbohydrate-protein complexes can be facilitated by electrostatic interaction. One example of this is sucrose octasulfate (SOS) bound to fibroblast growth factor [26]. The conformation of the bound-state sucrose octasulfate is significantly different from the low-energy conformation of the free-state or crystalline K_8SOS heptahydrate. This energy-unfavorable conformation distortion is partially offset by the strong electrostatic interaction between the negatively-charged SO_4^{2-} and the rich, positively-charged amino acids in the binding site. Strong electrostatic interactions between guest and host may be the dominant factor in the formation of this complex [26].

This same phenomenon is commonly seen in sialic acid complex formation in which negatively-charged sialic acid is required for the strong interaction with a positively-charged patch in the binding site. Structural studies confirm that SO_4^{2-} can substitute functionally for sialic acid.

Calcium-bound (C-type) lectins are distinguished from other animal lectins because they retain a highly conserved calcium-dependent carbohydrate recognition domain (CRD). The C-type lectins require disulphide-linked cysteines and a Ca^{2+} ion which encourages the specific recognition of carbohydrates. When a carbohydrate binds to a CRD, in addition to the hydrogen bonds, an unusual bond is formed between the Ca^{2+} ion and certain hydroxyl groups of the carbohydrate [1, 27]. The Ca^{2+} ion is integrally involved in the formation of bonds between the protein surface and the hydroxyl groups of the sugar.

1.1.4 Galectin-3 and its Galactose-Specific Recognition

In 1956, William C. Boyd from Boston University first proposed the term lectin to describe proteins that bind to carbohydrates. The observation that lectins agglutinate certain plant cell types led to one of the first tests for the carbohydrate specificity of lectins. It is now known that lectins differ widely in their specificity in cell-agglutination reactions and in their susceptibility to inhibition by saccharides [28]. Early studies in 1956 by Watkins and Morgan [29] and subsequent research demonstrated that type-specific lectins are inhibited best by saccharides that serve as part of the immuno-determinant of the corresponding blood group substance: lectins specific for group A antigen are inhibited by N-acetyl-D-galactosamine (or its α-glycosides) and lectins specific for group H (O) antigen are inhibited by L-fucose [28]. Not all of the known lectins are specific to human blood types. Of the large group of lectins not specific for human blood groups, a few are known to be highly specific to some saccharides. These lectins interact with erythrocytes by binding to saccharide receptors present on all red blood cell surfaces but not to the blood type determinants. Concanavalin A is a non-specific lectin that recognizes α-D-mannopyranosides, α -D-glucopyranosides, and α -N-acetyl-D-glucosaminides [30].

Galectins, a family of non-specific lectins found in animals, are proteins that recognize oligosaccharides which contain β-galactose and share particular amino acid sequences that are highly conserved as a carbohydrate recognition domain (CRD) [31, 32]. A free cysteine in the binding pocket is required for ligands to bind to all galectins. Due to this property, all of galectins and a number of related lectins are referred to as "S-type" lectins. To date, 15 galectins have been identified and classified into three groups: prototype, tandem-repeat, and chimera [33]. The prototype galectins consist of galectin-1, -2, -5, -7, -10, -11, -13, -14, and -15. The characteristic feature of prototype galectins is that the protein exists as a monomer or non-covalent dimer of CRDs. Tandem-repeat galectins (galectin-4, -6, -8, -9 and -12) consist of two different CRDs in a single polypeptide chain. The only known chimera type of galectin is galectin-3, which is composed of a non-lectin domain connected to a CRD.

Galectin-3 has been identified as an important regulator in many biological processes. Galectin-3 is a Ras escort/regulator protein and involved in the splicing and transport of mRNA and the regulation of immune cell homeostasis and inflammation. Information regarding the biological activity and importance of galectins in cancer research has been discussed thoroughly in a review by Kiss, et. al. [34].

The 2.1 Å crystal structure of the galectin-3 CRD reveals that, similar to other members of the galectin family, it is composed of 5-stranded and 6-stranded beta-sheets that associate in a sandwich arrangement (PDB ID: 1A3K) [35]. The carbohydrate specificity of galectin-3 has been studied to a significant extent by Leffler and his coworkers. Although galectin-3 specifically recognizes the galactose residue, the binding affinity it has for different ligands varies greatly. A series of galectin-3 inhibitors has been designed based on the structure of lactose and N-acetyl lactosamine-bound galectin-3 [22, 23, 36]. Preliminary results from these

studies suggest that molecules with subtle differences in carbohydrate structure may have the potential to retard tumor growth, angiogenesis, and metastasis. In the future, it is likely that members of the galectin family and/or their ligands will be used as therapeutic modalities for neoplastic and inflammatory disorders.

The conformation of the ligand bound to a macromolecular target can be studied by nuclear magnetic resonance spectroscopy (NMR) in solution for both high and low affinity complexes. However, the methods for different exchange rates are varied. The following section reviews how ligand exchange affects NMR observations and which NMR observations can be used to determine the structures of bound-state ligands or whole complexes.

1.2 Chemical Exchange and its Effects on NMR Spectra

The behavior of protein-ligand complexes in NMR measurements depends mainly on their thermodynamic and kinetic properties such as binding affinity (K_d), and exchange rates (k_{on} & k_{off}) [37, 38]. Appropriate approaches for NMR investigation are determined by the characteristics of the system under study and the type of information desired.

When studying the behavior of protein-ligand interactions, one of the most essential thermodynamic factors is the binding constant. When the interaction reaches its equilibrium (i.e., the balance between the binding and dissociation after infinite mixing time), the protein is present in ligand-free [P] and ligand-bound [PL] states. When the binding stoichiometry is 1:1, the chemical exchange rates and binding constants are related as follows:

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$$

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[P][L]}{[PL]} \tag{1.1}$$

NMR observables are sensitive to both exchange rates and equilibrium constants. Nuclei with magnetic properties resonate at a characteristic frequency when placed in a magnetic field. The resonance position (i.e. chemical shift) is very sensitive to the changes in the magnetic environment. As ligands undergo exchange on and off the protein binding site, this dynamic process takes a given nucleus from one magnetic environment to another. The effects can be dramatic and diverse depending upon the rate of the exchange. NMR parameters, such as the Nuclear Overhauser Effect (NOE) [39, 40] and transverse relaxation rate (R2) depend mainly on dipole–dipole interactions, generated between nuclear dipoles tumbling in solution [41]. This exchange results in a difference in the relative mass of the two states and a change in the molecular tumbling rates in solution. The free ligands have a shorter correlation time whereas the bound-state ligands have a longer correlation time. NOE of the bound-state ligand is negative, but it is positive for free-state ligands. The Nuclear Overhauser Effect and relaxation rate will be discussed further in section 1.3.

Chemical exchange is a dynamic process which occurs in all basic molecular interactions that underly biological function, including protein-protein, protein-nucleic acid, protein-carbohydate and enzyme-ligand interactions. The degree to which chemical exchange affects NMR data can provide information about the timescales and strength of these interactions. In terms of different binding constants, three regimes have been defined (providing that the stoichiometry of this equilibrium is 1:1).

1. Strong binding, with a dissociation constant of $K_d < 1\mu M$;
2. Weak binding, with a dissociation constant of $K_d > 1mM$;
3. Intermediate binding, with a dissociation constant between $1\mu M$ and $1mM$.

As shown in equation 1.1, the dissociation constant can be expressed as the ratio of k_{on} and k_{off}. where k_{on} is a bimolecular rate constant that defines the probability of a productive encounter between the free receptor and ligand. k_{on} is limited by a diffusion controlled rate of 10^7-10^8 $M^{-1}s^{-1}$. k_{off} is a unimolecular rate constant inversely proportional to τ_B – the lifetime of the bound state. As a rough approximation, if the ligand is tightly bound with the target protein, the ligand off rate (k_{off}) is slow. If the ligand is weakly bound with the target protein, the ligand off rate is fast.

1.2.1 Slow Exchange

If ligand binding is in the slowly-exchanging regime, k_{off} usually is far less than 10 s^{-1}. For strongly-binding ligands, the concentrations of free protein and free ligand are negligible under stoichiometric conditions. Both the protein and the ligand exist almost exclusively in the bound state. The complex behaves like a single stable molecule and its spectra will not show any signs of the free species or of exchange between the free and bound species.

In NMR, slow exchange also means that the exchange rate of the ligand is much smaller than the frequency difference between the free and bound states. In this case, the appearance of the peak is only slightly broadened and two separate peaks that correspond to these two states resonate at different frequencies.

1.2.2 Fast Exchange

A ligand is weakly bound to its target protein when the dissociation constant is in the millimolar range or $k_{off} > 10^4$ s^{-1}. In a weakly-binding fast-exchanging case, a large excess of the ligand has to be used to force the formation of a complex. Nevertheless, much information about the complex can be obtained from NMR measurements, such as an indication of conformational change upon complex formation and epitope mapping for both the ligand and the substrate.

In a weakly-binding fast-exchanging system, the exchange between the free and bound state is very fast and the lifetime of the complex is short. The NMR observable is the average of the individual observable of the free and bound states weighted by the fraction of these two states. This observation can generally be expressed as:

$$A = f_1 * A_1 + f_2 * A_2 \tag{1.2}$$

Here, f_1 and f_2 are the fractions representing the free and bound states. A_1 is the NMR observable of the free state and A_2 is that of the bound state. Instead of observing two peaks at two difference frequencies, a single peak is observed whose chemical shift is the average of free and bound states weighted by the fraction of free and bound. For a system with a binding stoichiometry of 1:1 and when the dissociation constant K_d is known, the fraction of the free and bound states of the protein relate to a binding constant and can easily be determined using this formula:

$$f_2 = \frac{1}{2[P_0]}\{(K_d + [L_0] + [P_0]) - \sqrt{((K_d + [L_0] + [P_0])^2 - 4[L_0][P_0])}\}$$

$$f_1 = 1 - f_2 \tag{1.3}$$

If the binding constant is unknown, the binding constant can first be determined using NMR spectroscopy, fluorescence, or other technologies such as surface plasma resonance. Several NMR techniques have been developed to determine binding constants. One of the most widely used techniques is monitoring the chemical shift perturbation of ligand resonances during titration of the ligand with proteins. In this experiment, a ligand or protein concentration is fixed and titrated with its complementary protein or ligand. The chemical shift of one peak is observed and plotted against the concentration of the ligand if the initial concentration of the protein is fixed using the following equation:

$$\Delta\delta_{obs} = \Delta\delta_{\max}\frac{1}{2[P_0]}\{(K_d + [L_0] + [P_0]) - \sqrt{((K_d + [L_0] + [P_0])^2 - 4[L_0][P_0])}\} \quad (1.4)$$

Other NMR methods such as relaxation T1/T2 and line shape analysis also can be used to determine the binding constant. The principles of these techniques are the same and rely on the measurement of NMR observable differences upon titration [42-44].

1.2.3 Intermediate Exchange (Moderate Binding)

On occasion, when the dissociation constant is within 1 μM and 1mM or k_{off} is between 10 s^{-1} and 10^4 s^{-1}, moderate binding occurs. In moderate binding, under stoichiometric conditions, both protein and ligand will exist as bound and free species in significant percentages. Depending upon the kinetic stability of the complex, exchange between the free and bound species may affect the NMR spectra.

When the exchange is not very fast, the lifetime of these complexes approaches the difference in frequency of the signals corresponding to the free and bound states. The signals become "exchange broadened" and can be very difficult to detect by NMR. In some cases the exchange rate can be changed from moderate to fast by increasing the temperature or changing other rate-influencing conditions such as pH or ionic strength. However, this can be done only if the changes do not compromise protein stability or relevance to the biological questions being answered.

As mentioned earlier, most lectins have an open-ended, groove-type binding site. Sugars generally bind to proteins weakly with binding affinities in the range of K_d= 1mM~ 1μM, and the ligands undergo fast exchange on and off the binding site. The research described in this paper focuses on low-affinity ligands with K_d, in the micro- to millimolar range, for which NMR can play a unique role in the determination of the bound structure. NMR experiments provide

information about the dynamics of the ligand molecule when it is bound to the protein target on both slow and fast timescales. Such studies can determine, for example, the geometry of the bound-state ligands, and the specificity of protein recognition. Current NMR methods used to study carbohydrate-protein interactions are reviewed below.

1.3 Current Experimental Methods for Studying Carbohydrate-Protein Interaction

Precise knowledge about the binding of ligands to receptors is key to understanding many biological processes. NMR can provide a complete picture of ligand-bound protein complexes for proteins and ligands in solution. In theory, all NMR spectroscopic data can provide some information related to the binding activity of a ligand complexed to a protein. In practice, only data that can be obtained easily and with a high degree of sensitivity to the molecular motions are useful [45]. This is especially true in structure-based drug design, when the process of identifying and optimizing new drug candidates is very time consuming and costly [45, 46].

1.3.1 Epitope Mapping by Chemical Shift Perturbation

When the ligand binds to the receptor protein, the chemical environments of the bound-state ligands and the receptor protein change. These changes in the chemical environment can easily be observed from the NMR spectrum since the chemical shift is very sensitive to this subtle change. For proteins, the nuclei located in the binding pockets usually show the largest chemical shift changes. For ligands, the nuclei involved in the recognition show the largest effects.

When screening ligand-protein interactions using chemical shift perturbation, one can either follow the chemical shift change of the receptor protein or the ligand. Ligands have a small number of protons and the proton NMR spectrum is simpler and less overlapping than the protein spectra. By comparing the spectra of samples with and without receptor proteins, it is

feasible to identify the moiety of the ligand involved in the binding process. For proteins, theoretically, a complete ^{1}H-NMR chemical-shift map can be obtained from a 2D homonuclear NMR experiment, such as TOCSY (Total Correlation Spectroscopy). However, the severe overlapping of proton signals usually makes assignments extremely difficult. This difficulty can be greatly alleviated by using heteronuclear correlation experiments. In fact, SAR (Structure-Activity Relationships) by NMR, one of the most important methods for combinatorial library screening in the pharmaceutical industry, is based on the availability of heteronuclear correlation experiments [47, 48].

SAR by NMR can detect ligand binding to the receptor protein and identify the amino acids of the receptor proteins involved in the binding activity if the corresponding resonances are assigned. The procedure relies on the acquisition of ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Coherence) spectra of protein with and without ligand. The chemical shift values of the particular ^{15}N/^{1} H signals in the two-dimensional correlation spectrum correspond to the particular amino acids in the target molecule. On comparison of the spectra, the peaks having significant chemical shift changes usually belong to the amino acids involved in the binding [45].

Chemical shift changes aid in identifying the residues or the atoms involved in the interaction [49] but it is not simple to deduce the geometry of the ligand-protein complex from the chemical-shift perturbation observed in an ^{1}H/^{15}N-HSQC spectrum. To determine the three dimensional structure of the ligand-bound complex, the relative distance between the ligand and the protein residues and the relative orientation between the protein and the ligand are required. Algorithms have been developed to determine the structure of the ligand-bound complex using chemical shift perturbation and other NMR or X-ray crystallography data when both the structures of the ligand and the protein are known. HADDOCK is one of the CNS-based] [50, 51]

docking programs for protein-protein, protein-DNA/RNA, peptides- phospholipids and protein-carbohydrate docking [52, 53].

In applying high-ambiguity data docking (HADDOCK), chemical shift perturbation data is used to define the ambiguity interaction restraints (AIRs) between the active and passive residues of both molecules. In practice, the active residues should have more than 40% of their surface solvent-accessible and chemical shift changes that are significantly larger than the average. Passive residues exist next to active residues and have more than 40% of their surfaces accessible to solvents. An AIR is defined as an ambiguous distance between all residues involved in the interaction and is set to be 2 Å, with a maximum 4 Å and minimum 2 Å. The application of HADDOCK for docking carbohydrates to proteins has been successful in our laboratory using both chemical shift perturbation data and other constraints. Further details of this process will be discussed in chapter 2.

Because chemical shift perturbation data does not provide any direct distance or orientation restraints for structure determination, high ambiguity docking that relies on chemical shift perturbation data is not sufficient. A subsequent NMR structural analysis is necessary to unambiguously determine the displacement or orientation of the ligand. The following sections will review additional NMR techniques which enable the measurement and collection of data related to intra- and intermolecular distance and orientation.

1.3.2 Nuclear Overhauser Effects

The field of NMR spectroscopy has seen tremendous growth in the last twenty years, particularly in advances related to Nuclear Overhauser Effect spectroscopy (NOESY). NOE spectroscopy is the most powerful technique for obtaining structural information on molecules in solution [41, 54, 55] and relies on the transfer of nuclear spin polarization from one spin

population to another via cross-relaxation [56]. NOEs are small but observable changes in resonance intensity that reflect internuclear distances (<5 Å). NOEs are extremely useful for protein structure determination, but can also reflect distances between ligand protons and protons on amino acid residues in protein binding sites.

An energy diagram of a two-spin system (I and S as shown in figure 1.2) is frequently used to illustrate the origin of the NOE. Resonance intensities for protons and other spin ½ nuclei are related to the population difference between α and β states. Once displaced from equilibrium populations, the possible transition pathways allowing recovery of an equilibrium state of a two-spin system can be classified into three groups:

1. Spin-lattice relaxation including $\alpha\beta \rightarrow \alpha\alpha$, $\beta\alpha \rightarrow \alpha\alpha$, $\beta\beta \rightarrow \beta\alpha$ and $\beta\beta \rightarrow \alpha\beta$. In this transition, only one spin (I or S) changes the population of α and β states;
2. Zero-quantum transition including $\alpha\beta \rightarrow \beta\alpha$.
3. Double-quantum transition including $\beta\beta \rightarrow \alpha\alpha$.

NOE enhancement is defined as the fractional change in the intensity of spin I on saturating spin S:

$$F=(I-I_0)/I_0 \tag{1.5}$$

Here I_0 is the signal intensity when spin I reaches a Boltzmann equilibrium. I is the signal intensity of spin I on saturating spin S and is proportional to the sum of the population differences ($N_{\alpha\alpha} - N_{\beta\alpha} + N_{\alpha\beta} - N_{\beta\beta}$).

The magnitude of NOEs are derived from the following equation [41]:

$$\varepsilon_{NOE} = \frac{\gamma_s}{\gamma_I} \frac{W2 - W0}{W0 + 2W1 + W2} \tag{1.6}$$

Here γ_s and γ_I are the gyromagnetic ratios of spin S and spin I. W2 and W0 are the rate of double-quantum and zero-quantum transition. W2-W0 describes the rate of spin relaxation which gives rise to an NOE signal [57, 58]. In order to identify the preferred transition pathway the molecular correlation time must be known. Shorter correlation times (as found for small molecules) spread fluctuation frequencies to higher values; longer correlation times (larger molecules) concentrate fluctuation frequencies to lower values. For small molecules, W2 is preferred and the NOE enhancement for small molecules is positive. For large molecules, W0 is greater than W2, so negative NOE enhancement is observed.

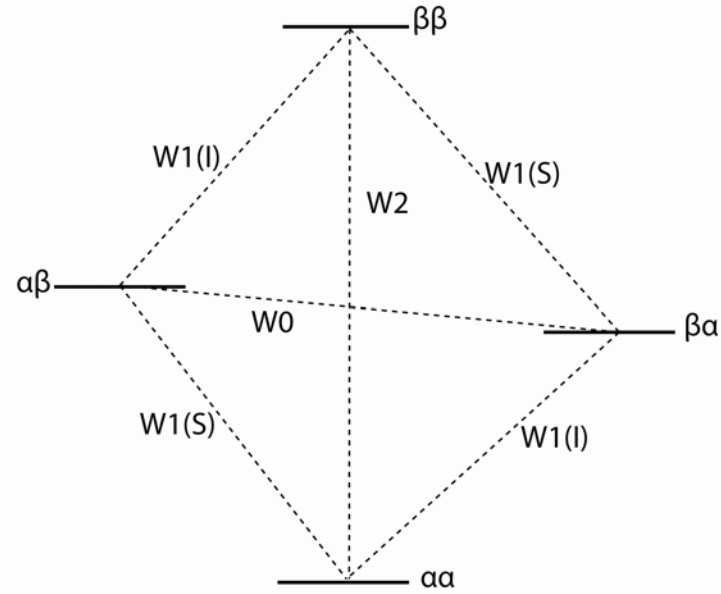

Figure 1.2 The transition pathways a of two-spin system (I and S). Here αα represents the ground state of both spin I and spin S. W0, W1 and W2 represents the zero, single and double quantum relaxation respectively.

Cross-relaxation rates [59, 60] and spectral density functions (J) [41] have been used to determine exactly how the NOE varies with molecular tumbling. The observed NOE is determined by the cross-relaxation rate as:

$$\sigma_{IS} \propto \frac{1}{r_{IS}^6}\{6J(\omega_I + \omega_S) - J(\omega_I - \omega_S)\} \qquad (1.7)$$

1.3.2.1 NOE Experiments

Three types of NOE experiments have been developed:

1. steady state

2. truncated driven NOE (TOE)

3. transient NOE (1D inversion and 2D)

A steady state NOE experiment is set up by saturating a single resonance at low power for approximately five times T1 before acquiring the FID. A truncated driven NOE is similar to steady state NOE but saturates a single resonance for shorter periods so the buildup of NOE can be observed. The 1D transition NOE experiment is set up by selective inversion of a single resonance and the 2D transition NOE is set up by frequency labeling all resonances by a 90 degree pulse and a variable delay.

When the product of proton nuclei frequency ω and the motional correlation time τ_c is close to 1, no NOE can be observed. A rotating-frame NOE experiment (ROESY) [61, 62] can be used to detect the NOE effects of the system when $\omega\tau_c$ is close to 1. For example, a molecule with a molecular weight of around 2 kDa has a very small NOE effect at 800 MHz, the field to be used most frequently in this thesis. For molecules like this, a ROESY experiment is desirable when observing the properties of the free molecule as opposed to those of a bound complex.

Any quantitative interpretation of steady state NOEs requires knowledge of the arrangement of all protons relative to each other. Steady-state NOE is normally used in cases of

well-resolved spectra. The steady-state NOE develops during saturation of one spin, and results in a change in the intensity of the signal from another spin. In the study of carbohydrate-protein complexes, the protein signals can be selectively irradiated and the intensity change of the ligand signals is detected. This intensity change can be quantitatively analyzed using a complex relaxation and conformational exchange matrix analysis package (CORCEMA) [63-66]. This package allows the simultaneous optimization of the conformation of the bound ligand and side chains in the binding site, providing that sufficient NOEs can be observed. In our research, few NOEs were observed between ligands and proteins; we did not utilize these more advanced analysis packages. For our purposes, initial rate measurements which display a $1/r^6$ relationship were adequate (Refer to equation 1.7).

NOEs (i.e. cross-peak integrals) can be converted into interproton distances. This is usually done with the help of a reference NOE corresponding to a known distance (e.g., CH2 groups, H-C-C-H fragments in aromatic rings etc.):

$$r_{ij} = r_{ref} \sqrt[6]{\frac{I_{ref}}{I_{ij}}} \qquad (1.8)$$

1.3.2.2 Transferred NOE

Nuclear dipole-dipole interaction drops off very fast with distance between two nuclei. NOESY spectra provide information about protons that are 5 Å or less apart in space. The presence of an NOE peak is considered to be direct evidence that two protons are within 5 Å of each other. The application of NOEs has been applied to the study of ligands bound to protein. The direct measurement of intermolecular distances between the receptor protein and the bound ligand provides the basis of bound-ligand geometry and conformation of the whole complex. This extension of 2D NOESY to exchanging systems, such as carbohydrate-protein complexes, is called transferred NOE (trNOE) [67-69].

In the presence of chemical exchange of the ligand between its free and bound states, the intramolecular trNOE allows the transfer of information in the bound ligand to the free ligand resonances concerning cross-relaxation between two nuclei. In the free state, the ligand is generally characterized by short correlation times ($\tau_c \sim 10^{-10}$ s) and thus may be either in the extreme narrowing limit ($\omega\tau_c << 1$), where the NOEs are small and positive, or at the boundary of the extreme narrowing limit ($\omega\tau_c \sim 1$), where the NOEs approach zero. When bound to the protein, the ligand is characterized by the long correlation time of the protein ($\tau_c > 10^{-8}$ s), and thus is in the spin diffusion limit ($\omega\tau_c > > 1$), where the NOEs are large and negative.

In studying large proteins (Mr > 20 kDa), distances derived from measuring of two-dimensional trNOESY cross-peak intensities can be modified by indirect cross-relaxation with other spins. One of the major indirect cross-relaxation effects is spin diffusion, which is very pronounced in large molecules. In this case, beside direct enhancements between protons close in space, other spins may mediate the exchange of magnetization, and produce negative cross-peaks between protons far apart in the macromolecule. These indirect trNOE effects may lead to interpretation errors in the analysis of the ligand-bound protein conformation.

As discussed previously, the ROESY experiment can be used to distinguish the three-spin from the real through-space NOE effects. This also applies to the transfer ROESY experiment to distinguish spin diffusion from the real NOEs. In the transferred ROESY experiment, spin-diffusion (three spin) effects appear as positive cross peaks. The application of this experiment permits one to distinguish direct from indirect enhancements, and complements those measured under regular conditions. The result is conformational information which is less contaminated by artifacts.

In general, one can observe inter- and intra-molecular trNOEs. Intramolecular trNOEs provide the basis of intramolecular distance restraints to define bound-ligand conformations. Intermolecular trNOEs occur between a ligand proton and a proton on a receptor protein, and therefore, in principle, can contribute to the definition of the geometry of bound ligands in protein binding pockets [70, 71]. However, the application of trNOE to fast-exchanging carbohydrate-protein system is limited by the fact that no adequate trNOEs are available. As seen in our case, there is only one trNOE available.

1.3.3 Saturation Transfer Difference (STD)

Saturation Transfer Difference NMR spectroscopy [45, 72] is a new member of bio-affinity NMR methods which allow the detection and identification of binding molecules directly from mixtures. Any small molecule in solution reversibly interacting with the binding domain of a protein shows characteristic differences in its relaxation behavior and molecular mobility from those molecules having no affinity for the protein. The principle of the STD experiment is well established. For a protein, it relies on a large system of protons tightly coupled by dipole-dipole interactions [73] . The selective longitudinal relaxation rate R1 of a proton that is discreetly excited in proteins is dominated by the cross-relaxation rate. Selective saturation of a single protein resonance will result in a rapid spread of the magnetization over the entire protein through spin diffusion. When a ligand is present, intermolecular transfer of magnetization from protein to ligand leads to progressive saturation of the ligand [72].

Although the mechanism of the STD experiment is the same as the NOE pumping experiment, which relies on the polarization transfer from the receptor protein to its bound compound, STD NMR spectroscopy does not require a diffusion filter, and the NOE mixing time

is replaced by the selective saturation time. The STD experiment is applicable to a wider range of K_d values (1 mM~10 nM) as it is the free fraction of the ligand that is observed.

STD NMR spectroscopy should be applicable to both weak-binding and strong-binding ligands although the absolute effect of the strong-binding ligands will be smaller than with weak binding ligands since fewer ligands are saturated and exchanged into solution. However, the STD experiment does not provide any information on the nature of the binding site in the receptor protein. This disadvantage limits its application to determining the structure of the ligand-bound protein complex.

A comparison of SAR and STD experiments for the identification and characterization of ligand binding is summarized in table 1.1. In our case, carbohydrate is weakly binding to the receptor protein with K_d in the range of milimolar. STD experiments can be performed to identify the protons close to the protein binding site.

Table 1.1 Comparison of SAR and STD NMR for the identification and characterization of ligand to protein

	SAR by NMR	STD
Protein size limit	Small protein (<10 kDa)	Large Protein (>30 kDa)
Protein isotope-labeling required	Yes	No
Ligand isotope-labeling required	Yes	No
Binding epitope on Protein	Yes	No
Binding epitope on Ligand	No	Yes
Binding affinity limit	<1 mM	10 pM-10 mM
Ligand identification	No	Yes
Sensitive to excess ligand	Yes	No

STD does not provide distance information directly. A CORCEMA-based procedure (CORCEMA-ST) has been developed as to utilize the STD information for the simulation annealing [74]. STD is dependent on the efficiency of the spin diffusion. Spin diffusion becomes more and more efficient with the increase of the molecular mass. In the system we study, the

galectin-3 carbohydrate recognition domain has a molecular mass of 16 kDa and the STD effect is very small.

1.3.4 Residual Dipolar Couplings

Until recently, the principal NMR data for structure determination were NOE and scalar J couplings. However, the weakness of these data as the structural restraints is obvious. They are short-range in nature, (i.e. NOE restraints provide the distance information of two nuclei only less than 5 Å apart), dihedral angle restraints only restrict groups of atoms separated by three bonds or less [75]. Over large distances, uncertainties in short-range restraints become more significant. Using only short-range constraints, the NMR structures of large, elongated systems are often very poor, even though individual regions of the structure are well-defined. Recently, more long-range constraints have been exploited to refine overall structures that were determined initially by using NOEs, dihedral angles or hydrogen bondings.

Residual dipolar coupling between two spins in a molecule occurs if the molecules in solution exhibit a partial alignment. This leads to an incomplete averaging of spatially anisotropic dipolar couplings [75-77]. Dipolar couplings have long been a mainstay in solid-state NMR, but only recently have they become routine in solution NMR.

RDC contains distance information as well as angle information. The latter relates to the angle formed by a vector connecting two interacting atoms and the axes of an alignment tensor. RDCs also can be expressed as where the effect of the alignment tensor is buried in the averaging brackets [76, 77]:

$$RDC = -\frac{\mu_0 \gamma_i \gamma_j \hbar}{4\pi^2 r^3} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \tag{1.9}$$

Here γ_i and γ_j are the gyromagnetic ratio of the nuclei studied. r is the inter-nuclei distance. θ is the angle between the inter-nuclear vector and the magnetic field. The bracket represents the averaging of the term $\frac{3\cos^2\theta - 1}{2}$. Under isotropic conditions, the molecules of interest are freely rotating in solution and θ is evenly distributed, so $\frac{3\cos^2\theta - 1}{2}$ is averaged to zero. In the isotropic condition, no residual dipolar coupling can be observed. However, if the molecule has a preferred orientation induced in the magnetic field and θ is no longer evenly distributed, this term is no longer averaged to zero. This always can be accomplished by observing the molecules in various alignment media, such as lipid-based liquid crystals (bicelles) and phage, or by indirectly aligning the molecules through the effects of magnetic fields on attached paramagnetic centers with highly anisotropic magnetic susceptabilities.

All of the aqueous liquid crystal alignment media used in NMR (Bicelle, PEG/Hexanol, stretched polyacrylamide gel, filamentous phage, cellulose crystallites) [78-85], share a common property in that they acquire a macroscopic order in a magnetic field. After mixing the molecules of interest with these media through steric collisions, they indirectly induce a slight orientation bias of the molecules. Up to tens of kHz of dipolar couplings on fully-aligned samples (such as solids) can be easily detected in solid state NMR. Dipolar couplings from partially-aligned solution samples are usually less than 100 Hz and do not cause severe dipolar couplings and distorted spectra. RDCs arise with this slight orientation bias in solution but the overall simplicity of solution NMR spectra remains [86].

In addition to the indirect induction of order by the liquid crystal media, RDCs can be observed in molecules with paramagnetic centers or diamagnetic systems such as DNA where the small anisotropy in each base is additive over the entire molecule. The inherent magnetic

susceptibility anisotropy of some metal ions (e.g. Ni^{2+}, Ln^{3+}, Fe^{3+}) leads to a sufficient order of the molecules in the magnetic field if these paramagnetic ions are integrated into the molecules either through the substitution of a diamagnetic ion or through some metal ion binding tags [87-94]. The paramagnetic metal ions raise not only a field-dependent alignment of the molecules by the magnetic susceptibility, but also distance-and orientation-dependent chemical shift changes due to the induced magnetic field by the unpaired electrons of the paramagnetic ions. The mechanism of these paramagnetic effects is complicated but these paramagnetism-based constraints have gained a lot of interest in studies of protein structure or protein-ligand interactions due to their rich source of structural information. We decided to devote a separate section to review the mechanism and application of paramagnetism, and it can be found section 1.3.5.

1.3.4.1 RDC and Order Tensors

From the equation above (equation 1.10), residual dipolar coupling is related to the inter-nuclei vector length and the angle between the inter-nuclear axis and the direction of the partial ordering of the protein. For protein backbone amide protons bonded to amide nitrogens, the distance is usually considered to be fixed at 1.04 Å. For carbohydrate pyranose ring H-C vectors, the distance is also considered to be a constant of 1.4 Å. In other words, RDCs reflect angles directly and, in principle, show us the relative orientation of two vectors even if they are on the opposite ends of a molecule. These long-range constraints have proven to be an important complement to the conventional NOE distance restraints and dihedral angle restraints which define short range order in structure biology.

When alignment results from a field-induced order, residual dipolar coupling is expressed by the following equation:

$$RDC = -\frac{\mu_0 \gamma_i \gamma_j \hbar}{4\pi^2 r^3}\left(\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi \right) \quad (1.10)$$

Here $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are axial and rhombic components of the magnetic susceptibilities.

When working with ligand-protein complexes, order tensors determined individually from data on bound ligands and proteins can provide powerful geometric constraints on the complexes. The directions and levels of the orienting force should appear the same from the point of view of each fragment when the whole complex is considered to be quite rigid. This provides a test for the validity of the proposed models for the orientation of a ligand in a protein binding site.

1.3.4.2 RDC Measurement

In NMR spectra, RDCs appear as an additional contribution to the scalar J coupling splitting. The magnitude of RDCs can be positive or negative and must be determined by taking the difference of the splitting under anisotropic conditions (J+D) and under isotropic conditions (J). There are numerous methods that have been designed to accurately measure dipolar couplings between nuclei [95-97]. They have been classified into two groups: frequency-based methods where separation of peaks centers (splitting) is measured in a frequency domain, and intensity-based methods where the coupling is extracted from the resonance intensity instead of splitting. One of the commonly used frequency based experiments in small systems is the IPAP (In-Phase Anti-Phase) experiment [98]. This is particularly valuable when an ^{1}H-^{15}N RDC is needed for smaller protein systems. For a large system, the interleaved HSQC/TROSY experiment is preferred [99, 100]., A constant-time COSY sequence for measuring ^{1}H-^{1}H dipolar couplings [101] has been developed by Tian et al. This method integrates the cross- and diagonal peaks and fits the ratio of cross- and diagonal peaks against different time delays using a tangent

function, instead of acquiring RDCs by measuring the splitting difference. Constant-time COSY is one of the intensity-based methods to be used in the studies which follow.

1.3.4.3 Transfer RDC

The geometry and conformation of bound-state ligand and ligand-protein complexes have traditionally been deduced from distance constraints between pairs of NMR active nuclei spanning the ligand-protein interface and the intramolecular NOEs of bound-state ligands. Especially for the weak binding, water mediated ligand recognition, common in carbohydrate-binding proteins, two protons of interest spanning the ligand-protein interface are far way (>5 Å). The steep inverse distance dependence ($1/r^6$) of the nuclear Overhauser effect (NOE) makes identification of sufficient numbers of distance constraints very difficult. In these cases the long-range orientation or distance information complementary to NOE-derived information can be very important.

In ligand binding systems, the measured ligand RDCs reflect an average degree of alignment from the free and bound orientations. The free-state RDCs can be measured in the alignment media for the ligand alone or can be zero for field-induced paramagnetic systems in isotropic buffers. Once the RDCs for ligands in the protein-bound conformation have been determined, the alignment tensor can be calculated and rotated to coincide with the protein's alignment tensor. The coincidence of the alignment tensors in the principle alignment frame allows the determination of the relative orientation of bound ligands to its receptor proteins. With few intermolecular distance restraints, the ligands can be well positioned in the protein binding pocket. There have been several RDC applications to determine bound ligand structure or the whole complex [75, 102]. Problems arise when studying the weak binding system under its natural abundance. The indirectly-induced alignment through the collisional interaction usually

is not large enough to make the RDCs of bound-state ligands substantially different from those in the free state. This leads to an indistinguishable observed average and free-state RDCs. One of the main objectives of this thesis research is to increase the difference between observed average and free-state RDC. The accurate bound state RDC can be extracted only when the difference between observed average and the free-state RDCs is large.

1.3.5 Paramagnetism-Based Constraints

Paramagnetic NMR is used to observe the magnetic effect of unpaired electrons on nuclear spin. For a molecule with a paramagnetic center, the observation of nuclei with a large gyromagnetic ratio, such as protons, becomes very difficult due to severe line broadening. For diamagnetic systems, both Fermi contact and dipolar coupling make significant contributions to the relaxation of the nuclear spin. For fast-relaxing paramagnetic systems at high magnetic fields, a third relaxation mechanism, Curie relaxation (due to the interaction between the nuclear spins and the induced electron magnetic moment) makes an additional contribution to the relaxation process. This additional contribution to the transverse relaxation is field-dependent and in a high magnetic field, leads to severe line broadening. Because the relaxation is dominated by coupling between the nucleus and unpaired electrons, rather by the coupling between two nuclei, the intensity of the Nuclear Overhauser Effect also is reduced.

However, the application of paramagnetic NMR to studying metalloproteins or ligand protein interaction is particularly attractive and feasible. The presence of a paramagnetic metal or spin labels in a system now offers opportunities for obtaining structural and mechanistic information through a variety of NMR observables which are not obtainable in the diamagnetic systems.

Paramagnetism-based constraints for NMR structure determination usually consist of hyperfine chemical shifts (mostly pseudo contact shifts (PCS)) [90], paramagnetic relaxation enhancements (PRE), and residual dipolar couplings (RDC) due to paramagnetically-induced self-orientation, and by cross correlations between Curie relaxation and dipolar relaxation (CCR) and some residual chemical shift anisotropy (RCSA) [103]. All these pieces of information provide either the distance or orientation constraints vital to structure refinement. The observations are long-range and suitable for the structure determination of protein-protein, protein-ligand complexes and studying the dynamic behavior of these complexes. Although only the pseudo-contact shifts and the field-induced residual dipolar couplings were utilized in this research, a brief review on all these constraints will be presented here.

1.3.4.1 Paramagnetic Relaxation Enhancement

In 1955, Solomon published the first mathematical equations of NOE and paramagnetic relaxation enhancement (PRE). Although longitudinal (R1) and transverse relaxation (R2) rates are capable of providing valuable information, more applications use transverse relaxation partially due to the easily measured inverse relationship with the line broadening of NMR signals. Here we also emphasize on the transverse relaxation effects.

For the molecules with a paramagnetic center, the electron magnetic moment is 658 times that for a proton so the electron-nuclear dipolar coupling is dominant in the contribution to the nuclear paramagnetic relaxation enhancement and the nuclear hyperfine shifts.

In terms of the transverse relaxation (R2), the Solomon equation associated with dipolar relaxation can be written as:

$$R_2^{dip} = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_N^2 g_e^2 \mu_B^2 S(S+1)}{r^6}\left[4\tau_c + \frac{\tau_c}{1+(\omega_I-\omega_S)^2\tau_c^2} + \frac{3\tau_c}{1+\omega_I^2\tau_c^2} + \frac{6\tau_c}{1+(\omega_I+\omega_S)^2\tau_c^2} + \frac{6\tau_c}{1+\omega_S^2\tau_c^2}\right] \tag{1.11}$$

Here τ_c, the effective correlation time, is given by the sum of the electron relaxation τ_s^{-1} and molecule rotational correlation τ_r^{-1} and chemical exchange rate τ_M^{-1}:

$$\tau_c^{-1} = \tau_s^{-1} + \tau_r^{-1} + \tau_M^{-1} \tag{1.12}$$

As with the paramagnetic shift, nuclear relaxation also may arise from contact interaction of the nuclear magnetic moments with the unpaired spin density delocalized at the nuclei. The equation for the transverse contact relaxation is

$$R_s^{contact} = \frac{1}{3} S(S+1)\left(\frac{A_c}{\hbar}\right)^2 \left(T_{1e} + \frac{T_{2e}}{1+\omega_s^2 T_{2e}^2}\right) \tag{1.13}$$

Because $\frac{A_c}{\hbar}$ is different from zero only for the nuclei coordinated to the paramagnetic center, it is usually neglected when longer distances are derived from the relaxation enhancement data.

A third contribution to the relaxation enhancement is the magnetic susceptibility relaxation, commonly known as Curie spin relaxation. The Curie relaxation is due to the interaction of the nuclear spins with the static magnetic moment generated by the small Boltzmann distribution difference in the population of the electron spin levels [104, 105]. The dipolar contribution to the transverse relaxation provided by this mechanism is

$$R_2^{Curie} = \frac{1}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\omega_I^2 g_e^4 \mu_B^4 S^2 (S+1)^2}{(3k_B T)^2 r^6} \left(4\tau_r + \frac{3\tau_r}{1+\omega_I^2 \tau_r^2}\right) \tag{1.14}$$

According to the above equations, each of these three contributions is capable of providing valuable information and all of these mechanisms contribute to the overall paramagnetic relaxation enhancement. The relative weights of dipolar, contact and Curie spin relaxation are variable. Under certain circumstances, the contribution from one or two mechanisms might be negligible. In fact, this always applies to spin labels (eg. nitroxide) or lanthanide ions.

PRE analysis for spin label is simple since Curie spin relaxation that could potentially exhibit significant cross-correlation with other relaxation mechanisms is negligible. The contact relaxation usually occurs when a given moiety is covalently bound to a paramagnetic center and efficient unpaired spin delocalization is present. So for spin labels such as nitroxide, in observed proton PRE effects, relaxation mechanisms other than dipolar relaxation can be neglected. The direct distance information between the unpaired electron and the nucleus can be directly derived based on the Solomon equation.

The case is different for lanthanide ions. For lanthanide-containing metalloproteins, the rotational correlation time usually is large and in a high magnetic field the Curie relaxation is significant due to the short electron relaxation. In fact since the electron relaxation time is much smaller than the rotational correlation time, the dipolar relaxation is much smaller than the Curie relaxation. The contact relaxation effect for lanthanide is as negligible as the spin label and only applies to the nuclei directly coordinated to the metal ion. The direct distance information can be directly derived using equation 1.14.

For both spin labels and lanthanides, the contact relaxation is no longer negligible for the nuclei close or directly coordinated with the paramagnetic center. The observed PRE effect is the sum of all three relaxation mechanisms.

Paramagnetic relaxation enhancement can be observed for the nuclei up to 35 Å away. PREs have been widely used in protein structure refinement and can be applied to ligand-protein system due to the relative long-range distance restraints.

1.3.4.2 Pseudo Contact Shifts (PCS)

For a paramagnetic molecule in an external magnetic field, the unpaired electron induces an averaged non-zero magnetic moment which influences the static magnetic field. A nucleus responds to this field difference which appears as chemical shift changes in the NMR spectrum. A nucleus will sense the induced magnetic field depending upon its position within the dipolar field. The anisotropic orbital contribution to the induced magnetic moment leads to intensity changes of the induced magnetic moment upon molecular rotation in an external magnetic field. So the magnetic susceptibility tensors for most paramagnetic molecules are anisotropic. The chemical shift changes [103, 106] can be expressed as

$$PCS = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}\left(3\cos^2\theta - 1\right) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi\right] \qquad (1.15)$$

Here $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic component of the magnetic susceptibility tensors. θ and ϕ are the polar angles with respect to the principle axes of the magnetic susceptibility tensors and r is the distance between the nucleus and the paramagnetic center.

Pseudo contact shifts provide both distance and orientation information. PCSs can be accurately measured for nuclei up to 40Å away from the paramagnetic center. This long-range structure information is a good complement to local structural constraints such as NOEs in protein structure refinement. Pseudo contact shifts are measured as the difference in chemical shifts between the paramagnetic and diamagnetic samples. Pseudo-contact shifts share the same magnetic susceptibility tensors used in describing residual dipolar couplings. Lanthanide ions, other than Gd, La and Lu, have large anisotropic magnetic susceptibility tensors [107] and short electron spin relaxation times. In our research, we used dysprosium to induce PCSs and determine the structure of carbohydrate-protein complexes. Details of these methods are provided in chapter 4.

1.3.4.3 Self-Oriented Residual Dipolar Couplings (RDC)

In paramagnetic molecules, the molecules are partially self-oriented in the external magnetic field due to the magnetic anisotropy of the electron magnetic moment. As residual dipolar couplings are induced by the alignment media, the self-oriented RDCs are dependent upon the size of magnetic anisotropy and the coupled nuclei vector orientation with the magnetic susceptibility tensors. Since the electron magnetic moment is induced by the external field, the size of the electron magnetic moment depends on the field strength. As a result, self-oriented RDCs also are field-dependent. The overall residual dipolar coupling [87, 108] is written as

$$RDC = -\frac{1}{4\pi}\frac{B_0^2}{15KT}\frac{\gamma_i\gamma_j\hbar}{2\pi r_{ij}^3}\left[\Delta\chi_{ax}\left(3\cos^2\theta-1\right)+\frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi\right] \qquad (1.16)$$

RDC share the same magnetic susceptibility tensors as pseudo contact shift provided that the RDC is measured through the subtraction of the J couplings of diamagnetic sample from the paramagnetic (J+D) couplings [109]. Except Gd, La and Lu, the large magnetic susceptibility of the other lanthanide ions provides sufficient alignment for RDC measurement if the metal ions show little relative motion to the molecule itself.

1.3.4.4 Cross Correlation (CCR) between Curie relaxation and Dipole-Dipole Relaxation

One important source of distance constraints for paramagnetic proteins is cross correlation between Curie spin relaxation and dipolar relaxation. Cross correlation rates between Curie spin relaxation and H-X (e.g. H-N) dipole-dipole coupling depend on the metal-to-proton distances and on the M-H-N angle. Under the assumption of isotropic χ tensors, the equation for CCR is written as [110]

$$CCR = \frac{\mu_0}{4\pi}\frac{B_0\gamma_H^2\gamma_N\hbar\chi}{10\pi^2 r_{HM}^3 r_{HN}^3}\frac{3\cos^2\theta_{MHN}-1}{2}\left(4\tau_c+\frac{3\tau_c}{1+\omega_I^2\tau_c^2}\right) \qquad (1.17)$$

Here r_{HM} is the distance between the proton and the metal ion. r_{HN} is the distance of H-N vector. θ_{MHN} is the angle between H-M vector and H-N vector.

These cross correlations can, in principle, be detected by the same pulse sequences as are commonly used in diamagnetic systems to detect cross correlations between DD and CSA (chemical shift anisotropy) [105, 110]. In the paramagnetic system, the cross correlation between dipole-dipole and chemical shift anisotropy [111] is present as well as the cross correlation between the dipole-dipole and Curie relaxation. The cross correlation between dipole-dipole and CSA can be determined in analogous diamagnetic systems. For a paramagnetic system, the cross correlation between dipole-dipole and Curie relaxation can be isolated by subtracting the effect due to cross correlation between dipole-dipole and CSA.

Considering the above observables, it is clear that a variety of new constraints associated with paramagnetic systems becomes available to solve structures. Most of the structural constraints derived from these observations of the electron-nucleus interactions are long-range and can be an important complement to the local NOE constraints to refine protein, protein-protein, protein-ligand interactions.

1.4 Significance of the Development of New NMR Methods to Study Weakly-Binding, Fast-Exchanging Carbohydrate-Protein Interactions

Although X-ray crystallography provides the fastest route to high-resolution structures of proteins, there are limitations when it comes to the examination of geometries of ligand-protein interactions, particularly when the interactions are weak. The intrinsic flexibility of the carbohydrate can lead to disordered segments, and crystal packing can distort or exclude certain bound-ligand geometries. An example is wheat germ agglutinin/ GlcNAc-β(1,6)-Gal-β(1,4)-Glc complex. The Glc residue exhibits several crystal packings with other wheat germ agglutinin

[112]. For carbohydrate-protein complexes, ligands usually have at least some atoms that protrude from the protein binding site, leaving them exposed to the solvent. These atoms do not interact with the protein but they often participate in crystal packing and distort the ligand conformation in its crystal form [113]. X-ray crystallography also is limited by the fact that a number of proteins undergo large changes in structure upon ligand binding, making crystal soaking techniques impractical for the introduction of ligands into pre-formed crystals. Although X-ray crystallography is valuable, it is essential to complement this data with information from other sources (*e.g.*, NMR) to confirm the structure of ligand-protein complexes in solution.

NMR measurements offer new insights into the conformation of bound oligosaccharides at an atomic level and deliver a complete picture of ligand binding to its receptor. However, special problems arise when the ligands are carbohydrates. When NMR is applied to study carbohydrate-protein interactions, existing NMR techniques are inefficient due to the fact that hydrogen bonding networks dominate the carbohydrate-specific recognition and the carbohydrate-protein interaction is weak. This results in few observable NOE constraints and difficult analysis of other averaged observables. The isotope labeling of carbohydrates is very difficult and carbohydrates are studied in their natural abundance, greatly reducing the sensitivity of carbon-resolved NMR techniques. In order to increase the signal-to-noise ratio, an excess of the ligand must be used. However, the excess of the ligand leads to a very small fraction of bound-state ligand due to weak binding affinity. This becomes a problem when RDCs are going to be applied to structure refinement, as contributions by the free-state ligand to the averages can be significant. The difference between the observed RDC average and the free-state RDC of the ligand is within the range of experimental errors (+/-1 Hz). The bound-state RDC extracted using equation (1.2) is not accurate enough. This is particularly a problem when carbohydrates that are

modified to improve their properties are targeted and unreliable structures of lead bound protein are available.

The alignment of the ligand or the receptor protein usually was achieved through the interaction with liquid crystal media aligned in the magnetic field. Most of the drug molecules are lipid soluble and usually carry some hydrophobic moieties. The strong interaction between the ligand and external alignment media prevents the extraction of bound-state RDCs accurately.

The following chapters present improved methods for studying weakly-binding fast-exchanging carbohydrate protein systems using NMR.

References

1. Sharon, N., *Atomic Basis of Protein–Carbohydrate Interactions: An Overview*. Protein-Carbohydrate Interactions in Infectious Diseases., ed. C.A. Bewley. 2006: RSC Publishing.

2. Werz, D.B., et al., *Exploring the structural diversity of mammalian carbohydrates ("Glycospace") by statistical databank analysis.* Acs Chemical Biology, 2007. **2**(10): p. 685-691.

3. Ratner, D.M., et al., *Probing protein-carbohydrate interactions with microarrays of synthetic oligosaccharides.* Chembiochem, 2004. **5**(3): p. 379-382.

4. Sharon, N., *Carbohydrates.* Scientific American, 1980. **243**(5): p. 90-&.

5. Komath, S.S., M. Kavitha, and M.J. Swamy, *Beyond carbohydrate binding: new directions in plant lectin research (vol 4, pg 973, 2006).* Organic & Biomolecular Chemistry, 2006. **4**(8): p. 1611-1611.

6. Elgavish, S. and B. Shaanan, *Lectin-carbohydrate interactions: different folds, common recognition principles.* Trends in Biochemical Sciences, 1997. **22**(12): p. 462-467.

7. Kamp R.M., C.J.J., T. Choli-Papadopoulou, *Methods in Proteome and Protein Analysis*. 2000: Spinger.

8. Mcdonald, I.K. and J.M. Thornton, *Satisfying Hydrogen-Bonding Potential in Proteins.* Journal of Molecular Biology, 1994. **238**(5): p. 777-793.

9. Gaiser, O.J., et al., *Structural basis for the substrate specificity of a Bacillus 1,3-1,4-beta-Glucanase.* Journal of Molecular Biology, 2006. **357**(4): p. 1211-1225.

10. Adhikari, P., et al., *Mutational analysis at Asn-41 in peanut agglutinin - A residue critical for the binding of the tumor-associated Thomsen-Friedenreich antigen.* Journal of Biological Chemistry, 2001. **276**(44): p. 40734-40739.

11. Toone, E.J., *Structure and Energetics of Protein Carbohydrate Complexes.* Current Opinion in Structural Biology, 1994. **4**(5): p. 719-728.

12. Garcia-Hernandez, E. and A. Hernandez-Arana, *Structural bases of lectin-carbohydrate affinities: Comparison with protein-folding energetics.* Protein Science, 1999. **8**(5): p. 1075-1086.

13. Asensio, J.L., et al., *The Interaction of Hevein with N-Acetylglucosamine-Containing Oligosaccharides - Solution Structure of Hevein Complexed to Chitobiose.* European Journal of Biochemistry, 1995. **230**(2): p. 621-633.

14. Asensio, J.L., et al., *NMR investigations of protein-carbohydrate interactions: refined three-dimensional structure of the complex between hevein and methyl beta-chitobioside.* Glycobiology, 1998. **8**(6): p. 569-577.

15. Bouckaert, J., et al., *Receptor binding studies disclose a novel class of high-affinity inhibitors of the Escherichia coli FimH adhesin.* Molecular Microbiology, 2005. **55**(2): p. 441-455.

16. Poveda, A. and J. Jimenez-Barbero, *NMR studies of carbohydrate-protein interactions in solution.* Chemical Society Reviews, 1998. **27**(2): p. 133-143.

17. Levitt, M. and M.F. Perutz, *Aromatic Rings Act as Hydrogen-Bond Acceptors.* Journal of Molecular Biology, 1988. **201**(4): p. 751-754.

18. Gabius, H.J., *The how and why of protein-carbohydrate interaction: A primer to the theoretical concept and a guide to application in drug design.* Pharmaceutical Research, 1998. **15**(1): p. 23-30.

19. Muraki, M., *The importance of CH/pi interactions to the function of carbohydrate binding proteins.* Protein and Peptide Letters, 2002. **9**(3): p. 195-209.

20. Umezawa, Y., et al., *CH/pi interaction in the crystal structure of organic compounds. A database study.* Bulletin of the Chemical Society of Japan, 1998. **71**(5): p. 1207-1213.

21. Umezawa, Y. and M. Nishio, *CH/pi interactions as demonstrated in the crystal structure of guanine-nucleotide binding proteins, Src homology-2 domains and human growth hormone in complex with their specific ligands.* Bioorganic & Medicinal Chemistry, 1998. **6**(4): p. 493-504.

22. Cumpstey, I., et al., *C-2-symmetrical thiodigalactoside bis-benzamido derivatives as high-affinity inhibitors of galectin-3: efficient lectin inhibition through double arginine-*

arene interactions. Angewandte Chemie-International Edition, 2005. **44**(32): p. 5110-5112.

23. Cumpstey, I., et al., *Synthesis of a phenyl thio-beta-D-galactopyranoside library from 1,5-difluoro-2,4-dinitrobenzene: discovery of effcient and selective monosaccharide inhibitors of galectin-7.* Organic & Biomolecular Chemistry, 2005. **3**(10): p. 1922-1932.

24. Tejler, J., H. Leffler, and U.J. Nilsson, *Synthesis of O-galactosyl aldoximes as potent LacNAc-mimetic galectin-3 inhibitors.* Bioorganic & Medicinal Chemistry Letters, 2005. **15**(9): p. 2343-2345.

25. Sorme, P., et al., *Low micromolar inhibitors of galectin-3 based on 3 '-derivatization of N-acetyllactosamine.* Chembiochem, 2002. **3**(2-3): p. 183-189.

26. Shen, J. and L.E. Lerner, *A Comparison of the Conformation of Sucrose Octasulfate, Free and Bound to Acidic Fibroblast Growth-Factor.* Carbohydrate Research, 1995. **273**(2): p. 115-127.

27. Kelm, S., et al., *Use of Sialic-Acid Analogs to Define Functional-Groups Involved in Binding to the Influenza-Virus Hemagglutinin.* European Journal of Biochemistry, 1992. **205**(1): p. 147-153.

28. Sharon, N. and H. Lis, *Lectins - Cell-Agglutinating and Sugar-Specific Proteins.* Science, 1972. **177**(4053): p. 949-&.

29. Watkins, W.M. and W.T.J. Morgan, *Neutralization of the Anti-H Agglutinin in Eel Serum by Simple Sugars.* Nature, 1952. **169**(4307): p. 825-826.

30. Goldstei.Ij and L.L. So, *Protein-Carbohydrate Interaction .3. Agar Gel-Diffusion Studies on Interaction of Concanavalin a a Lectin Isolated from Jack Bean with Polysaccharides.* Archives of Biochemistry and Biophysics, 1965. **111**(2): p. 407-&.

31. Cooper, D.N.W., *Galectinomics: finding themes in complexity.* Biochimica Et Biophysica Acta-General Subjects, 2002. **1572**(2-3): p. 209-231.

32. Angata, T. and E.C.M. Brinkman-Van der Linden, *I-type lectins.* Biochimica Et Biophysica Acta-General Subjects, 2002. **1572**(2-3): p. 294-316.

33. Liu, F.T. and G.A. Rabinovich, *Galectins as modulators of tumour progression.* Nature Reviews Cancer, 2005. **5**(1): p. 29-41.

34. Danguy, A., I. Camby, and R. Kiss, *Galectins and cancer.* Biochimica Et Biophysica Acta-General Subjects, 2002. **1572**(2-3): p. 285-293.

35. Seetharaman, J., et al., *X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution.* Journal of Biological Chemistry, 1998. **273**(21): p. 13047-13052.

36. Aplander, K., et al., *Synthesis of a 3 '-naphthamido-LacNAc fluorescein conjugate with high selectivity and affinity for galectin-3.* Carbohydrate Research, 2006. **341**(10): p. 1363-1369.

37. Carlomagno, T., *Ligand-target interactions: What can we learn from NMR?* Annual Review of Biophysics and Biomolecular Structure, 2005. **34**: p. 245-266.

38. Gemmecker, G., *NMR as a Tool in Drug Research*, in *NMR Spectroscopy in Drug Development and Analysis*, P.D.I.W.D.B.D. Prof. Dr. U. Holzgrabe, Editor. 2007. p. 135-154.

39. Solomon, I., *Relaxation Processes in a System of 2 Spins.* Physical Review, 1955. **99**(2): p. 559-565.

40. Overhauser, A.W., *Polarization of Nuclei in Metals.* Physical Review, 1953. **91**(2): p. 476-476.

41. David Neuhaus, M.P.W., *The Nuclear Overhauser Effect in Structural and Conformational Analysis Second Edition*, in *Methods in Stereochemical Analysis*, A.P. Marchand, Editor. 2000.

42. Fielding, L., S. Rutherford, and D. Fletcher, *Determination of protein-ligand binding affinity by NMR: observations from serum albumin model systems.* Magnetic Resonance in Chemistry, 2005. **43**(6): p. 463-470.

43. Fielding, L., *NMR methods for the determination of protein-ligand dissociation constants.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2007. **51**: p. 219-242.

44. Fielding, L., *Determination of association constants (K-a) from solution NMR data.* Tetrahedron, 2000. **56**(34): p. 6151-6170.

45. Meyer, B. and T. Peters, *NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angewandte Chemie-International Edition, 2003. **42**(8): p. 864-890.

46. Henry, C.M., *Structure-based drug design.* Chemical & Engineering News, 2001. **79**(23): p. 69-+.

47. Shuker, S.B., et al., *Discovering high-affinity ligands for proteins: SAR by NMR.* Science, 1996. **274**(5292): p. 1531-1534.

48. Hajduk, P.J., *SAR by NMR: Putting the pieces together.* Molecular Interventions, 2006. **6**(5): p. 266-+.

49. Medek, A., et al., *The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands.* Journal of the American Chemical Society, 2000. **122**(6): p. 1241-1242.

50. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallographica Section D-Biological Crystallography, 1998. **54**: p. 905-921.

51. Brunger, A.T., *Version 1.2 of the Crystallography and NMR System.* Nature Protocols, 2007(2): p. 2728-2733.

52. Dominguez, C., R. Boelens, and A.M.J.J. Bonvin, *HADDOCK: A protein-protein docking approach based on biochemical or biophysical information.* Journal of the American Chemical Society, 2003. **125**(7): p. 1731-1737.

53. De Vries, S.J., et al., *HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets.* Proteins-Structure Function and Bioinformatics, 2007. **69**(4): p. 726-733.

54. Wagner, G., et al., *Protein Structures in Solution by Nuclear-Magnetic-Resonance and Distance Geometry - the Polypeptide Fold of the Basic Pancreatic Trypsin-Inhibitor Determined Using 2 Different Algorithms, Disgeo and Disman.* Journal of Molecular Biology, 1987. **196**(3): p. 611-639.

55. Bax, A., *Two-Dimensional Nmr and Protein-Structure.* Annual Review of Biochemistry, 1989. **58**: p. 223-256.

56. Wikipedia, t.f.e. *Nuclear Overhauser effect*. [cited; Available from: http://en.wikipedia.org/wiki/Nuclear_Overhauser_effect.

57. Noggle, J.H. and R.E. Schirmer, *The Nuclear Overhauser Effect*, A. Press, Editor. 1971: New York.

58. Jardetzky, O. and G.C.K. Roberts, *NMR in molecular biology*, A. Press, Editor. 1981: New York.

59. Kumar, A., R.R. Ernst, and K. Wuthrich, *A Two-Dimensional Nuclear Overhauser Enhancement (2d Noe) Experiment for the Elucidation of Complete Proton-Proton Cross-Relaxation Networks in Biological Macromolecules.* Biochemical and Biophysical Research Communications, 1980. **95**(1): p. 1-6.

60. Ravindranathan, S., et al., *Transferred cross-relaxation and cross-correlation in NMR: effects of intermediate exchange on the determination of the conformation of bound ligands.* Journal of Magnetic Resonance, 2003. **163**(2): p. 199-207.

61. Sanders, J. and B.K. Hunter, *The Modern NMR Spectroscopy—A Guide for Chemists. 2nd Edition*, in *The Modern NMR Spectroscopy—A Guide for Chemists*. 1993, Oxford University Press: New York. p. 160–176.

62. Hull, W.E., *Experimental Aspects of Two-Dimensional NMR*, in *The Two-Dimensional NMR Spectroscopy—Applications for Chemists and Biochemists (2nd ed.)*, W.R.C.a.R.M.K. Carlson, Editor. 1994, VCH: New York. p. 337–343.

63. Jayalakshmi, V. and N.R. Krishna, *Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes.* Journal of Magnetic Resonance, 2002. **155**(1): p. 106-118.

64. Curto, E.V., H.N.B. Moseley, and N.R. Krishna, *CORCEMA evaluation of the potential role of intermolecular transferred NOESY in the characterization of ligand-receptor complexes.* Journal of Computer-Aided Molecular Design, 1996. **10**(5): p. 361-371.

65. Moseley, H.N.B., E.V. Curto, and N.R. Krishna, *Complete Relaxation and Conformational Exchange Matrix (Corcema) Analysis of Noesy Spectra of Interacting Systems - 2-Dimensional Transferred Noesy.* Journal of Magnetic Resonance Series B, 1995. **108**(3): p. 243-261.

66. Krishna, N.R., H.N.B. Moseley, and E.V. Curto, *Complete Relaxation and Conformational Exchange Matrix (Corcema) Analysis of Noesy Spectra - Application to Transferred Noesy and Protein-Folding Studies.* Journal of Cellular Biochemistry, 1995: p. 74-74.

67. Balaram, P., Bothnerb.Aa, and E. Breslow, *Localization of Tyrosine at Binding-Site of Neurophysin Ii by Negative Nuclear Overhauser Effects.* Journal of the American Chemical Society, 1972. **94**(11): p. 4017-&.

68. Balaram, P., Bothnerb.Aa, and J. Dadok, *Negative Nuclear Overhauser Effects as Probes of Macromolecular Structure.* Journal of the American Chemical Society, 1972. **94**(11): p. 4015-&.

69. Clore, G.M. and A.M. Gronenborn, *Theory of the Time-Dependent Transferred Nuclear Overhauser Effect - Applications to Structural-Analysis of Ligand Protein Complexes in Solution.* Journal of Magnetic Resonance, 1983. **53**(3): p. 423-442.

70. Li, D.W., et al., *Interligand overhauser effects in type II dihydrofolate reductase.* Biochemistry, 2001. **40**(14): p. 4242-4252.

71. Moseley, H.N.B., et al., *Quantitative determination of conformational, dynamic, and kinetic parameters of a ligand-protein/DNA complex from a complete relaxation and conformational exchange matrix analysis of intermolecular transferred NOESY.* Biochemistry, 1997. **36**(18): p. 5293-5299.

72. Mayer, M. and B. Meyer, *Characterization of ligand binding by saturation transfer difference NMR spectroscopy.* Angewandte Chemie-International Edition, 1999. **38**(12): p. 1784-1788.

73. Kalk, A. and H.J.C. Berendsen, *Proton Magnetic-Relaxation and Spin Diffusion in Proteins.* Journal of Magnetic Resonance, 1976. **24**(3): p. 343-366.

74. Jayalakshmi, V. and N.R. Krishna, *CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities.* Journal of Magnetic Resonance, 2004. **168**(1): p. 36-45.

75. Lipsitz, R.S. and N. Tjandra, *Residual dipolar couplings in NMR structure analysis.* Annual Review of Biophysics and Biomolecular Structure, 2004. **33**: p. 387-413.

76. Prestegard, J.H., C.M. Bougault, and A.I. Kishore, *Residual dipolar couplings in structure determination of biomolecules.* Chemical Reviews, 2004. **104**(8): p. 3519-3540.

77. Bax, A., *Weak alignment offers new NMR opportunities to study protein structure and dynamics.* Protein Science, 2003. **12**(1): p. 1-16.

78. Sanders, C.R., et al., *Magnetically-Oriented Phospholipid Micelles as a Tool for the Study of Membrane-Associated Molecules.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1994. **26**: p. 421-444.

79. Sanders, C.R. and J.P. Schwonek, *Characterization of Magnetically Orientable Bilayers in Mixtures of Dihexanoylphosphatidylcholine and Dimyristoylphosphatidylcholine by Solid-State Nmr.* Biochemistry, 1992. **31**(37): p. 8898-8905.

80. Clore, G.M., M.R. Starich, and A.M. Gronenborn, *Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses.* Journal of the American Chemical Society, 1998. **120**(40): p. 10571-10572.

81. Hansen, M.R., L. Mueller, and A. Pardi, *Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions.* Nature Structural Biology, 1998. **5**(12): p. 1065-1074.

82. Ishii, Y., M.A. Markus, and R. Tycko, *Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel.* Journal of Biomolecular Nmr, 2001. **21**(2): p. 141-151.

83. Sass, H.J., et al., *Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes.* Journal of Biomolecular Nmr, 2000. **18**(4): p. 303-309.

84. Sass, J., et al., *Purple membrane induced alignment of biological macromolecules in the magnetic field.* Journal of the American Chemical Society, 1999. **121**(10): p. 2047-2055.

85. Ruckert, M. and G. Otting, *Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments.* Journal of the American Chemical Society, 2000. **122**(32): p. 7793-7797.

86. Tjandra, N. and A. Bax, *Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium.* Science, 1997. **278**(5340): p. 1111-1114.

87. Barbieri, R., et al., *Paramagnetically induced residual dipolar couplings for solution structure determination of lanthanide binding proteins.* Journal of the American Chemical Society, 2002. **124**(19): p. 5581-5587.

88. Gochin, M. and H. Roder, *Protein-Structure Refinement Based on Paramagnetic Nmr Shifts - Applications to Wild-Type and Mutant Forms of Cytochrome-C.* Protein Science, 1995. **4**(2): p. 296-305.

89. Prudencio, M., et al., *A caged lanthanide complex as a paramagnetic shift agent for protein NMR.* Chemistry-a European Journal, 2004. **10**(13): p. 3252-3260.

90. Bertini, I., C. Luchinat, and M. Piccioli, *Paramagnetic probes in metalloproteins.* Nuclear Magnetic Resonance of Biological Macromolecules, Pt B, 2001. **339**: p. 314-340.

91. Wohnert, J., et al., *Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings.* Journal of the American Chemical Society, 2003. **125**(44): p. 13338-13339.

92. Ikegami, T., et al., *Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions.* Journal of Biomolecular Nmr, 2004. **29**(3): p. 339-349.

93. Ma, C. and S.J. Opella, *Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy.* Journal of Magnetic Resonance, 2000. **146**(2): p. 381-384.

94. Su, X.C., et al., *Site-specific labelling of proteins with a rigid lanthanide-binding tag.* Chembiochem, 2006. **7**(10): p. 1599-1604.

95. Bax, A., G. Kontaxis, and N. Tjandra, *Dipolar couplings in macromolecular structure determination.* Nuclear Magnetic Resonance of Biological Macromolecules, Pt B, 2001. **339**: p. 127-174.

96. de Alba, E. and N. Tjandra, *NMR dipolar couplings for the structure determination of biopolymers in solution.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2002. **40**(2): p. 175-197.

97. Prestegard, J.H., H.M. Al-Hashimi, and J.R. Tolman, *NMR structures of biomolecules using field oriented media and residual dipolar couplings.* Quarterly Reviews of Biophysics, 2000. **33**(4): p. 371-424.

98. Ottiger, M., F. Delaglio, and A. Bax, *Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra.* Journal of Magnetic Resonance, 1998. **131**(2): p. 373-378.

99. Kay, L.E., P. Keifer, and T. Saarinen, *Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity.* Journal of the American Chemical Society, 1992. **114**(26): p. 10663-10665.

100. Kontaxis, G., G.M. Clore, and A. Bax, *Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times.* Journal of Magnetic Resonance, 2000. **143**(1): p. 184-196.

101. Tian, F., P.J. Bolon, and J.H. Prestegard, *Intensity-based measurement of homonuclear residual dipolar couplings from CT-COSY.* Journal of the American Chemical Society, 1999. **121**(33): p. 7712-7713.

102. Jain, N.U., S. Noble, and J.H. Prestegard, *Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings.* Journal of Molecular Biology, 2003. **328**(2): p. 451-462.

103. Bertini, I., C. Luchinat, and G. Parigi, *Paramagnetic constraints: An aid for quick solution structure determination of paramagnetic metalloproteins.* Concepts in Magnetic Resonance, 2002. **14**(4): p. 259-286.

104. Koenig, S.H., *A Classical Description of Relaxation of Interacting Pairs of Unlike Spins - Extension to T1p, T2, and T1poff, Including Contact Interactions.* Journal of Magnetic Resonance, 1982. **47**(3): p. 441-453.

105. Bertini, I., et al., *Cross correlation between the dipole-dipole interaction and the curie spin relaxation: The effect of anisotropic magnetic susceptibility.* Journal of Magnetic Resonance, 2001. **152**(1): p. 103-108.

106. Mcgarvey, B.R., *Theory of Contact and Pseudo Contact Shift in Tetrahedral Nickel(Ii) Complexes - Inor.* Abstracts of Papers of the American Chemical Society, 1970: p. 32-&.

107. Pintacuda, G., et al., *NMR structure determination of protein-ligand complexes by lanthanide labeling.* Accounts of Chemical Research, 2007. **40**(3): p. 206-212.

108. Bertini, I., C. Luchinat, and G. Parigi, *Solution NMR of Paramagnetic Molecules*. 2001, Amsterdam.: Elsevier.

109. Banci, L., et al., *Paramagnetism-based restraints for Xplor-NIH.* Journal of Biomolecular Nmr, 2004. **28**(3): p. 249-261.

110. Bertini, I., et al., *Cross correlation rates between Curie spin and dipole-dipole relaxation in paramagnetic proteins: The case of Cerium substituted Calbindin D-9k.* Journal of Biomolecular Nmr, 2002. **23**(2): p. 115-125.

111. Goldman, M., *Interference Effects in the Relaxation of a Pair of Unlike Spin-1/2 Nuclei.* Journal of Magnetic Resonance, 1984. **60**(3): p. 437-452.

112. Muraki, M., M. Ishimura, and K. Harata, *Interactions of wheat-germ agglutinin with GlcNAc beta 1,6Gal sequence.* Biochimica Et Biophysica Acta-General Subjects, 2002. **1569**(1-3): p. 10-20.

113. Moore, J.M., *NMR techniques for characterization of ligand binding: Utility for lead generation and optimization in drug discovery.* Biopolymers, 1999. **51**(3): p. 221-243.

Chapter 2

Enhancement of Bound-State Residual Dipolar Couplings: Conformational Analysis of Lactose Bound to Galectin-3

2.1 Introduction

Defining the conformation of ligands as they exist when bound to proteins, and defining the structural interactions that exist between bound ligands and residues in the active site of a protein provide a useful starting point for the rational design of many drugs. NMR has proven to be a major source of experimental of information needed for defining these properties [1]. Transfer Nuclear Overhauser Effects (NOE) [2-4] and saturation transfer difference (STD) experiments [5, 6] have been particularly effective at providing information on the properties of bound ligands. These are both based on cross-relaxation between pairs of spins via a dipole-dipole interaction. The steep r^{-6} distance dependence of cross-relaxation provides both contact information and distance constraints useful in conformational analysis. For rapidly exchanging systems, these experiments are typically applied with a 10-20 fold excess of ligand over protein, a condition which makes observation of cross-relaxation involving ligand protons a highly sensitive process. An average of free and bound-state cross relaxation effects is observed, but the dependence of cross relaxation efficiency on a correlation time that scales with the size of the complex allows bound state contributions to dominate to an extent that makes analysis of bound state conformations straightforward [7]. Nevertheless, there are limitations. Data only provide constraints for proton-proton pairs with distance separations shorter than 4-5 Å. For ligands with small numbers of such inter-residue pairs and for proteins that use hydrogen-bonding arrays as

opposed to van der Waals contacts for recognition, the number of constraints is often inadequate to define a bound geometry. This is particularly a problem for carbohydrate-recognizing proteins.

For cases with small numbers of NOE constraints, residual dipolar couplings (RDCs) provide a potentially useful alternative in that they are orientational constraints, independent of distances between protons on ligand residues and the protein. However, applications to exchanging systems have been few [8, 9] because there is no preferential weighting of bound state information. Here we provide one solution to this problem. It involves a simple modification of the protein to enhance its orientation, a factor that is required for measurement of RDCs and a factor that scales up contributions from any bound ligand.

Dipolar couplings between pairs of spin ½ nuclei display the angular dependence shown in equation 2.1 [10, 11],

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h}{(2\pi r)^3} \langle \frac{3\cos^2\theta - 1}{2} \rangle \qquad (2.1)$$

where θ is the angle between the magnetic field and the spin-spin interaction vector and r is the distance between interacting spins. In our case the pairs will be directly bonded or be in a semi-rigid ring structure for which the interaction distance is known and θ becomes the primary variable. The difficulty is the averaging that takes place with molecular tumbling in solution. When an isotropic distribution of vector directions is sampled, the function goes to zero and no dipolar couplings are observed [11]. The solution most frequently involves use of a liquid crystal medium. Typical media for molecules that dissolve in aqueous environments are phospholipid bicelles, alkyl-ethyleneglycol-alcohol bilayers (C12E5-hexanol, for example) and bacterio-phage [11-13]. These media normally cause departures from isotropic orientational distributions through collisional interactions with anisotropically shaped molecules. The dipolar

contribution does not go to zero and adds to normal through-bond J-couplings where they are measured as departures in multiplet splittings from their isotropic values.

The problem for exchanging systems is that the anisotropy of shape is independent of size and the induced RDCs for a free ligand can be just as large as the induced RDCs for a ligand in a protein complex. Observed RDCs are given by equation 2.2. If RDCs for the free state (RDC_F) and bound state (RDC_B) are similar, and the fraction of ligand in the bound state (f_B) is of order 0.1, it will clearly be difficult to extract RDC_B from a measured coupling. This had been tested in a mannose-binding protein trimannoside complex [8] and our system as data shown in column 4 of table 2.1. The difference between the observed average and the free-state RDC is usually below measurement accuracy (1Hz). The bound-state RDCs extracted are not reliable. These inaccurate data could lead to incorrect interpretation of the ligand orientation.

$$RDC_{obs}=f_F*RDC_F + f_B*RDC_B \qquad (2.2)$$

One approach to resolving this problem is to make RDC_B larger than RDC_F. At least one attempt to do this has been reported in the literature; it relies on using paramagnetic alignment of a protein, rather than a liquid crystal medium. RDC_F in this case is nearly equal to zero and RDC_B can dominate [14]. This is something that we will return to in chapter 4. The approach we present here relies on making RDC_B large so measured RDCs will be easier to measure. The magnitude of an RDC_B is directly dependent on the magnitude of alignment of the complex. In our case we will enhance the alignment of the protein (and protein-ligand complex) by switching from collisional alignment to specific association with elements of the liquid crystal medium. This has precedent in the membrane attached protein system studied by Koenig et al to measure the dipolar couplings of a transducin peptide fragment weakly bound to photo-activated rhodopsin [15]. The alignment medium was Rhodopsin-containing disks which were isolated

from rod outer segments of bovine retina limited the application to other weakly binding ligand-protein systems [15]. Our medium is a C12E5 medium with aligned elements having a lipid-bilayer-like hydrophobic interior. Association is promoted by addition of a hydrophobic alkyl tail to the protein. The addition of a hydrophobic alkyl chain to a carbohydrate disaccharide had been applied to the study of glycosidic torsional motions using residual dipolar couplings by Yi et al [16] . The order of the alignment of the carbohydrate can be significantly enhanced through specifically anchoring the hydrophobic tail to the hydrophobic interior of the alignment media. The order of the alignment varies with the length of the alkyl tail. The magnitude of RDCs of a butylated disaccharide is more than 10 fold that of a methylated disaccharide under the same experimental conditions. The same strategy is applied to the galectin-3 CRD. The length of the alkyl chain varies from methyl to propyl. Only a short propyl chain proves necessary to sufficiently enhance the bound-state RDCs. It will be added to the terminus of a protein by adding a cysteine residue and using a cysteine specific reaction as shown below (Figure 2.1) [17-19] . The details of this reaction will be described later.

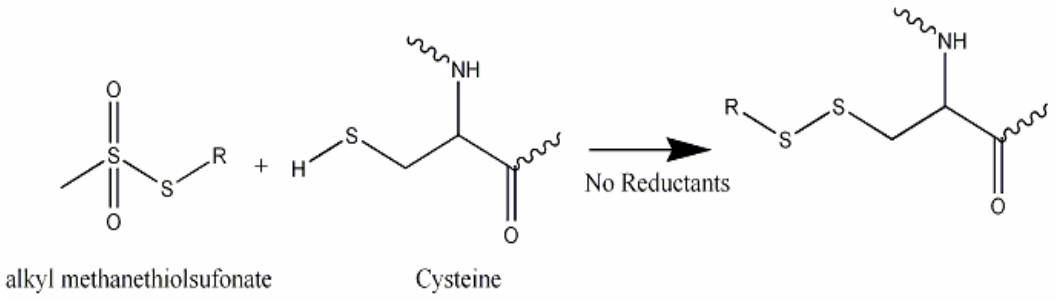

Figure 2.1: Cysteine specific alkylation with alkyl methanethiolsufonate under a reductant free condition.

The protein and ligand we use as an illustration of the methods proposed are the galectin-3 Carbohydrate Recognition Domain (CRD) (PDB ID: 1A3K) and lactose respectively [20]. Galectin-3 has two functional domains including an N-terminal domain and a C-terminal CRD (residues117-250). The CRD of galectin-3 is evolutionarily conserved within the galectin family and carries a galactose-specific binding site [20, 21]. It is this domain that we study here. Although the CRD of galectin-3 already has one cysteine (C173), this cysteine is buried and quite inaccessible to the solvent (<5%, calculated by NACCESS) [22, 23], we expect it to be unreactive to a bulky reagent. Another reactive cysteine is introduced by adding the nucleic acids coding for this amino acid to the C-terminus of the cDNA for the domain. Because the C-terminus is flexible and far away from the binding site it is unlikely to change the protein conformation or interfere with ligand binding. The CRD has been characterized by NMR previously [24] and we have previously studied its carbohydrate binding properties, focusing more on protein structural response to ligand binding [25]. Here we present a study focused on the conformation of a bound ligand (lactose) as opposed to that of the protein.

2.2 Materials and Methods

2.2.1 Preparation of the Galectin-3 C251- CRD Plasmid

Since we are interested in only the CRD of galectin-3, the polymerase chain reaction (PCR) was applied to amplify the desired sequence from a pET-3c plasmid that encodes intact galectin-3 [26] to generate a construct encoding mainly the CRD (aa 114-250) with a new start site (Met) replacing aa 113, and a new cysteine introduced at the C-terminus (position 251) by adding the codon for this amino acid to the appropriate primer. Two restriction sites (Nde1 & BamH1) are also included in the primers. The primers used in PCR were: 5'-

AACGAGCGGCATATGCTGATTGTGCC-3' (forward) and 5'-*GCGGGATCCTCAACATATCATGG*-3' (reverse) (IDT, IL) [26].

The amplified PCR product and vector pET-9a (Novagen) were double-digested by Nde1 and BamH1 and purified using a Qiaquick Gel Extraction Kit (Qiagen). The double-digested galectin-3 CRD gene was then ligated into the pET-9a vector with a molar ratio of the galectin-3 CRD gene and the PET-9a vector of 3:1 using T4 DNA ligase (Promega).

2.2.2 Protein Expression and Purification

Galectin-3 C251-CRD was expressed in *Escherichia coli* BL21 (DE3) cells using the following protocol for expressing the ^{15}N isotopically labeled protein: After transformation with the plasmids containing the galectin-3 C251-CRD gene and plating on kanamycin agar plates, one colony was selected to inoculate 50 mL LB medium containing 100 ug/ml kanamycin. This was incubated at 37ºC overnight with shaking at 225-250 rpm. The cells were harvested and transferred to 1 L pre-sterilized M9 minimal media with 1g $^{15}NH_4Cl$ (added for ^{15}N labeling). The medium also contained 6.5 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 4% glucose. The culture was then incubated in a 37ºC incubator with shaking at 225-250 rpm. Monitoring cell growth by measuring the absorbance at wavelength 600 nm, 1 mL of 1 mM isopropyl-β,D-thiogalactopyranoside (IPTG) was added to induce the protein expression when the absorbance reached 0.9 ± 0.1. The cells were then grown for additional 2-4 hours and were spun down in a Sorvall SL-34 rotor at 5000 rpm for 30 min.

To extract the protein, the harvested cell pellet was resuspended in a lysis buffer containing 75 mM KH_2PO_4,75 mM NaCl, 2 mM EDTA, 5 mM NaN_3 ,1 mM DTT and 4 mM β-mecaptoethanol (pH=7.4) [27]. The lysates were then spun down in an ultracentrifuge (Beckman) at 30,000 rpm for 30 min and the supernatant was loaded onto a lactosyl-agarose affinity

column[27]. The column was washed with 250 mL lysis buffer to remove extraneous proteins and galectin-3 C251-CRD was eluted with 100 mL lysis buffer+ 300 mM lactose. The protein was concentrated using an Amicon concentrator (Millipore) and the concentrated solution was stored at -20^{o}C. Typically 25-40 mg pure protein per liter can be obtained.

2.2.3 Preparation of Galectin-3 S-propyl-C251-CRD

Protein alkylthiolation had to be performed under reductant-free conditions. Hence, the reductants present during isolation (DTT & β-mecaptoethanol) were first removed by ultracentrifugal filtration (Millipore) using reductant-free, N_2-saturated buffer C (75 mM KH_2PO_4, 75 mM NaCl, 2 mM EDTA,5 mM NaN_3, pH=7.4); the initial sample (4 mL) was reduced to 0.5 mL and 3.5 mL of buffer C was added in 3-4 successive cycles. Propyl methanethiolsulfonate (PMTS) (Toronto Research Chemicals) was dissolved in the same reductant-free buffer C. The protein was then mixed with PMTS in a stoichiometry of 1:2 and incubated at room temperature for 10~15 min. The excess PMTS was removed by buffer exchange (reductant-free buffer C) using ultracentrifugal filtration (Millipore) immediately after the reaction. One of the advantages of this reaction is the ability to recover protein in its native form by adding appropriate reductants to cleave the newly-formed disulfide bond.

The extent of reaction was assayed by LC-MS. A Perkin Elmer Sciex API I plus quadrupole mass spectrometer (Applied Biosystems) was used for the assay. A molecular ion was easily identified for both monomeric and dimeric species. There was an additional peak displaced by 74 mass units for the monomeric species. This was characteristic of addition of a single thiopropane group. Purification of alkylthiolated protein by size exclusion chromatography can be used to remove a small amount of dimer, however this was not a necessary step for the current application.

2.2.4 NMR Spectroscopy

All NMR experiments were performed on a Varian Inova 800 MHz spectrometer equipped with a cryogenic triple resonance probe. Two parallel experiments were performed for each sample: one in isotropic phosphate buffer (75 mM KH_2PO_4, 75 mM NaCl, 2 mM EDTA, 5 mM NaN_3, pH=7.4, 10% ^2H_2O) and one in an aligned 3%C12E5 bicelle medium prepared in the same buffer. The aligned medium was made approximately as described by Ruckert et al [12]. In general 50 ul of C12E5 (Sigma) was dissolved in 500 ul of phosphate buffer and hexanol (app. 12 ul) was added in 3 ul increments until the medium became clear and viscous. An equal volume of protein in buffer was added to produce a final sample approximately 0.5 mM in protein and 3% in bicelle. For measurements on ligand, unlabeled lactose (Sigma) was first dissolved in phosphate buffer to make a stock solution at 200 mM and then added to protein containing samples to reach a final concentration of 5 mM. A control set of samples containing just lactose in isotropic buffer and bicelle medium was also prepared. Homogeneous alignment of C12E5 containing samples was ascertained by monitoring the deuterium quadrupole splitting of deuterons on $^2H^1HO$ in all cases.

For the measurement of ^1H-^{15}N backbone RDCs of the protein, an interleaved set of ^1H-^{15}N HSQC and TROSY experiments was run [28, 29]. Offsets of paired peaks in the ^{15}N direction were taken to be ½ the ^{15}N-^1H coupling. Differences in measurements in isotropic and aligned sets were taken to be RDCs. Normally 256 t1 quadrature pairs were acquired over 3000 Hz with 36 acquisitions each. The entire acquisition required 12 hours.

For the measurement of ligand ^1H-^{13}C RDCs a sensitivity-enhanced ^1H-^{13}C coupled HSQC was used [30]. Three sets of experiments were carried out to get the dipolar couplings of the ligands; these included an isotropic and aligned spectrum of lactose only and an aligned

spectrum of the lactose-protein mixture. Generally, 512 t1 points over 9500 Hz were acquired. Acquisitions were recorded for 18 hr with 48 acquisitions per t1 point. These long acquisitions are required since all ^{1}H-^{13}C couplings are measured at nature abundance.

In addition to the collection of ^{1}H-^{13}C couplings, ^{1}H-^{1}H couplings were obtained using the intensity-based constant-time ^{1}H-^{1}H COSY experiment [31]. An array of constant times from 0.1 s to 0.42 s with time increment 0.04 s was used for collecting the ^{1}H-^{1}H couplings of lactose alone, both in phosphate buffer and aligned in 3% C12E5 bicelle medium. For the ^{1}H-^{1}H couplings of a lactose-protein mixture, an array of constant times from 0.018 s to 0.153 s with time increment 0.015 s was used. Spectra used 32 t1 quadrature pairs over 2000 Hz and acquisitions required 14 hrs with 30 acquisitions per t1 pair.

All the data were processed using NMRPipe [32]. For measuring the ^{1}H-^{15}N couplings of the protein, NMRPipe autoPeak picking and Nonlinear Lineshape Fitting (nLinLS) were used. Errors in the final RDCs are estimated at ±0.5 Hz. For measuring the ^{1}H-^{13}C couplings of the ligands, the Bayesian fitting program, XRAMBO [33], was used to extract splittings from the f1 dimension of the coupled spectra. Errors in the final RDCs are estimated at ±0.5 Hz. For ^{1}H-^{1}H couplings, the CT-COSY was processed in Magnitude mode and the data were fit to the function, $I_{cross}/I_{auto}=A|\tan((J+D)*\pi*t))|$ using routines in Matlab 6.0. Here I_{cross} and I_{auto} are the intensity of cross-peaks and auto-peaks respectively. The precision of couplings is estimated at ±0.1Hz.

Analysis of RDCs in terms of molecular structure and orientation was conducted with the program REDCAT [34]. The molecular coordinates from the crystal structure of the wild type CRD (PDB 1A3K) [20] were used with hydrogens being added to the structure prior to analysis using the program Chimera [35].

2.2.5 Ligand Docking to Galectin-3 CRD

The structure of lactose-bound galectin-3 CRD was determined using HADDOCK, an experimental data driven docking program. Unlike other methods, HADDOCK drives the docking process based on the experimental chemical shift perturbation or biochemical mutagenesis data rather on energetics and the shape complementarities. The chemical shift perturbation data or mutagenesis data are defined as ambiguous interaction restraints (AIR), i.e., an ambiguous distance between all residues shown to be involved in the interaction.

The active residues were determined based on two criteria: significant chemical shift perturbation and high solvent accessibility. No chemical shift perturbation experiment was carried out in our case. Instead we used the known crystal structure and the NMR structure to identify the residues that form hydrogen bonds or have substantial van der Waals contacts with the ligand as the residues having significant chemical shifts changes. The residues were further filtered by the solvent accessibility. NACESS was used to calculate the solvent accessibility of the protein based on the crystal structure. Those residues that were more than 40% solvent accessible (i.e. W181, N164, E184, R186, N222) were identified as active residues. The passive residues were those residues that neighbored the active residues and had more than 40% solvent accessibility. They were N180, N179, G182, R183, N167, N166, R168, N143 and D207. For the ligand lactose, from the crystal structure, both galactose and glucose were involved in the recognition of the lactose ligand and were regarded as active residues.

The AIR restraints were defined as the distance between any atom of an active residue of one molecule to any atom of an active or passive residue of the second molecules. An example can be seen in Appendix A. Normally, the average AIR is set to be 2 Å with upper distance limit 4 Å.

In addition to the AIR restraints, RDCs of the protein and carbohydrate were included as the orientation restraints. The order tensor elements (Sxx, Syy, Szz) calculated from REDCAT were used for the determination of Da and R which were required for the SANI module. The simple relationship between order tensor elements, Da and R follows the equations (2.3 and 2.4):

$$D_a = \frac{D_{\max}}{2} S_{zz} \qquad (2.3)$$ (Here D_{max} is 24350 for N-H, -60400 for C-H and -240200 for H-H)

$$R = \frac{S_{xx} - S_{yy}}{S_{zz}} \qquad (2.4)$$

Due to the lack of intramolecular NOEs for lactose, the pyranose ring of glucose and galactose ring were fixed to their most stable chair forms during the whole docking procedure. The dihedral angles (ψ and ϕ) of the glycosidic bond from an energy-minimized geometry of lactose (ϕ=50^o, ψ=-116^o) generated using the GLYCAM_04 force field were restrained to be 50^o and -116^o with variation +/-2^o. Here ψ is defined by H1 (GAL)-C1 (GAL)-O4 (GLC)-C4 (GLC) and ϕ by C1 (GAL)-O4 (GLC)-C4 (GLC)-C5 (GLC).

The whole docking procedure consisted of three main steps: rigid body energy minimization, semi-flexible simulation annealing and flexible refinement in explicit water. During semi-flexible simulated annealing, the side chains of those residues in the binding interfaces were set to be free. One thousand structures were generated after rigid-body energy minimization and 200 structures with minimal energy were chosen for semi-flexible simulated annealing. Following the second step, 50 structures were chosen for the final flexible refinement in explicit water. The RMSD of the ligand found by superimposing the protein was calculated from 15 structures with minimal energy after the final step. The simulated structure was validated by checking the violation of the hydrogen bonding, van der Waals contacts and experimental restraints.

2.3 Results

2.3.1 Cysteine Specific Alkylthiolation of Galectin-3 C251-CRD

Both the addition of the C-terminal cysteine and the formation of a disulfide bond between this cysteine and 1-propanethiol could lead to unwanted perturbation of protein properties. Moreover, the galectin-3 CRD has a native cysteine that we preferred to keep as a free sulfhydryl**.** To avoid undesirable outcomes, we added the cysteine and directed the alkythiolation at the C-terminal end because the C-terminus of the galectin-3 CRD appears to be an unstructured segment in the crystal structure (PDB 1A3K) and this segment is well removed from the carbohydrate binding site.

The lack of significant perturbation upon addition of the cysteine can easily be confirmed by comparing ^{15}N-^{1}H HSQC spectra for mutant and native proteins. ^{15}N and ^{1}H chemical shifts are quite sensitive to local conformation changes and cross-peaks in these spectra from residues near perturbed regions should move. Figure 2.2 presents a comparison of spectra for native and mutant C251-CRDs of galectin-3. The native spectrum is reconstructed from the work by Umemoto et al. [24] and the C251-CRD spectrum was taken as described in the methods section. With few exceptions the two spectra are identical within the errors expected from small differences in buffer and other environmental factors. The major differences are for the amino acids of the N-terminus and C-terminus. The difference at the N-terminus is explainable because the galectin-3 CRD constructs differ slightly in the number of pre-CRD residues (before aa 117). In the work by Umemoto et al., the CRD was prepared by collagenase digestion to yield a protein containing aa 108-250, whereas our construct encodes a protein containing aa 114-250 plus an added start Met (aa 113) and the C –terminal cysteine (aa 251). Our galectin-3 C251-CRD contains only 138 amino acids (114-251) if the Met is cleaved. The difference at the C-

terminus arises from our addition of cysteine-251. From the similarity in shifts of peaks from other residues, we can conclude that the CRD conformation was not changed significantly by the introduction of a cysteine at the C-terminus.

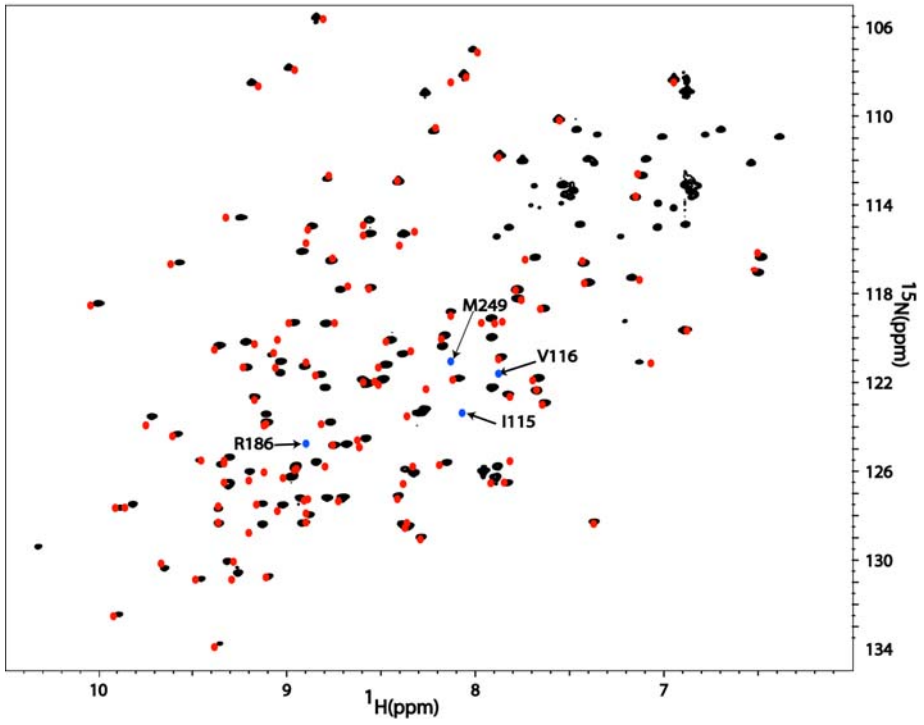

Figure 2.2: Overlay of the ^{1}H-^{15}N HSQC spectrum of galectin-3 C251-CRD (aa 114-251) (black) with the native galectin-3 CRD (red). The spectrum of galectin-3 CRD-CRD was collected at 800 MHz, as a 1 mM sample in phosphate buffer, pH 7.4. The spectrum of the native galectin-3 CRD was reconstructed based on the deposited chemical shifts from BMRB (accession number: 4909) Amino acids 108-113 were not plotted here for easy comparison. The blue peaks represent peaks where no partner can be found.

We can also draw a definitive conclusion with respect to the issue of modification of the internal cysteine. Cysteine (C173) is actually buried and inaccessible to the solvent based on the crystal structure [23] and an NACCESS calculation (< 5%) [22]. So, it is safe to assume that this

cysteine would not readily alkylate under native folded conditions for the protein. To confirm this, we used a combination of mass spectrometry and NMR spectroscopy to prove that the reaction occurred only at the C-terminal cysteine. Figure 2.3A shows the molecular ion region of an LC-MS spectrum for the C251 protein before the reaction. The major peak is at an appropriate mass for the protein. The minor peak with higher mass is very likely from reaction with some residual β-mercaptoethanol in the buffer. Figure 2.3 B shows the protein after the reaction of the C251 protein with a 1:2 ratio of protein to PMTS as described in the Methods section. The major peak now appears to be displaced to a higher mass by 76 Da. This is very close to the theoretical value of 74 Da expected for the addition of a single thiopropane chain. Only a very minor peak occurs at a mass that could correspond to the addition of two thiopropane groups. This suggests that only one of the two cysteines is reacting.

The particular cysteine involved, C251 or C173, can easily be identified using NMR spectra. ^{1}H-^{15}N HSQC spectra were collected for both galectin-3 S-propyl-C251-CRD and galectin-3 C251-CRD. These are superimposed in Figure.2.4. The cross-peak arising from C173 is easily assigned based on published assignments for the native protein. There is essentially no shift of this cross-peak between spectra. Three peaks were observed to have chemical shift changes. Two of them are new peaks that arose after the addition of C251 to the native protein; one is probably I250 which was not assigned in the native protein. M249 also has a small chemical shift change after the alkylthiolation. This is strong evidence that only the newly-added cysteine (C251) is involved in the alkylthiolation and that the remainder of the protein is not perturbed by the addition of the thiopropane chain.

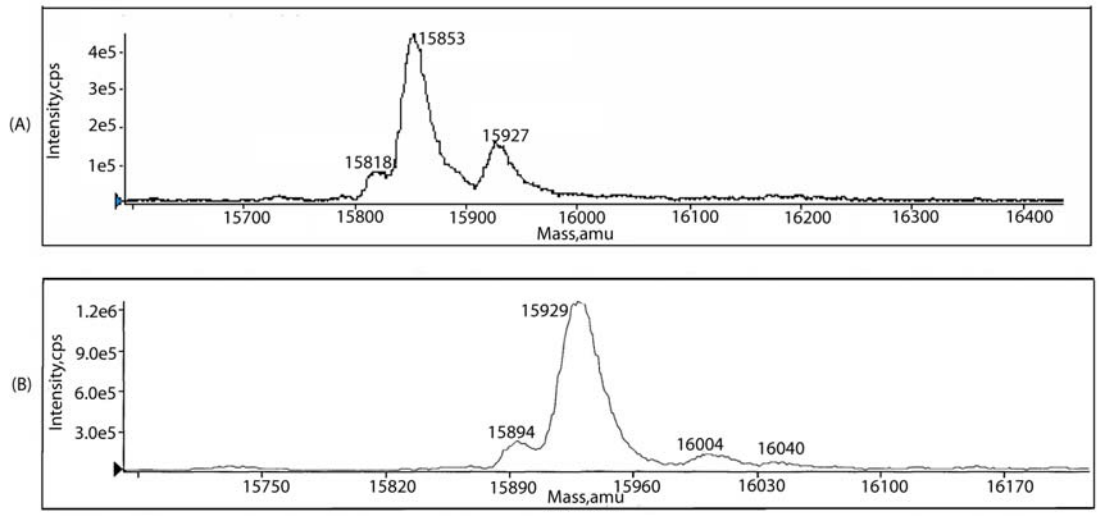

Figure 2.3: LC-MS spectra. (A) ^{15}N-galectin-3 C251- CRD (MW~15852 Da) in phosphate buffer (75 mM KH_2PO_4, 75 mMNaCl, 2 mM EDTA, 5 mM NaN_3 ,1 mM DTT and 4 mM β-mecaptoethanol, pH=7.4). (B) ^{15}N-galectin-3 S-propyl-C251-CRD (MW~15926 Da) in phosphate buffer (75 mM KH_2PO_4, 75 mMNaCl, 2 mM EDTA, 5 mM NaN_3 ,no DTT& β-mecaptoethanol, pH=7.4).

2.3.2 Assessment of Protein Alignment in a Bicelle Medium

We had anticipated that the addition of a thiopropane chain to the galectin-3 CRD would increase the degree of association with the bicelles of our orientation media and enhance the size of RDCs for both the protein and the associated ligand. We first evaluated this increase for the protein, both to validate its association and to provide data that provides an orientation for the protein that can then be used as a reference for the orientation of the ligand. Figure 2.5 depicts the variation in ^{1}H-^{15}N RDCs measured for galectin-3 C251-CRD (red bars) and galectin-3 S-propyl-C251-CRD (blue bars) in a 3% C12E5 bicelle medium. Note that the RDCs measured in the galectin-3 S-propyl-C251-CRD and the galectin-3 C251-CRD are significantly different. The range of RDCs for galectin-3 S-propyl-C251-CRD (+/- 40 Hz) is almost three time that seen in the C251-galectin-3-CRD (+/-15 Hz) sample.

Analysis of these data was conducted using the program REDCAT [34] and the crystal structure of galectin-3 CRD (PDB 1A3K) [20]. Given a molecular structure and RDC data, REDCAT uses a singular value decomposition algorithm to solve for the elements of an order tensor which best represents the RDC data. It then diagonalizes the tensor to give principle order parameters and orientations of axes of a principle alignment frame. It repeats this thousands of times randomly selecting combinations of RDCs within estimated error limits of measurements. In our case we used 40 RDCs excluding measurements from highly overlapped regions and regions known to be in flexible loops. An error estimate of 2 Hz was used for galectin-3-CRD-C251 and 7 Hz for galectin-3 S-propyl-C251-CRD. These errors represent both structural imperfections (estimated at 10% of the RDC range) and estimated experimental errors. The best set of principal order parameters found for the thiopropyl protein are Sxx (-2.7E-06), Syy (1.2E-03) and Szz (-1.2E-03). These are approximately three times the values calculated from data on the unmodified protein. Again this confirms enhanced alignment.

The distribution of directions for allowed alignment frames are depicted in Figure 2.6A for galectin-3 S-propyl-C251-CRD. The spread gives an indication of the precision in determining the orientation of protein alignment. REDCAT also can back-calculate RDCs given elements of an order tensor and a probable structure. For the galectin-3 S-propyl-C251-CRD a plot of experimental versus back-calculated RDCs is given in Figure 2.6 B. The correlation is reasonably good. This correlation can be quantitated in terms of a Q factor [10, 36]. This is 0.21 for our case, something representative of agreement with the structure to within 2 Å over the back-bone atoms [10]. This is confirmation that the modified protein is not significantly structurally perturbed.

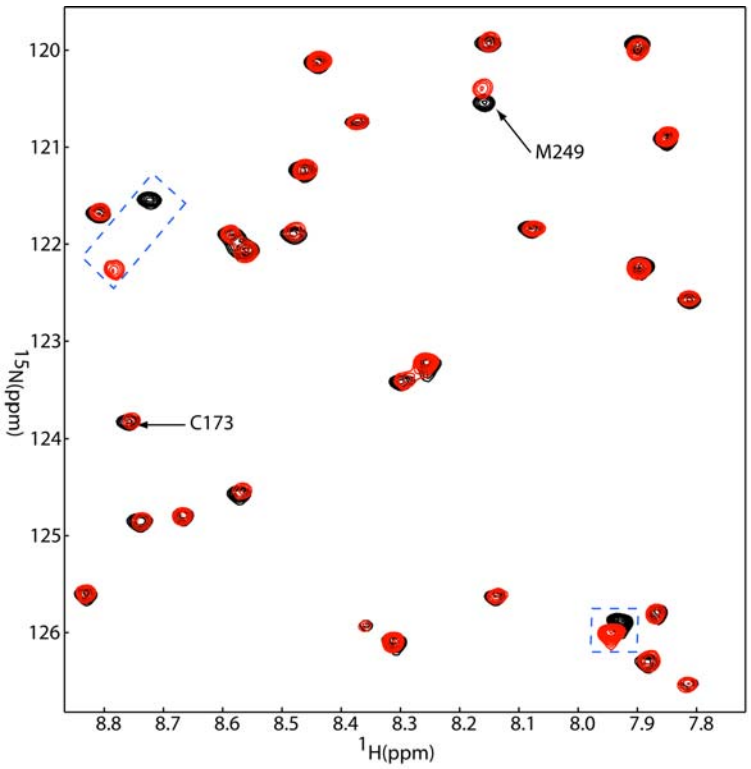

Figure 2.4: Overlay of the ^{1}H-^{15}N HSQC spectra of 1 mM galectin-3 C251-CRD (red) on spectra of galectin-3 S-propyl-C251-CRD (black) in 75 mM phosphate buffer (pH=7.4). Spectra recorded at 25^{o}C with a ^{1}H resonance frequency of 800 MHz. C173 chemical shift remains unchanged. Three peaks (blue squares) were observed to have chemical shift changes where M249 shows a small degree of chemical shift variance and the other two were new peaks after the addition of C251.

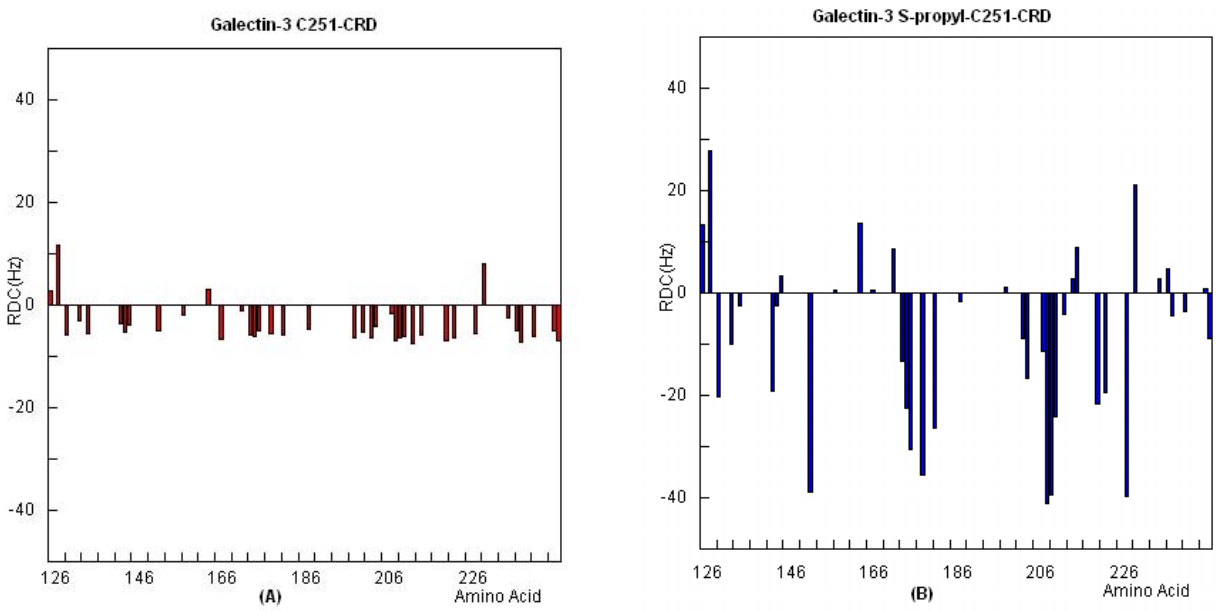

Figure 2.5: RDC distributios Plots: (A) RDCs from galectin-3 C251-CRD (red bars); (B) RDCs from S-propyl-C251-galectin-3 CRD (blue bars). 40 RDCs were collected and plotted against amino acids number (125-247) from left to right.

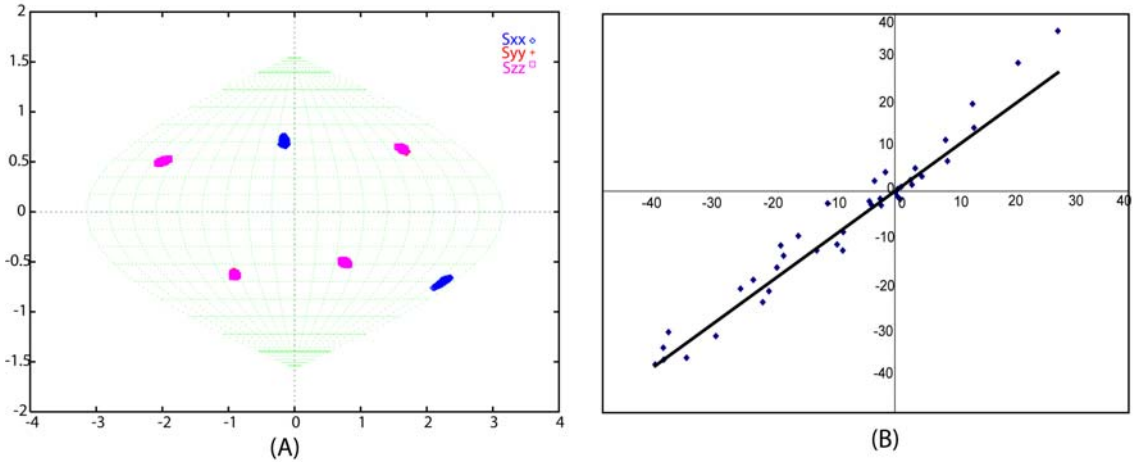

Figure 2.6: (A) Sauson-Flamsted projection of allowed directions for alignment frame axes; (B) Correlation plot of experimentally measured RDCs vs. back-calculated RDCs of galectin-3 S-propyl-C251- CRD based on the crystal structure of galectin-3 CRD using REDCAT.

2.3.3 Assessment of Ligand Alignment

Ligand alignment can be assessed by the measurement of one-bond ^{13}C-^{1}H and intra-ring ^{1}H-^{1}H RDCs. ^{13}C-^{1}H couplings are easily measured with an f1- coupled ^{13}C-^{1}H HSQC, even at natural abundance, because the ligand is in large excess over the protein and an averaged spectrum is measured. From lactose we expect to see 14 one-bond ^{13}C-^{1}H couplings, seven from the glucose ring and seven from the galactose ring. There may be some resolution of sets from the alpha and beta anomers, especially for sites near the anomeric carbon of glucose. However, we do not expect couplings to be different except at the anomeric carbon, because we do not expect them to bind with significant differences, as the anomeric site is distant from the part of galactose recognized by galectin-3. While the number of couplings that could be measured is significant, five of the ^{13}C-^{1}H vectors in β-glucose and four in galactose are nearly parallel to one another because they occupy axial positions on the preferred $^{1}C_4$ ring conformation. Also, two more vectors in each ring (H_6-C_6) are hard to use because of potential internal motions. α-glucose is somewhat better than β-glucose because of the equatorial anomeric ^{13}C-^{1}H pair. However, the minimal number of completely independent vectors necessitated an attempt to measure intra-ring ^{1}H-^{1}H couplings. This was done with a constant time COSY experiment [31].

Examples of segments from the coupled HSQC spectra are presented in Figure 2.7 for the C_4-H_4 and C_5-H_5 pairs from the galactose ring. Doublets are shown for the ligand under isotropic conditions (A), aligned in the absence of protein (B), aligned in the presence of galectin-3 C251-CRD (C) and aligned in the presence of galectin-3 S-propyl-C251-CRD (D). The RDCs (seen as the difference between isotropic and aligned conditions) are obviously measurable, but small for lactose in the absence of protein. They remain small and not very different for lactose in the presence of galectin-3 C251-CRD. This is the result of the influence of the excess free lactose in

the average RDC measured, and the fact that the RDCs are not significantly different in bound and free forms of lactose in this case. The situation is different in the presence of galectin-3 S-propyl-C251-CRD. The couplings are larger, and while the lines broaden, couplings can still be accurately measured. More importantly, the observed RDCs are heavily weighted by the bound state RDCs in the presence of the propyl modification.

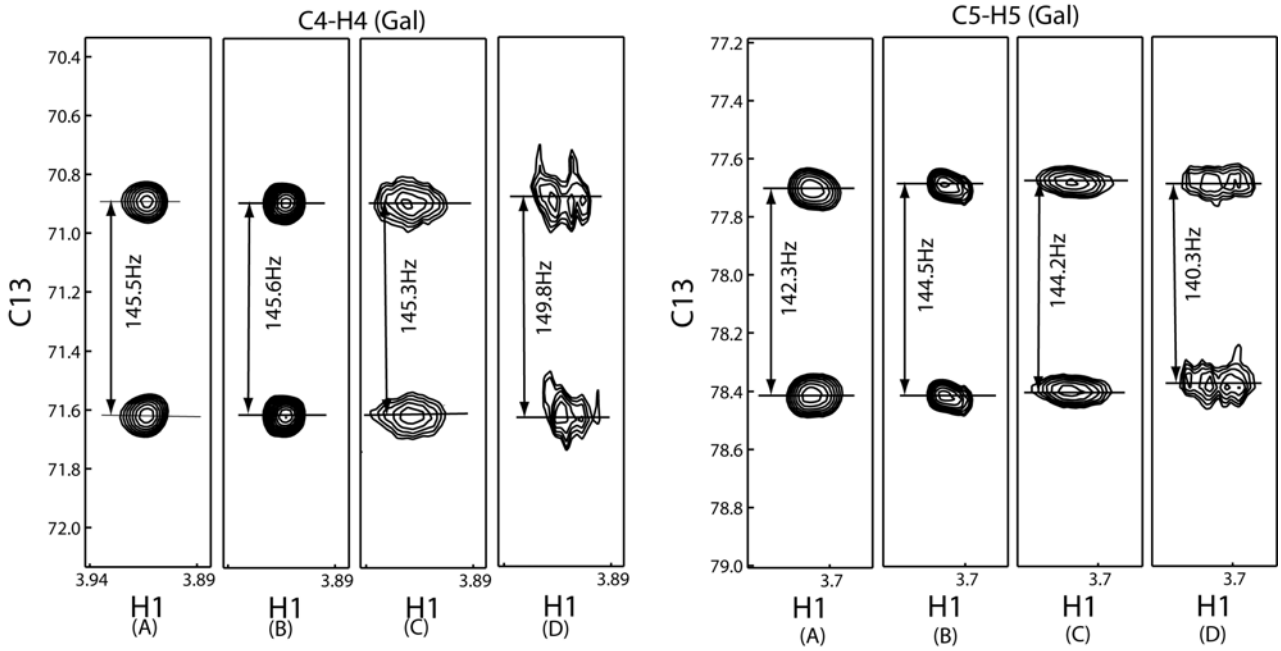

Figure 2.7: Examples of measurement couplings of lactose from coupled ^{1}H-^{13}C HSQC spectra. (A) lactose in phosphate buffer; (B) lactose in the absence of protein in 3%C12E5; (C) lactose in the presence of galectin-3 C251-CRD in 3%C12E5 10:1 ratio; (D) lactose in the presence of galectin-3 S-propyl-C251-CRD in 3%C12E5 10:1 ratio. All spectra were collected at 298 k with a proton resonance frequency of 800 MHz. All samples were prepared using 75 mM phosphate buffer (PH=7.4).

Table 2.1 presents a summary of ligand RDCs. The bound state values have been separated from the average measurements using published data on lactose binding constants [20]. We have used REDCAT and an energy-minimized geometry of lactose (φ=50^{o}, ψ=-116^{o}) generated by GLYCAM_04 force field [37] to extract an order tensor and back-calculate RDCs

from the bound state data. Note that except for one case where measurement was difficult because of overlap with media resonances, overall agreement is very good. Also note that the principal order parameters extracted (-1.0E-03 for S_{zz}) agree very well with the order parameters extracted for the bicelle-bound galectin-3 S-propyl-C251-CRD itself. This suggests that the lactose is well modeled by the conformation used and that the there is relatively little motion in the binding site.

2.4 Discussion

2.4.1 Orientation of Galectin-3 S-propanyl-C251-CRD on Bicelle Surfaces

The ^{1}H-^{15}N RDC data collected on galectin-3 thiopropanyl-C251-CRD in an oriented bicelle medium confirms retention of an active structure for the modified protein and provides information on the actual orientation of the protein in the medium. The latter is useful for understanding the nature of the protein's interaction with the medium and for providing a reference for understanding ligand orientation. The level of order is nearly of equal in magnitude (but opposite in sign) for z and y axes, but order along the z axis of the principal alignment frame (PAF) is slightly greater. The order parameter for this axis also is negative. Order parameters derived from RDCs can be considered to be the average of $(3\cos^2\theta'-1)/2$ where θ' is the instantaneous angle between a molecular PAF axis and the magnetic field. To make this negative for the z axis the angle relative to the magnetic field must have a preference for 90°. Bilayer-like elements also are known to orient with their normals near 90° with respect to the magnetic field. As we believe the elements of the C12E5 medium are bilayer-like, we can construct a model for protein- bilayer interaction as depicted in Figure 2.8. The PAF is defined in the coordinate system of the molecule (PDB 1A3K) and is depicted as the set of axes in the figure. The bilayer is drawn with its normal parallel to the PAF z axis.

Table 2.1: RDCs of free-state and bound-state lactose at a ^{1}H frequency 800 MHz spectrometer

			Galectin-3 C251 CRD	Galectin-3 S-propyl-C251-CRD		
		RDC^{a}_{free}(Hz)	RDC^{a}(Hz)	RDC^{a}(Hz)	RDC^{b}_{bound}(Hz)	RDC^{c}_{cal}(Hz)
Glc(α)	C1-H1	-2.8	-3.1	-3.7	-12	-12
	C2-H2	2.3	3.3	0.8	-13	-19
	C3-H3	3.9	3.1	0.2	-33	-27
	C5-H5	4.2	3.8	2	-19	-17
	H1-H2	-0.1	0.2	0.6	15	15
Gal	C1-H1	-0.2	0.7	-3.5	-33	-36
	C2-H2	3.0	2.3	-0.6	-33	-37
	C3-H3	1.7	1.8	-3.5	-51*	-31
	C4-H4	0.1	-0.2	4.3	42	42
	C5-H5	2.2	1.9	-2.0	-40	-35
	H1-H2	0.2	0.4	-0.4	-11	-8

Order tensor elements	Sxx	Syy	Szz
Bound-state lactose	2.6E-05	1.0E-03	-1.0E-03
S-Propyl Protein	2.7E-06	1.2E-03	-1.2E-03

* Deviation due to peak overlap with peaks from alignment media C12E5. Peak picking manually
a Peak splittings measured by *XRAMBO* with error ±0.5 Hz
b Bound state RDCs were calculated using equation 1 with fraction of bound-state 10%
c Back-calculated RDCs using *REDCAT*

Note that the orientation of the molecule projects the C-terminal propyl chain along the bilayer normal, in a manner that would allow insertion into the hydrophobic core. Actually we cannot distinguish this orientation from those produced by the rotation of the protein by 180º about the x or y axis of the PAF on the basis of RDCs alone, but these other structures would seem physically unreasonable. The structure is certainly not static or permanently fixed to the bilayer. If it were, the order parameters would approach 1×10^{-1} instead of 1×10^{-3}. There is very likely an equilibrium between free and associated states. The level of order can be

increased by modifying the protein with longer alkyl chains, but these would produce more line broadening and may prohibit collecting data that allows determination of protein orientation.

2.4.2 Geometry of Bound Lactose

The RDC data on bound lactose allows determination of molecular orientation. As mentioned above the similarity in magnitudes of order parameters suggests minimal mobility of the ligand within the binding pocket and opens the possibility of building a model with lactose docked in the protein binding site. In principle, orientational constraints are provided by the knowledge that the PAFs, as seen from the point of view of the two molecules, must coincide if they are part of the same rigid complex. Thus the ligand can be positioned on the protein by

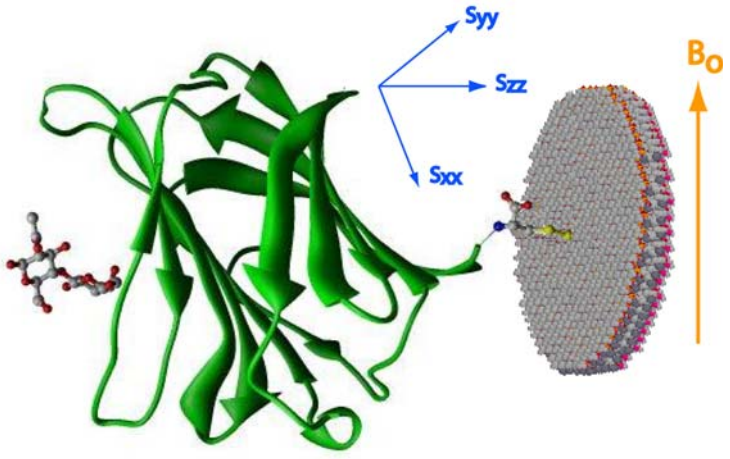

Figure 2.8: Model for association between an alkylated protein and a bilayer-like medium. The propyl chain embeds into the hydrophobic inner layer. The structure of the protein was created based on known crystal structure (1A3K). Chimera [35] was used to optimize propyl bond angles and make the graphic.

translational movements only. In practice we used the program HADDOCK to accomplish docking [38]. HADDOCK was initially developed for docking protein-protein complexes, but with slight modification of the input topology and parameter files of the carbohydrate, it also

works for protein-carbohydrate complexes. Here both glucose and galactose residues of lactose were considered to be active residues. The active and passive residues in the protein were identified based on the crystal structure of galectin-3 CRD (PDB 1A3K) [20] and the approach of residues to the C4 region of the terminal galactose. This program allows the translation of a ligand relative to a protein under orientational constraints. The molecule is translated to optimize a list of contacts and non-contacts for the two molecules without having van der Waals violations.

The determined structure (around the binding site) is depicted in Figure 2.9. The picture shows good van der Walls contacts between a tryptophan and the more hydrophobic sides of galactose and glucose. H-bonding contacts with the 4-hydroxyl have been well maintained. In both the crystal structure and our docked structure, the C-4 hydroxyl group is in a position where it can form hydrogen bonds with H158, R162 and D160. There are, however, some differences outside of this primary interaction region. In particular, the lactose is rotated by about 30° toward a contact with tryptophan in our structure. The crystal structure has a gap between the ligand and the protein at this site. Of course, the ligands are not identical: the crystal structure contained an N-acetyl-lactosamine and the docked structure contained lactose. In the crystal structure, one can see interactions between the acetyl group and the protein that may be leading to this orientational difference.

We have demonstrated an ability to amplify bound-state RDCs in situations where ligands are exchanging rapidly on and off a protein and only average parameters are measured. This puts RDC measurements on a par with STD and trNOE measurements commonly used to determine bound state geometries of protein ligands. Not only is the bound state emphasized, but the sensitivity of measurements is adequate to examine ligands that have not been

isotopically labeled, even when the primary data are ^{13}C-^{1}H couplings. The RDC information is orientational in nature as opposed to distance dependent. For oligosaccharide ligands this is useful when few NOEs across glycosidic bonds can be observed. Even though we have not done so here, glycosidic torsional angles in the bound state could be determined from the orientation of individual rings in an oligosaccharide when sufficient data are available.

In our experiment, the enhanced orientation was obtained by modifying the protein by mutagenesis and subsequently adding a short alkyl chain. Other strategies are similar to this in principle [14, 39], but the ability to vary association by lengthening the alkyl chain is important. The larger the association, the more the bound state data on the ligand will dominate, and the less protein in the ligand:protein mix can be used. This will be particularly advantageous in other applications where data on protein orientation is not required, and the protein need not be isotopically labeled.

The method we developed allows for the accurate measurement of bound-state RDCs of weakly-bound ligand in nature abundance by specifically enhancing the alignment of the protein. This approach is applicable to other weakly-binding ligand-protein system with high off-rates. The limitation of this method is that the ligand must not be very hydrophobic. When ligands are strongly hydrophobic, the ligand and alignment medium interact to such a degree that assessment of bound-state RDC is difficult.

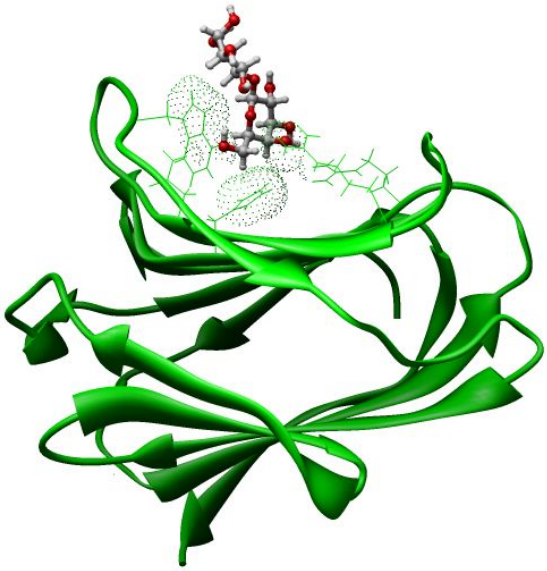

Figure 2.9: Lactose-galectin-3 CRD complex docked using RDCs constraints

References

1. Moore, J.M., *NMR techniques for characterization of ligand binding: Utility for lead generation and optimization in drug discovery.* Biopolymers, 1999. **51**(3): p. 221-243.

2. Poveda, A. and J. Jimenez-Barbero, *NMR studies of carbohydrate-protein interactions in solution.* Chemical Society Reviews, 1998. **27**(2): p. 133-143.

3. Campbell, A.P. and B.D. Sykes, *The 2-Dimensional Transferred Nuclear Overhauser Effect - Theory and Practice.* Annual Review of Biophysics and Biomolecular Structure, 1993. **22**: p. 99-122.

4. Post, C.B., *Exchange-transferred NOE spectroscopy and bound ligand structure determination.* Current Opinion in Structural Biology, 2003. **13**(5): p. 581-588.

5. Mayer, M. and B. Meyer, *Characterization of ligand binding by saturation transfer difference NMR spectroscopy.* Angewandte Chemie-International Edition, 1999. **38**(12): p. 1784-1788.

6. Meyer, B. and T. Peters, *NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angewandte Chemie-International Edition, 2003. **42**(8): p. 864-890.

7. Moseley, H.N.B., et al., *Quantitative determination of conformational, dynamic, and kinetic parameters of a ligand-protein/DNA complex from a complete relaxation and conformational exchange matrix analysis of intermolecular transferred NOESY.* Biochemistry, 1997. **36**(18): p. 5293-5299.

8. Jain, N.U., S. Noble, and J.H. Prestegard, *Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings.* Journal of Molecular Biology, 2003. **328**(2): p. 451-462.

9. Lipsitz, R.S. and N. Tjandra, *Residual dipolar couplings in NMR structure analysis.* Annual Review of Biophysics and Biomolecular Structure, 2004. **33**: p. 387-413.

10. Bax, A., *Weak alignment offers new NMR opportunities to study protein structure and dynamics.* Protein Science, 2003. **12**(1): p. 1-16.

11. Prestegard, J.H., C.M. Bougault, and A.I. Kishore, *Residual dipolar couplings in structure determination of biomolecules.* Chemical Reviews, 2004. **104**(8): p. 3519-3540.

12. Ruckert, M. and G. Otting, *Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments.* Journal of the American Chemical Society, 2000. **122**(32): p. 7793-7797.

13. Gaemers, S. and A. Bax, *Morphology of three lyotropic liquid crystalline biological NMR media studied by translational diffusion anisotropy.* Journal of the American Chemical Society, 2001. **123**(49): p. 12343-12352.

14. Ikegami, T., et al., *Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions.* Journal of Biomolecular Nmr, 2004. **29**(3): p. 339-349.

15. Koenig, B.W., et al., *Measurement of dipolar couplings in a transducin peptide fragment weakly bound to oriented photo-activated rhodopsin.* Journal of Biomolecular Nmr, 2000. **16**(2): p. 121-125.

16. Yi, X.B., et al., *Glycosidic torsional motions in a bicelle-associated disaccharide from residual dipolar couplings.* Journal of the American Chemical Society, 2004. **126**(42): p. 13636-13638.

17. Bruice, T.W. and G.L. Kenyon, *Alkyl Alkanethiolsulfonate Sulfhydryl-Reagents - Beta-Sulfhydryl-Modified Derivatives of L-Cysteine as Substrates for Trypsin and Alpha-Chymotrypsin.* Bioorganic Chemistry, 1985. **13**(2): p. 77-87.

18. Smith, D.J., E.T. Maggio, and G.L. Kenyon, *Simple Alkanethiol Groups for Temporary Blocking of Sulfhydryl Groups of Enzymes.* Biochemistry, 1975. **14**(4): p. 766-771.

19. Mascia, M.P., J.R. Trudell, and R.A. Harris, *Specific binding sites for alcohols and anesthetics on ligand-gated ion channels.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(16): p. 9305-9310.

20. Seetharaman, J., et al., *X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution.* Journal of Biological Chemistry, 1998. **273**(21): p. 13047-13052.

21. Houzelstein, D., et al., *Phylogenetic analysis of the vertebrate galectin family.* Molecular Biology and Evolution, 2004. **21**(7): p. 1177-1187.

22. Hubbard, S.J. and J.M. Thornton, *NACCESS.* Computer Program, Department of Biochemistry and Molecular Biology, University College London, 1993.

23. Frigeri, L.G., M.W. Robertson, and F.T. Liu, *Expression of Biologically-Active Recombinant Rat Ige-Binding Protein in Escherichia-Coli.* Journal of Biological Chemistry, 1990. **265**(34): p. 20763-20769.

24. Umemoto, K. and H. Leffler, *Letter to the Editor: Assignment of H-1, N-15 and C-13 resonances of the carbohydrate recognition domain of human galectin-3.* Journal of Biomolecular Nmr, 2001. **20**(1): p. 91-92.

25. Umemoto, K., et al., *Conformational differences in liganded and unliganded states of Galectin-3.* Biochemistry, 2003. **42**(13): p. 3688-3695.

26. Massa, S.M., et al., *L-29, an Endogenous Lectin, Binds to Glycoconjugate Ligands with Positive Cooperativity.* Biochemistry, 1993. **32**(1): p. 260-267.

27. Leffler, H., F.R. Masiarz, and S.H. Barondes, *Soluble Lactose-Binding Vertebrate Lectins - a Growing Family.* Biochemistry, 1989. **28**(23): p. 9222-9229.

28. Kay, L.E., P. Keifer, and T. Saarinen, *Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity.* Journal of the American Chemical Society, 1992. **114**(26): p. 10663-10665.

29. Kontaxis, G., G.M. Clore, and A. Bax, *Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times.* Journal of Magnetic Resonance, 2000. **143**(1): p. 184-196.

30. Schleucher, J., et al., *A General Enhancement Scheme in Heteronuclear Multidimensional Nmr Employing Pulsed-Field Gradients.* Journal of Biomolecular Nmr, 1994. **4**(2): p. 301-306.

31. Tian, F., P.J. Bolon, and J.H. Prestegard, *Intensity-based measurement of homonuclear residual dipolar couplings from CT-COSY.* Journal of the American Chemical Society, 1999. **121**(33): p. 7712-7713.

32. Delaglio, F., et al., *Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes.* Journal of Biomolecular Nmr, 1995. **6**(3): p. 277-293.

33. Andrec, M. and J.H. Prestegard, *Metropolis Monte Carlo implementation of Bayesian time-domain parameter estimation: Application to coupling constant estimation from antiphase multiplets.* Journal of Magnetic Resonance, 1998. **130**(2): p. 217-232.

34. Valafar, H. and J.H. Prestegard, *REDCAT: a residual dipolar coupling analysis tool.* Journal of Magnetic Resonance, 2004. **167**(2): p. 228-241.

35. Pettersen, E.F., et al., *UCSF chimera - A visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.

36. Cornilescu, G. and A. Bax, *Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase.* Journal of the American Chemical Society, 2000. **122**(41): p. 10143-10154.

37. Woods, R.J., *GLYCAM_04.* http://glycam.ccrc.uga.edu/index.html.

38. Dominguez, C., R. Boelens, and A.M.J.J. Bonvin, *HADDOCK: A protein-protein docking approach based on biochemical or biophysical information.* Journal of the American Chemical Society, 2003. **125**(7): p. 1731-1737.

39. Wohnert, J., et al., *Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings.* Journal of the American Chemical Society, 2003. **125**(44): p. 13338-13339.

Chapter 3

Bound State Residual Dipolar Couplings for Rapidly Exchanging Ligands of His-Tagged Proteins[1]

3.1 Introduction

Historically, structure-based drug design incorporates library screening and lead optimization technologies, which are guided by the geometry of a ligand bound protein complex [2, 3]. The binding network, derived from the bound-state conformation of the ligand and three-dimensional structure of the receptor protein, determines the specificity of molecular recognition and binding affinity. In the case of fast-exchanging, weakly-binding systems, NOE-based NMR methods (such as trNOEs), have proven especially useful in determining the bound-state conformation of the ligand [2, 4]. These methods are applicable to large proteins under rather dilute conditions as long as the ligand is present in substantial quantities (i.e. at least ten-fold molar excess). The NOE is distance dependent and offers information related to distance constraints between pairs of protons on ligands that lie less than 5 Å apart [5]. However, a sufficient number of NOE distance constraints are required for the accurate determination of the bound-state geometry and it is not always possible to observe that many trNOEs. The intramolecular NOE does not provide the distance between the ligand and its receptor. And, the intermolecular trNOEs, in the case of carbohydrate-protein systems, are very difficult to observe due to the dominance of hydrogen bonding networks. NOEs help clarify local but not global geometry. In these cases, alternative local or long-distance constraints are usually needed.

Residual dipolar coupling (RDC) has been applied widely to structure determination and refinement because it provides data that describe the orientation between remote domains of the protein. These long-range structural constraints are useful for studying ligand-protein interactions when sufficient NOEs are unavailable. There have been a few applications of RDCs to the determination of bound ligand geometry [6, 7]. However, the extraction of accurate bound-state RDCs for weak binding systems has been very difficult due to the dominance of free ligands in the observed average RDCs. Recently, we employed an alternate approach to enhancing the RDC contribution of protein-bound ligands by anchoring the protein to a C12E5/hexanol alignment medium by adding a short alkyl chain to the amino terminus of the protein [8]. Although the RDCs enhancement is significant, this approach is limited when the cysteines are functionally conserved and by the fact that only one reactive cysteine is allowed in the protein. When multiple reactive cysteines are present, alkylthiolation can introduce multiple alkyl chains on the protein. This alkylthiolated protein can adopt different orientation preferences which complicate the interpretation of the order tensors and reference frame rotation. To overcome this complication, we have developed a method that utilizes the presence of strong electrostatic interactions between His-tagged protein and a nickel-chelated liquid crystal media. This approach, originally proposed by Dr. Ron Seidel [1], provides a universal method for enhancing bound-state RDCs and is described below.

For large complexes, the efficiency of the magnetization transfer producing the NOE is proportional to the molecular weight of the complex. The NOEs from a ligand bound in a 100 kDa complex are heavily weighted and easily dominate NOEs from a 1000 Da free ligand, even when the ligand is present in excess (e.g. ten-fold). This is not the case for residual dipolar couplings. Measurement of RDCs in solution requires weak alignment in order to keep dipolar

couplings from averaging to zero. This weak alignment is most often achieved through use of nematic liquid crystalline media. The dissolved macromolecule and ligand bias their isotropic orientation (a few tenths of a percent) through weak collisional or steric interactions with the medium. Unlike trNOEs, these interactions often depend on size-independent properties, such as anisotropy of shape. As a result, measured RDCs of ligands are nearly equally weighted by populations of bound and free species. The bound-state RDCs show no substantial difference from the free-state RDCs. Therefore, the bound-state RDCs that are extracted might not accurately reflect the conformation and orientation changes once the ligand is bound.

In addition to our own studies [8], there have been a few attempts to accentuate RDC contributions of the bound ligand by relying on orientational properties specifically originating in the proteins. For example, the large magnetic anisotropy of ordered peptide fibrils [9], ordered protein assemblies [10], and proteins with paramagnetic centers [11, 12] exhibit free-state ligand RDCs of zero and the observed RDCs simply reflect the bound-state values. Since this approach relies upon intrinsic properties of the protein, it is successful only when magnetic anisotropies exist naturally or where proteins can be engineered to contain highly anisotropic metal centers. Measured RDCs also are very small in most of these cases. Other line-broadening effects introduce additional difficulties for precise measurement.

In chapter 2, we described a process of attaching alkyl chains to the substrate protein to enhance the order of the preferred orientation. Here we examine the possibility of increasing the weighting of the bound state information by associating the target protein with an alignment medium using a poly-histidine-tag (His-tag) added to the protein terminus.

Many proteins can tolerate the addition of a poly-Histidine tag (3-6 Histidines) to the C- or N-terminal without significant structural perturbation or alteration of biological function. His-

tag addition is a very common modification of proteins and is routinely used for affinity purification by Immobilized Metal Affinity Chromatography (IMAC). Purification proceeds through ionic interactions of poly-Histidine residues with chelated nickel (NTA-Ni), bound to an immobile resin. The binding affinity of His-tag to nickel ion can be reduced through competitive binding of imidazole or the alteration of protonation of the imidazole moiety of histidine by changing the pH. In general, binding affinity increases when pH increase and vice versa. Recently, the same chelating agent (NTA-Ni) that is used in purification columns has been covalently attached to fatty acyl chains [13] to provide a chelate that can reside on membrane-like surfaces (DOGS-NTA). DOGS-NTA is available in two forms: 1,2-dioleoyl-*sn*-glycero-3-{[N(5-amino-1-carboxypentyl) iminodiacetic acid]succinyl} (nickel salt)) [14] and ammonium salt. The structure of DOGS-NTA ammonium salt is shown as Figure 3.1. The commercial DOGS-NTA nickel salt might have nickel ion overloaded. The excess of nickel ion can cause some non-specific interactions with the protein or compete with DOGS-NTA-Ni for the poly-Histidine tag. The 1:1 DOGS-NTA nickel salt can be prepared by mixing nickel chloride with DOGS-NTA ammonium salt with stoichiometry 1:1.

NTA is a chelating agent that can form several bonds to some transient metal ions such as Ni^{2+}, and Co^{2+}. The coordination number of Ni^{2+} usually is 6. Among all 6 ligand atoms four come from the NTA and the other two are from water, assuming that there are no other chelating agents present. When a His-tagged protein is mixed with DOGS-NTA nickel salt, the two water molecules are readily substituted by the side chain nitrogen atom of two Histidines. Figure 3.2 depicts this process.

The commercial availability of DOGS-NTA-Ni and ammonium salts has encouraged the use of immobilizing His-tagged proteins on lipid surfaces, a method for two-dimensional protein

crystallization [13, 15]. The lipid-like structure of DOGS-NTA allows the incorporation of these nickel-chelating lipids into liquid crystal media, such as the bicelles used for sample alignment in RDC measurements [16]. The strong electrostatic interaction between the His-tag and nickel ion may allow for the enhancement of His-tagged protein alignment, and improved RDC measurement for bound ligands.

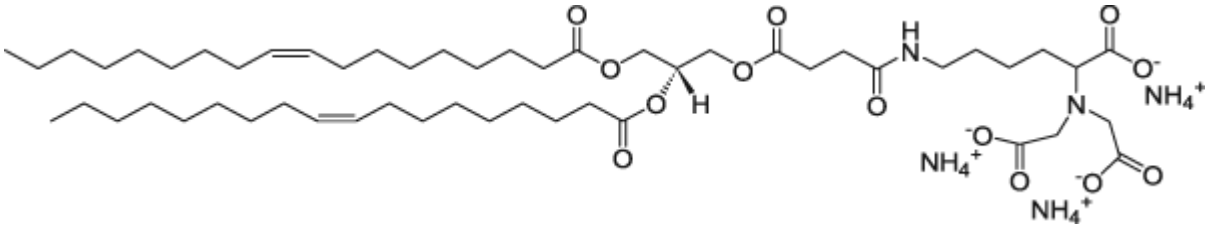

Figure 3.1 the chemical structure of DOGS-NTA ammonium salt

To validate our hypothesis about the measurement of enhanced RDCs, we employed the galectin-3 system discussed in the previous chapter. Galectin-3 is one of the well-studied galectins and it specifically recognizes galactose-terminated oligosaccharides [17]. The expression of galectin-3 is frequently seen in the inhibition of apoptosis, progression of cancer, and mediation of inflammation through the interaction with a variety of galactose-contained carbohydrates [18, 19]. The structures of galectin-3 bound with different carbohydrates are of considerable interest in drug discovery. The galectin-3 carbohydrate recognition domain has a crystal structure with the ligand N-acetyl lactosamine in the binding site [17] as well as an NMR structure with lactose in the binding site [20]. The next section describes how we applied the nickel chelated lipid to determine the bound geometry of a simple disaccharide, α-lactose.

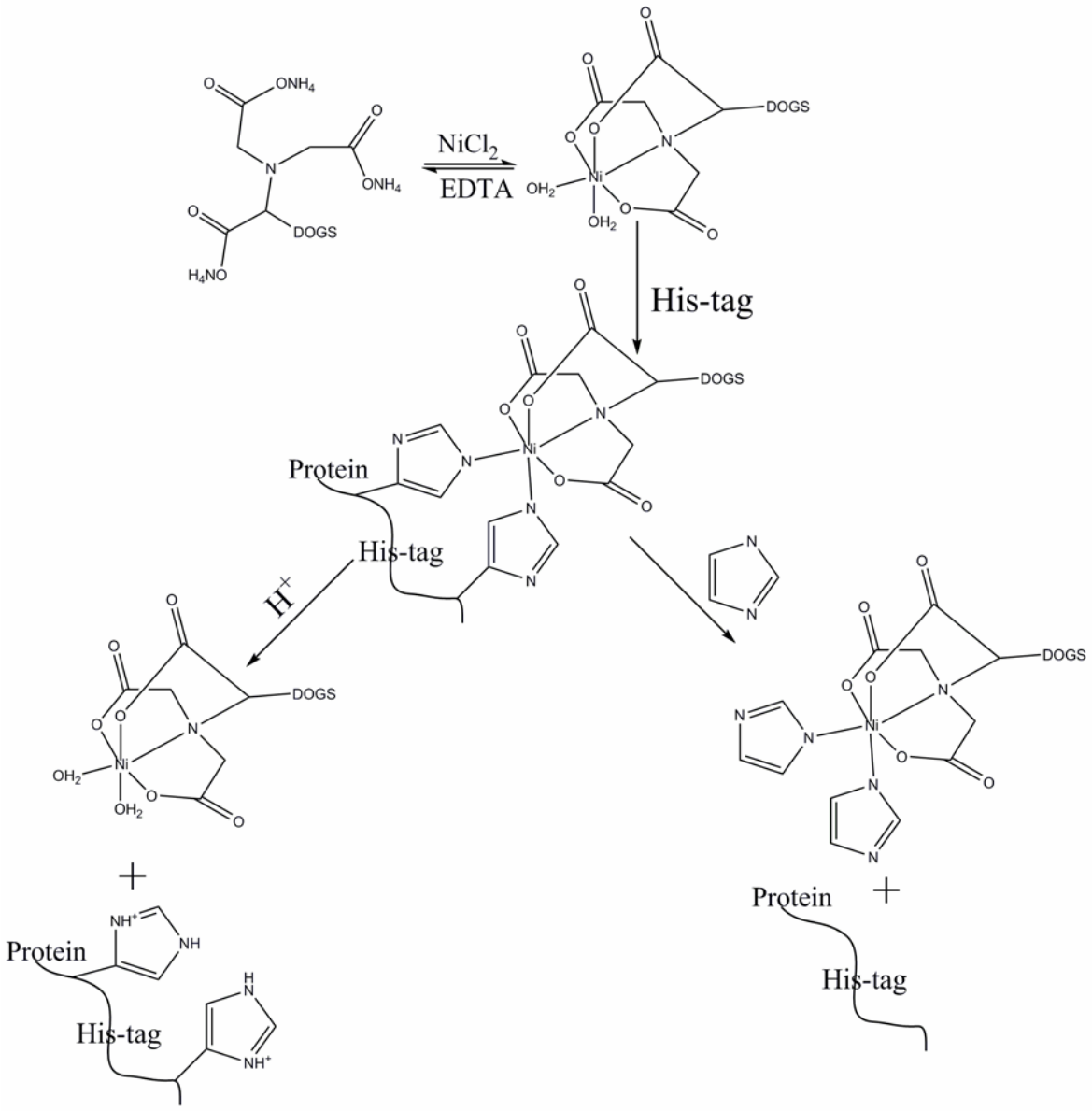

Figure 3.2 Binding activity of nickel-chelated DOGS-NTA. NTA readily chelates the free Ni^{2+} with a strong binding constant. When the His-tagged protein is present, two water ligands are substituted by two imidazole side chains of histidines. This chelation can be inversed by adding imidazole and changing the pH.

3.2 Materials and Methods

3.2.1 ^{15}N His-tagged Protein Preparation

cDNA encoding the carbohydrate recognition domain of galectin-3 (residues 114-250) was cloned from an existing full-length human galectin-3 gene (pET-3c). The cloned gene, after digestion by two restriction enzymes (Nde1 and BamH1), was inserted into the pre-digested *Escherichia coli* expression vector pET-23a containing a C-terminal 6x His-tag. Proper cloning was verified by DNA sequencing. BL21 (DE3) codon plus cells were transformed with the expression plasmid for galectin-3-CRD (pET-23a) in 50 ml of LB rich media containing 100 μg/ml ampicillin and incubated at 37°C overnight. Cells were collected, resuspended in M9 media and added to 1 L M9 media, containing 100 μg/ml ampicillin, 1 g/L $^{15}NH_4Cl$ and 4 g/L glucose. Protein induction was initiated by the addition of IPTG to 0.5 mM (at OD_{600} of 1.0) and incubated for an additional 2-4 hours. Cells were harvested by centrifugation at 5000 rpm for 20-30 min. Cell pellets were frozen at -80^{o}C before lysis. The cell pellets were thawed at room temperature and cells were resuspended in 10 ml phosphate buffer (75 mM KH_2PO_4, 75 mM NaCl, 2 mM EDTA, 1 mM DTT, 5 mM NaN3 pH 6.8) at room temperature, broken using sonication (8 min, 10 s pulse and 20 s pause), and subsequently clarified by centrifugation (100k x g, 1 hr). The supernatant was loaded into a 25 ml lactosyl-agarose affinity column (Sigma) which was pre-equilibrated in a phosphate buffer. Elution with 300 mM lactose yielded 25-40 mg pure galectin-3-CRD as assessed by SDS-PAGE. To verify His-tag presence and accessibility, the galectin-3-CRD was passed over a 25 ml nickel affinity column (Amersham) and eluted by the addition of 300 mM imidazole. The UV absorbance was monitored to verify the binding of the protein with the nickel column. Overnight dialysis against a phosphate buffer

(at 4°C) effectively removed the residual imidazole. NMR samples were made to be 0.3 mM ^{15}N His-tagged galectin-3-CRD in 75 mM KH_2PO_4, 75 mM NaCl, 5 mM NaN_3 (pH=6.8).

3.2.2 Nickel Chelated Bicelle Preparation

The DOGS-NTA nickel salt is hydrophobic and by itself insoluble in water. However, in the presence of C12E5 or hexanol, it is readily incorporated into the alignment media. The nickel-chelated $C_{12}E_5$ media was made in the same fashion as previously described by Ruckert *et. al*. [21], with DOGS-NTA-Ni being added directly to the appropriate volume of hexanol prior to mixing with lipids. $C_{12}E_5$ was mixed thoroughly with the buffer. Hexanol/DOGS-NTA-Ni was added in 2-3 μL increments and vortexed after each addition. The alignment of this medium was checked by measuring the deuterium splitting at 25°C on a spectrometer operating at 600 MHz for protons.

The DOGS-NTA nickel salt was successfully incorporated into DMPC/DHPC bicelle although ligand RDCs were not shown here. DMPC is as insoluble as nickel chelate in aqueous buffer. It is important to distribute the nickel chelate evenly along the bicelle surface. The following protocol was used to make 1ml nickel-chelated 15% DMPC/DHPC bicelle. 123 mg DMPC and the appropriate amount of DOGS-NTA-Ni were dissolved in 500 μl of a volatile organic solvent (chloroform/ methanol). The organic solvent was removed by vacuum overnight. The residue was resuspended in the desired buffer. At this stage, it did not dissolve but formed a milk-white suspension. 26.9 mg DHPC was dissolved in the desired buffer. DHPC, DMPC and nickel chelate were mixed together and vortexed very well and frozen with liquid nitrogen or acetone/dry ice mixture. The mixture was then thawed at room temperature and sonicated to disperse the mixture . The freeze-thaw-sonicate procedure was repeated at least 3 times until the mixture was clear. This chelate bicelle functions almost the same as the regular DMPC/DHPC

bicelle. The alignment was checked by monitoring the deuterium splitting when the temperature increased.

Deuterium splittings were measured from both DMPC/DHPC bicelle preparations, in the presence and absence of DOGS-NTA-Ni, with liquid crystal concentrations ranging from 16 to 2 percent.

3.2.3 NMR Spectroscopy

All NMR spectra were recorded with a Varian Inova 800 MHz spectrometer equipped with a cryogenic triple resonance probe. One-dimensional ^{1}H-^{15}N HSQC spectra were collected to monitor the association of His-tag galectin-3-CRD with DOGS-NTA-Ni. Spectra were recorded at a temperature of 298 K using 1000 scans on samples containing 0.3 mM galectin-3-CRD in 3% $C_{12}E_5$, 0.3 mM galectin-3-CRD in 3% $C_{12}E_5$/0.5 mM DOGS-NTA, and 0.3 mM galectin-3-CRD in 3% $C_{12}E_5$/0.5 mM DOGS-NTA-Ni.

Two parallel coupled gradient HSQC experiments were performed for the measurement of RDCs from protein resonances under isotropic (298 K) and partially-aligned (298 K) conditions on 3% $C_{12}E_5$ / 0.5 mM DOGS-NTA-Ni samples. Data collection typically included 1024 t_2 points and 256 t_1 points, subsequently extended with linear prediction and zero filling to 2048 and 1024 points respectively. The residual dipolar couplings were the differences in the splittings from both aligned and isotropic conditions (298 K).

The sensitivity-enhanced ^{1}H-^{13}C coupled HSQC (natural abundance) [22] was used to observe RDCs emanating from ligand resonances. Spectra were recorded under isotropic (298 K) and partially-aligned (298 K) conditions with samples containing either 6 mM free lactose, 6 mM lactose in 3% $C_{12}E_5$/0.5 mM DOGS-NTA-Ni, or 6 mM lactose with 0.3 mM His galectin-3-CRD in 3% $C_{12}E_5$/0.5 mM DOGS-NTA-Ni. Data collection included 1024 t_2 points and 512 t_1

points with 16 scans each for free lactose and lactose in media. Due to the line broadening observed for the membrane association in the presence of protein, 64 acquisitions per t_1 point were acquired in samples containing lactose with His-galectin-3-CRD in 3% $C_{12}E_5$/hexanol/0.5 mM DOGS-NTA-Ni.

Due to the near parallel orientation of directly bonded C-H vectors in lactose, only four independent RDCs could be obtained. In order to measure at least five independent RDCs, constant-time ^{1}H-^{1}H COSY (CT-COSY) experiments [23] were run to obtain the remaining ^{1}H-^{1}H couplings needed for order tensor analysis. As above, spectra were recorded under isotropic (298 K) and partially-aligned (298 K) conditions with samples containing either 6 mM free lactose, 6 mM lactose in 3% $C_{12}E_5$/hexanol/0.5 mM DOGS-NTA-Ni, or 6 mM lactose with 0.3 mM His galectin-3-CRD in 3% $C_{12}E_5$/hexanol/0.5 mM DOGS-NTA-Ni. Constant time periods were arrayed from 0.1 ms to 0.42 ms with time increments of 0.04 ms (for free lactose and lactose in media) or 0.018 ms to 0.153 ms with time increments of 0.015 ms (for lactose with His-galectin-3-CRD in 3% $C_{12}E_5$/hexanol/0.5 mM DOGS-NTA-Ni).

NMRPipe/NMRDraw [24] was used for data processing including the extraction of RDCs of both protein and ligand. For RDC measurement, the NMRDraw frequency-domain nonlinear least squares module (NLinLS) was used. The resulting measured couplings had estimated errors of 1 Hz (20 % of the overall range). For precise measurement of ligand couplings, XRAMBO, a Bayesian fitting program, was used for each C-H pair [25]. XRAMBO allowed for measurement of ligand RDCs with estimated errors less than 0.5 Hz. Intensity-based CT-COSY spectra were processed and dipolar couplings extracted as described in the literature [23]. The principal order frame and the order parameters were determined using all experimental

dipolar couplings analyzed by REDCAT [26]. Further details of data processing are described in chapter 4.

3.3 Results

3.3.1 Assessment of His-tag Addition and Protein Structure Perturbation

Galectin-3, a chimera-type galectin, consists of an N-terminal collagen-like domain and C-terminal carbohydrate recognition domain (CRD). The N-terminal domain does not have a stable secondary structure and is less studied. The C-terminal domain of galectin-3 (residues 114-250) comprises the CRD and carries the galactose-specific binding site. After examining the crystal structure of galectin-3 CRD (PDB ID: 1A3K), the His-tag was introduced onto the C-terminal end [17] to minimize paramagnetic broadening effects of the nickel ion on resonances from the bound ligand and to prevent masking of the oligosaccharide-binding site by His-tag-Nickel-chelate interactions. ^{15}N isotope-labeled protein was expressed in M9 minimal media supplemented with ^{15}N ammonium chloride. Due to the presence of a galactose-specific CRD and a C-terminal His-tag, the protein could be isolated using a lactosyl-agarose affinity column or a standard Ni-chelate as described in the "Materials and Methods" section.

Although many proteins can tolerate the addition of a terminal His-tag, it is still necessary to address the potential structural alteration resulting from His-tag addition. This usually is accomplished by comparing ^{1}H-^{15}N HSQC spectra recorded on His-tagged and non-His-tagged samples. HSQC spectra are commonly used to report on the characteristics of soluble macromolecules isotopically enriched in ^{15}N [27]. Any structural change causes an alteration of the local electronic environment and is reflected in the HSQC spectra. Overlaying the ^{1}H-^{15}N HSQC spectra for the galectin-3-CRD and the 6xHis-tag galectin-3-CRD showed little change in observed dispersion patterns.

3.3.2 Nickel-Chelated Lipid Alignment and Stability

The quality of an alignment medium is determined by its homogeneity, stability and the order of induced alignment. A good alignment medium must align in a magnetic field for a long enough period of time for the collection of NMR data. The properties of DOGS-NTA-Ni doped alignment media were tested and, as described in "Materials and Methods", DOGS-NTA had been successfully incorporated into both DMPC/DHPC and $C_{12}E_5$/hexanol alignment media.

DOGS-NTA was added to $C_{12}E_5$ /hexanol (pentaethylene glycol monododecyl ether) bicelles [21] at a ratio of approximately 0.5 mM DOGS-NTA to 3%(m/m) lipid. Splitting of the deuterium resonance from 10% ^2H_2O included in the medium was used to monitor alignment, homogeneity and stability. A 3% $C_{12}E_5$/hexanol preparation showed splittings of 12-14 Hz with no residual isotropic peak. The deuterium splitting of the same sample was measured after being stored at room temperature for 14 days. The splitting remained in the same range (12-14 Hz for 3% lipid). Moreover, the addition of galectin-3-CRD to levels of 0.3 mM did not substantially change these splittings. This behavior is very similar to that of undoped 3% $C_{12}E_5$/hexanol bicelles.

The same procedure was applied to DOGS-NTA doped DMPC/DHPC. At 25^{o}C, the alignment media was isotropic and showed an unsplit deuterium peak the same as the undoped DMPC/DHPC bicelle. At temperatures higher than 30^{o}C, the splitting of the deuterium appeared. For 16% DMPC/DHPC bicelle doped with DOGS-NTA, the deuterium splitting is 27 Hz at 35^{o}C and is similar to the undoped DMPC/DHPC bicelle (25 Hz at 35 ^{o}C). The addition of DOGS-NTA to C12E5/Hexanol and bicelle shows no changes in the homogeneity, stability and order of the alignment.

3.3.3 Association of Protein with Nickel Doped $C_{12}E_5$/hexanol

We expected strong chelation between the side chains of the His-tag and the NTA-Ni from the bicelle surface. The accessibility of the NTA-Ni was tested by monitoring the intensity change of one-dimensional spectra. One-dimensional versions of HSQC spectra provide a good indicator of association. In isotropic solution, each amide ^{1}H-^{15}N pair gives rise to a sharp line. With membrane association, the correlation time which is proportional to effective molecular size increases and numerous through-space dipolar couplings appear. The line width is also proportional to the total correlation time and will contribute to signal loss. Changes in the signal-to-noise ratio in spectra with and without the chelating metal can be indicative of selective lipid anchoring of His-tagged macromolecules in solution. Figure 3.3 shows a 1D ^{1}H-^{15}N HSQC spectrum collected on a 0.3 mM sample of galectin-3-CRD (pH 7.4) in a 3% $C_{12}E_5$/Hexanol bicelle preparation doped with (A) 0.5 mM DOGS-NTA ammonium salt (no chelating metal) and (B) 0.5 mM DOGS-NTA-Ni. A substantial reduction of intensity of resonances in the presence of the Ni-loaded DOGS-NTA is obvious. However, the line widths are only slightly broader in the presence of the activated chelator. RDC data on the protein from 2D HSQC spectra under "B" conditions are identical to those for protein in bicelles without DOGS-NTA-Ni. These phenomena can be interpreted by the slow exchange between bound and free states of the protein. The severe line broadening prevented the detection of the resonances from the bound protein and only resonances from free protein are detected. This will have little consequence on measurement of transferred RDCs for ligands since RDCs will be averaged over all states and it is only important that RDCs from ligand in association with the bicelle bound protein be accentuated.

In the case of a fast exchanging system, NMR observables such as RDCs and line widths will be averaged over bound and free forms. Accurate RDC measurement usually requires reasonable line widths. Hence, it is convenient to be able to adjust the fraction of protein associated with the alignment medium. Following the principles of His-tagged protein purification, the binding affinity can be adjusted by the addition of competitors (e.g. imidazole when compatible with the protein under study) or buffer pH. The effect of pH is shown in Figure 3.3 C. It is known that pH affects His-tag–nickel association due to the altered protonation state of the histidine side chain. This results in a reduced affinity as the pH is decreased. At low pH (6.8, Figure 3.3 C), the loss of intensity due to the fraction of anchored protein is relatively small compared to the loss at a higher pH (7.4, Figure 3.3 B). The smaller fraction of associated protein results in observable resonances for ligand, which still show large RDCs. The fraction of bound protein can be adjusted by the competitive imidazole but galectin-3 is less stable in imidazole buffer so the competitive effect by imidazole is not shown.

3.3.4 Measurement of Ligand RDCs

Lactose (Galβ1-4Glc) was chosen as a ligand for validation of our proposed method for measurement of transferred RDCs. The reducing end of lactose is not protected so when dissolved in water, lactose is present in two isomers: α- and β–lactose. Both forms have little overlap in ^{1}H-^{13}C HSQC spectra except H4-C4 of glucose. Only the resonances from α-lactose were chosen because there is one more independent RDC available from the reducing end. ^{1}H-^{13}C RDCs, as well as some ^{1}H-^{1}H RDCs, were measured at a 20:1 ratio of ligand to protein, or 6mM lactose in the presence of 0.3 mM galectin-3-CRD in pH 6.8, 3% $C_{12}E_5$/ 0.5 mM DOGS-NTA. At this relatively high concentration of lactose, the signal-to-noise ratio is good enough to measure one-bond ^{1}H-^{13}C RDCs at natural abundance. RDCs, as an extra contribution to the ^{1}H-

^{13}C J-couplings, were measured by taking the difference between ^{1}H-^{13}C splittings in the indirect dimensions of ^{1}H-^{13}C in coupled HSQC spectra taken under isotropic conditions (no alignment media) and aligned conditions (DOSGS-NTA doped media with and without Ni). ^{1}H-^{1}H couplings were measured using a ^{1}H-^{1}H CT COSY experiment [23] and the splittings were

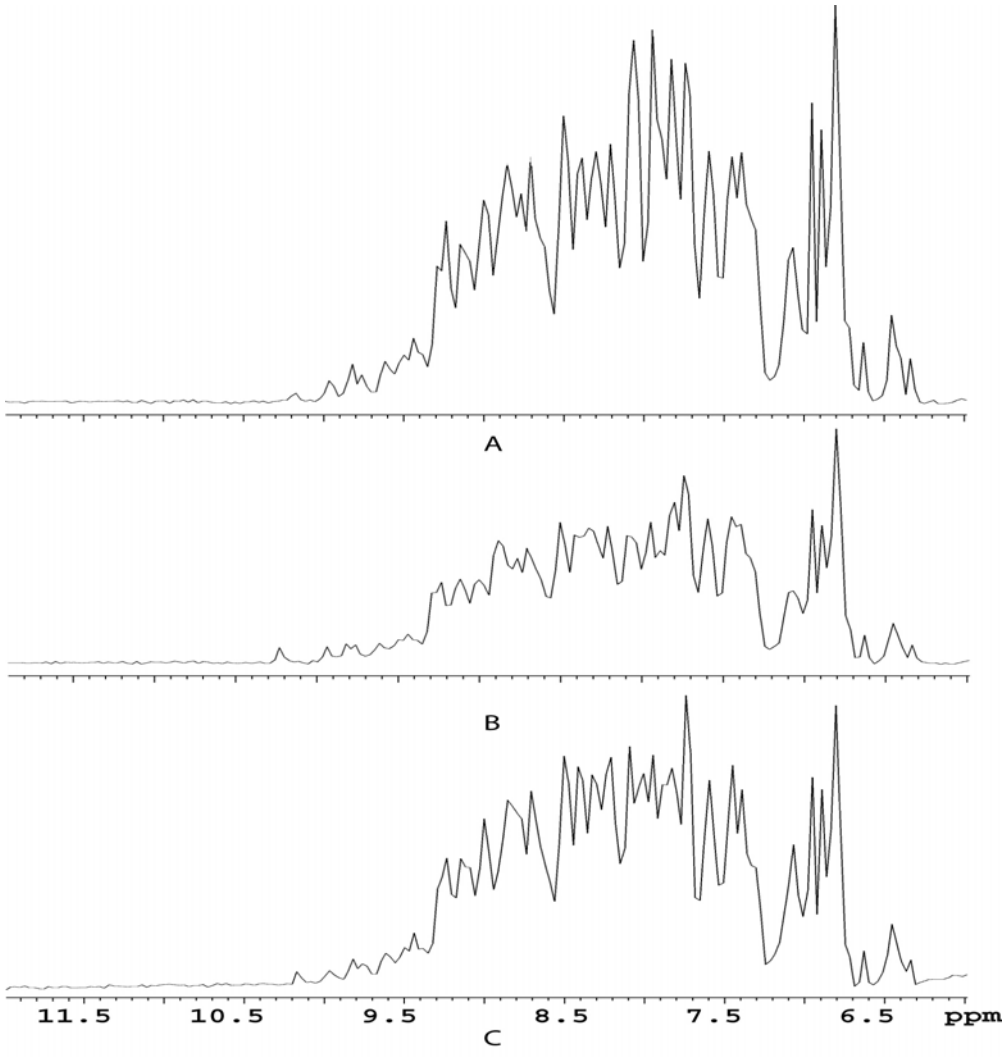

Figure 3.3. 1D HSQC spectra of ^{1}H-^{15}N His-galectin-3-CRD. (A) 0.3 mM Galectin-3 (pH 7.4) in a 3% $C_{12}E_5$/hexanol bicelle preparation doped with 0.5 mM DOGS-NTA ammonium salt (no chelating metal); (B) 0.3 mM galectin-3 (pH 7.4) in a 3% $C_{12}E_5$/hexanol bicelle preparation doped with 0.5 mM DOGS-NTA-Ni; (C) 0.3 mM galectin-3 (pH 6.8) in a 3% $C_{12}E_5$/hexanol bicelle preparation doped with 0.5 mM DOGS-NTA-Ni.

extracted by fitting the ratio of intensity of cross-peak and diagonal peak against constant time using a tangent function. The errors reported are the fitting errors. All measured RDCs are summarized in table 3.1.

As expected, there is no significant change in RDCs when comparing those from the ligand in the absence of protein with those from the ligand in the presence of galectin 3-CRD and uncharged DOGS-NTA (columns 1 and 2, Table 1). Given a dissociation constant of 1mM for the lactose–galectin-3 interactions and the initial concentrations of 0.3 mM galectin-3 and 6 mM lactose, we would expect about 5 % of the ligand to be bound under the conditions of our experiment. Even if the level of order for the protein were twice as large as that for the free ligand, changes would be less than 1 Hz. The measured RDCs difference is in the range of measurement error. It is very difficult to extract RDCs for the bound state from the measured differences. An increase in the difference between observed RDCs for the free-state and for the protein-complexed state (from less than 1 Hz to nearly 5 Hz) is seen when DOGS-NTA is charged with nickel (column 3, Table 1). This difference in observed RDCs is caused by the change in weighting of contributions from the free- and bound-states.

As an illustration, splittings of C2-H2 and C5-H5 of the galactose residue are shown in Figure 3.4. Changes in splittings in the presence and absence of protein in media lacking the Ni required for enhanced protein alignment are almost indistinguishable (Figures 3.4A and 3.4B). A substantial change in splitting is observed, however, when the protein is associated with a different medium (lactose with His-tagged galectin-3-CRD and DOGS-NTA nickel chelated lipid –see in Figure 3.4.). The cross peaks in Figure 3.4 C are broadened and show partially resolved ^{1}H-^{1}H splittings, which is expected for a more strongly oriented molecule. While RDCs

are somewhat harder to measure, the fact that they are more strongly influenced by the protein in a bound state makes data on the bound state easier to extract and interpret.

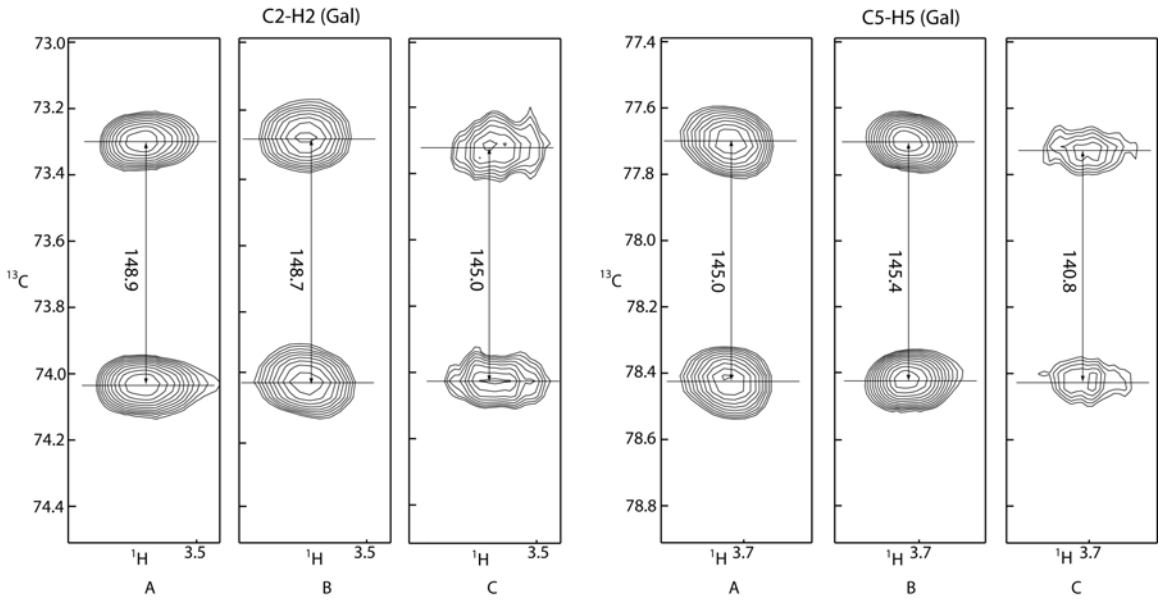

Figure 3.4. Splittings of C2-H2 and C5-H5 of the lactose galactose residue. (A) 6 mM lactose, 0.3 mM galectin-3 (pH 6.8) in 3% $C_{12}E_5$/Hexanol bicelle preparation doped with 0.5 mM DOGS-NTA ammonium salt (no chelating metal); (B) 6 mM lactose (pH 6.8) in 3% $C_{12}E_5$/Hexanol bicelle preparation doped with 0.5 mM DOGS-NTA-Ni; (C) 6 mM lactose, 0.3 mM galectin-3 (pH 6.8) in 3% $C_{12}E_5$/Hexanol bicelle preparation doped with 0.5 mM DOGS-NTA-Ni.

Free and bound forms of ligand were present in the solution. The observed RDC is the average of free-state RDCs and bound-state RDCs weighted by the proportion of the free-state to the bound-state. RDCs for the bound state (column 4, Table 3.1) can be extracted using the known binding constant for lactose (1 mM) [17] and equating a simple population weighted average of RDCs in the free state (column 1) and RDCs of the bound state to measured RDCs (column 3). The results are given in column 4. The magnitude of order observed in the bound state, as well as a test for consistency with bound geometries observed in the crystal structure,

can be assessed using the known structure of lactose along with RDCs to calculate an order tensor [28]. Once an order tensor is determined, RDCs can be back-calculated for the examination of the correlation between observed couplings and those calculated from the model. A good correlation or a small Q-factor usually gives an indication of successful sampling of the bound-state conformation.

Given the free-state lactose geometry (phi, psi = 50, -116 respectively), the measured free-state RDCs of the ligand, and analysis by REDCAT, order matrix parameters were determined to be Sxx = -1.1E-05, Syy = -1.7E-04, Szz = 1.8E-04. Order matrix parameters derived from the bound lactose geometry found in the crystal structure of galectin-3-CRD (PDB ID 1ULC, phi, psi = 48, -95 respectively) gave order matrix parameters of Sxx =- 2.8E-4, Syy = -2.9E-03, Szz = 3.2E-03). This suggests an average increase in the order of a factor of 15 due to the association of a fraction of the galectin-3-CRD with the alignment medium. The increased order of the proteins dominated the contribution of bound-state ligand and allowed for more accurate extraction of RDCs from the bound state.

The bound-state RDCs were back-calculated using REDCAT and the results are given in column 5 of table 3.1. The experimental bound-state RDCs were consistent with the back-calculated bound-state RDCs except for Glc C2-H2. The resonances from this site could not accurately be estimated due to overlapping of the background from the $C_{12}E_5$ alignment media. The root-mean square deviation (RMSD) excluding C2-H2 is 5 Hz .We also attempted back-calculation using a slightly different solution structure, but the agreement was not as good with an RMSD of 8 Hz. While these differences are not large, the better agreement in the former case suggests consistency with the structure found in the crystal structure.

3.4 Discussion

The dominance of free-state RDCs in the observed average normally limits one's ability to accurately assess bound-state RDCs. However, the incorporation of target proteins containing poly-histidines tags into lipid bilayer–like preparations doped with DOGS-NTA-Ni can result in sufficient enhancement of bound-state dipolar couplings and improve the accuracy of measuring bound-state RDCs,. The extraction of a set of bound-state RDCs for lactose bound to galectin-3-CRD is quite consistent with the bound-state geometry of lactose found in a galectin-3-CRD crystal structure [17]. Our experiments have shown that is it better to have both protein and ligand structural information to determine the structure of a ligand-bound protein complex. In our experience, signal-to-noise levels of resonances from the protein were dramatically reduced upon association with the alignment media but no RDC was observed. One explanation is that the protein is slowly exchanging on and off the bicelle. Once the protein is bound, resonances are too broad to detect. Resonances observed from the fraction of the free-state protein do not provide information on protein-bicelle association. However, this does not invalidate information on conformations of bound ligands. In fact, most transferred NOE data are obtained on proteins too large to characterize by NMR. Transferred RDC data on ligands for proteins that cannot be observed is equally valuable.

In fact, if the crystal structure of a complex is known, it is still possible to obtain a representation of protein orientation indirectly from dipolar couplings emanating from bound-state lactose. When order parameters in the principal order frame are derived for the bound ligand, a rotation matrix for taking protein coordinates into the principal frame also result. The model produced by moving the protein to the principal frame agrees well with one having the His-tag pointing towards the membrane surface. Such a model could not be derived in cases

where appropriate protein structures did not already exist. However, in the present case, it does support the view that interaction of the His-tag with a lipid-anchored chelate is responsible for the enhanced level of order.

The order of the alignment was enhanced through the strong association of the His-tagged protein with the chelated lipid. Like the method we present in chapter 2, the spin distribution anisotropy was obtained through the external liquid crystal media. They are suitable for the hydrophilic ligands but might not be applicable to the hydrophobic ligands. Strong interactions between the hydrophobic ligands and the alignment media deviate the fast exchanging regime of the ligand. It has been known that most of the drug candidates are hydrophobic.

In order to solve the issue above, self-oriented alignment by the paramagnetic metal ion is being exploited. It will be covered in the next chapter. Self-oriented alignment stems from the magnetic susceptibility anisotropy of the paramagnetic metal ion and can be obtained without the aid of external alignment media. This advantage allows the study of both hydrophobic and hydrophilic ligands.

Table 3.1 Lactose RDCs measured at a ^{1}H frequency 800 MHz

			With DOGS-NTA Ammonium	With DOGS-NTA Nickel salt		
		RDC_{free}(Hz)	RDC(Hz)	RDC(Hz)	RDC_{bound}(Hz)	RDC_{cal}(Hz)
Glc(α)	C1-H1	-4	-3.9	-4.5	-14	-13
	C2-H2	2.7	2.5	0.7	-37	-21
	C3-H3	3.7	3.4	2.9	-12	-22
	C5-H5	3.9	3.2	2.8	-18	-26
	H1-H2	-0.1		-1.5	-29	-24
Gal	C1-H1	1.7	1.5	-2.6	-84	-86
	C2-H2	1.9	2.1	-1.8	-72	-70
	C3-H3	2.4	1.8	-2.3	-91	-96
	C4-H4	0.1	1	1.4	26	23
	C5-H5	2.8	2.4	-1.8	-89	-88
	H1-H2	0.2		0.9	15	13

Order tensor elements	Sxx	Syy	Szz
Free state lactose	-1.1E-04	-1.7E-04	1.8E-04
Bound-state lactose	-2.8E-04	-2.9E-03	3.2E-03

References

1. Seidel, R.D., T.D. Zhuang, and J.H. Prestegard, *Bound-state residual dipolar couplings for rapidly exchanging ligands of His-tagged proteins.* Journal of the American Chemical Society, 2007. **129**(15): p. 4834-4839.

2. Moore, J.M., *NMR techniques for characterization of ligand binding: Utility for lead generation and optimization in drug discovery.* Biopolymers, 1999. **51**(3): p. 221-243.

3. Homans, S.W., *NMR spectroscopy tools for structure-aided drug design.* Angewandte Chemie. International Ed. In English, 2004. **43**(3): p. 290-300.

4. Jiménez-Barbero, J. and T. Peters, *Nmr Spectroscopy of Glycoconjugates*: Wiley-Vch,Weinheim,Germany,2002.

5. Tugarinov, V., P.M. Hwang, and L.E. Kay, *Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins.* Annual Review of Biochemistry, 2004. **73**: p. 107-146.

6. Shimizu, H., A. Donohue-Rolfe, and S.W. Homans, *Derivation of the bound-state conformation of a ligand in a weakly aligned ligand-protein complex.* Journal of the American Chemical Society, 1999. **121**(24): p. 5815-5816.

7. Jain, N.U., S. Noble, and J.H. Prestegard, *Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings.* Journal of Molecular Biology, 2003. **328**(2): p. 451-462.

8. Zhuang, T.D., H. Leffler, and J.H. Prestegard, *Enhancement of bound-state residual dipolar couplings: Conformational analysis of lactose bound to Galectin-3.* Protein Science, 2006. **15**(7): p. 1780-1790.

9. Chen, Z.J. and B. Reif, *Measurements of residual dipolar couplings in peptide inhibitors weakly aligned by transient binding to peptide amyloid fibrils.* Journal of Biomolecular NMR, 2004. **29**(4): p. 525-530.

10. Koenig, B.W., et al., *Measurement of dipolar couplings in a transducin peptide fragment weakly bound to oriented photo-activated rhodopsin.* Journal of Biomolecular Nmr, 2000. **16**(2): p. 121-125.

11. Ikegami, T., et al., *Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions.* Journal of Biomolecular Nmr, 2004. **29**(3): p. 339-349.

12. Feeney, J., et al., *Calmodulin tagging provides a general method of using lanthanide induced magnetic field orientation to observe residual dipolar couplings in proteins in solution.* Journal of Biomolecular Nmr, 2001. **21**(1): p. 41-48.

13. Schmitt, L., C. Dietrich, and R. Tampe, *Synthesis and Characterization of Chelator-Lipids for Reversible Immobilization of Engineered Proteins at Self-Assembled Lipid Interfaces.* Journal of the American Chemical Society, 1994. **116**(19): p. 8485-8491.

14. Prosser, R.S., et al., *Lanthanide chelates as bilayer alignment tools in NMR studies of membrane-associated peptides.* Journal of Magnetic Resonance, 1999. **141**(2): p. 256-260.

15. Kubalek, E.W., S.F.J. Legrice, and P.O. Brown, *2-Dimensional Crystallization of Histidine-Tagged, Hiv-1 Reverse-Transcriptase Promoted by a Novel Nickel-Chelating Lipid.* Journal of Structural Biology, 1994. **113**(2): p. 117-123.

16. Sanders, C.R., et al., *Magnetically-Oriented Phospholipid Micelles as a Tool for the Study of Membrane-Associated Molecules.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1994. **26**: p. 421-444.

17. Seetharaman, J., et al., *X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution.* Journal of Biological Chemistry, 1998. **273**(21): p. 13047-13052.

18. Nakamura, M., et al., *Involvement of galectin-3 expression in colorectal cancer progression and metastasis.* International Journal of Oncology, 1999. **15**(1): p. 143-148.

19. Almkvist, J., et al., *Activation of the neutrophil nicotinamide adenine dinucleotide phosphate oxidase by galectin-1.* Journal of Immunology, 2002. **168**(8): p. 4034-4041.

20. Umemoto, K., et al., *Conformational differences in liganded and unliganded states of Galectin-3.* Biochemistry, 2003. **42**(13): p. 3688-3695.

21. Ruckert, M. and G. Otting, *Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments.* Journal of the American Chemical Society, 2000. **122**(32): p. 7793-7797.

22. Schleucher, J., et al., *A General Enhancement Scheme in Heteronuclear Multidimensional Nmr Employing Pulsed-Field Gradients.* Journal of Biomolecular Nmr, 1994. **4**(2): p. 301-306.

23. Tian, F., P.J. Bolon, and J.H. Prestegard, *Intensity-based measurement of homonuclear residual dipolar couplings from CT-COSY.* Journal of the American Chemical Society, 1999. **121**(33): p. 7712-7713.

24. Delaglio, F., et al., *Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes.* Journal of Biomolecular Nmr, 1995. **6**(3): p. 277-293.

25. Andrec, M. and J.H. Prestegard, *Metropolis Monte Carlo implementation of Bayesian time-domain parameter estimation: Application to coupling constant estimation from antiphase multiplets.* Journal of Magnetic Resonance, 1998. **130**(2): p. 217-232.

26. Valafar, H. and J.H. Prestegard, *REDCAT: a residual dipolar coupling analysis tool.* Journal of Magnetic Resonance, 2004. **167**(2): p. 228-241.

27. Wuthrich, K., *Nmr of Proteins and Nucleic Acids*: Wiley-Interscience, New York, Ny USA 1986.

28. Losonczi, J.A., et al., *Order matrix analysis of residual dipolar couplings using singular value decomposition.* Journal of Magnetic Resonance, 1999. **138**(2): p. 334-342.

Chapter 4

Structure Determination of a Galectin-3 – Carbohydrate Complex Using Paramagnetism-based NMR Constraints

4.1 Introduction

Carbohydrate-protein interactions are primary mediators in the communication of a cell with its environment. They influence cell adhesion, modulate intracellular signaling, and provide a pathway for entry of a number of pathogens. Recently, these interactions have become targets for the development of drugs that can modulate cellular processes, and the determination of the structural characteristics of the complexes has become an issue [1-4]. Structure determination of these complexes, particularly those of lectins with natural ligands, can be problematic. As mentioned previously, the interactions are often weak, with binding affinities in the range of 100 μM-1 mM. This precludes the use of X-ray crystallography due to the difficulty in obtaining the crystal of whole complex. NMR structure determination is also problematic since these interactions are often dominated by hydrogen-bonding networks that place non-exchangeable proton pairs across the protein-ligand interface at distances too great for accurate measurement of NOEs. Below, we present an alternative NMR approach that is not NOE based. Instead, it relies on a combination of orientationally-sensitive measurements (RDCs) and long-range distance and orientation dependent measurements (PCSs) to conformationally constrain and locate a carbohydrate ligand within a protein binding site.

The conventional NMR approach to structure determination for weakly-binding ligands involves a combination of transferred NOEs (trNOEs) to determine bound ligand geometry [5-7]

and saturation transfer difference (STD) spectra to determine parts of the ligand in closest contact with the protein surface [8, 9]. Both of these approaches work very well, for most complexes of small ligands in which the ligands are rapidly exchanging on and off the surface of large proteins. Here, the very efficient cross relaxation properties of large systems allow measurements to be dominated by bound ligands, even with a 10-100 fold excess of ligand over protein. However, neither of these methods provides much information about the placement of the ligand on the protein surface.

Residual dipolar couplings can provide a good complement to NOE measurements and several examples of their measurement under conditions of rapid ligand exchange have appeared in the literature [10-14] . Under ideal circumstances RDC measurements will help clarify the relative orientations of ligand and protein. However, the development of reliable methods for RDC measurement is difficult. RDC measurements are normally achieved by aligning the protein and ligand in liquid crystal media, and rely on steric interactions of the anisotropically-shaped molecules with elements of the alignment medium. Since the complex and free ligand are equally likely to have anisotropic shapes, one cannot count on preferential orientation of the bound complex and the domination of measurements by information on the ligand in this complex as is the case for transferred NOEs,

To circumvent this problem, our lab has developed methods for enhancing the alignment of proteins by adding components that have specific interactions with the hydrophobic elements of the alignment medium. One approach was to attach a short hydrophobic chain to a cysteine at the C-terminus of the protein; another was to incorporate a chelate-carrying lipid into the orientation medium so that it would bind his-tag terminated proteins [10, 12]. In both cases, the

enhanced protein alignment resulted in the enhanced alignment of bound ligands and the domination of measured ligand RDCs by the protein-bound fraction.

Applications of the above RDC enhancement strategies are, however, limited. Incorporating a hydrophobic chain requires a single reactive cysteine in the protein sequence. Integrating a chelate-carrying lipid onto the orientation medium encourages an interaction of the protein with the alignment medium that results in severe line broadening of protein signals and an inability to determine protein orientation. In both cases, the requirement for an alignment medium with hydrophobic properties excludes applications to more hydrophobic ligands which would have competitively strong interactions with the medium even as free molecules. This is particularly restrictive for drug discovery, in which the addition of hydrophobic moieties is commonly used to increase affinity. Also, methods based purely on orientational constraints cannot position ligands translationally. The following methods were developed to address these limitations.

Alignment without the use of hydrophobic liquid crystal media is possible if a molecule has a sufficiently high magnetic susceptibility anisotropy. Any substance, in the presence of a magnetic field, will have a magnetic moment induced in proportion to its magnetic susceptibility. This moment, in turn, interacts with the magnetic field to produce an energy of interaction that is proportional to the magnetic field squared. When the susceptibility is anisotropic, different orientations will have different energies, and orientational distributions become non-isotropic in accordance with Boltzmann statistics. For a protein-ligand system, the susceptibility of the protein can be made an order of magnitude or more larger than that of the ligand by attaching an appropriate paramagnetic tag. This enhances the alignment of the bound ligand without the use of a hydrophobic alignment medium.

The use of paramagnetic tags has the additional advantage of providing relatively long-range distance constraints (i.e. the distance between the paramagnetic center and NMR observable sites of interest). Both paramagnetic enhanced relaxation and pseudo-contact shifts provide this distance information. However, we chose to use PCS (proportional to $1/r^3$) because the distance dependence is less steep than for paramagnetic relaxation enhancement (proportional to $1/r^6$). In combination with a large paramagnetic moment, this allows pseudo-contact shifts to be observed for nuclei up to 35 Å away from the metal center [15, 16].

Paramagnetic tags can be a chemical or a short polypeptide. The magnetic anisotropy usually arises from the paramagnetic metal ion chelating with the tag. Among all the paramagnetic metal ions, lanthanide series are extremely attractive. They behave similarly due to their similar size and outer shell 4f electron configuration. But their magnetic properties range from a ferromagnetic ground state in Gd, a diamagnetic ground state in La and Lu, to complicated incommensurate spin spiral structures of all other elements [17]. Studying calbindin D_{9k} complexed with different lanthanide ions suggests that all lanthanide ions share the same coordination configuration but different magnetic susceptibility anisotropy. Dy^{3+} and Tb^{3+} have the largest magnetic susceptibility anisotropy while Gd^{3+}, due to its half-filled 4f electron orbital, is magnetic susceptibility isotropic [18]. Lanthanide ions other than La^{3+}, Lu^{3+} and Gd^{3+} have larger magnetic susceptibility anisotropy than other transient metals such as Ni^{2+}, Fe^{3+}.

There have been several applications of paramagnetic constraints previously. Many of the early applications were to metalloproteins, since an appropriate paramagnetic metal ion can often be incorporated through replacement of an original ion with a suitable paramagnetic ion [18-21]. However, among all proteins, only an estimated one third are metalloproteins [22], and it is useful to be able to introduce additional metal-binding sites. More recent applications have

used metal binding tags. A metal-binding chelate can be chemically attached to a cysteine through a cysteine-specific reaction [23, 24] or a metal-binding peptide can be incorporated into the expression construct of the protein of interest [25-27]. Both approaches had been exploited by our lab. In terms of the magnitude of NMR observables, the latter approach induced a much larger order of alignment of the protein and was further investigated to determine the structure of the ligand-protein complex. However, a review of NMR observations using chemical tag is presented here as well.

The EDTA- or DTPA-based lanthanide binding tag was successfully synthesized in Griesinger's lab [24, 28, 29]. The chiral center was modified to remove the steorochemical chirality when the metal ion was bound to the EDTA moiety [24, 29, 30]. This improvement greatly simplified the spectra merging two sets of peaks into one. One of the synthesized EDTA-based tags (EDTA-CA-MTS) is shown as Figure 4.1. EDTA-CA-MTS has two functional moieties.

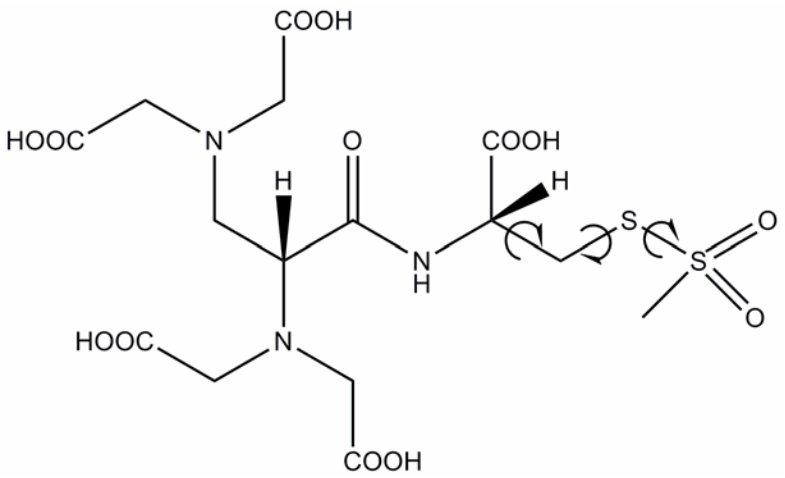

Figure 4.1. S-Methanesulfonyl-L-cysteaminyl-1R-carbonic-acid-ethylenediamine-N, N, N, N-tetraacetic acid (EDTA-CA-MTS) structure. The σ bonds linking the tag and the protein allow substantial free internal motion of the tag.

The EDTA end readily chelates metal ions with very strong binding affinity (~nM). The methanethiosulfonate end, similar to PMTS (refer to section 2.2.3), specifically reacts with free sulfhydryl groups to incorporate the EDTA moiety into the protein. This lanthanide binding tag was attached to the trigger factor mutant (S100C) and significant protein alignment was observed. Up to 8 Hz ^{1}H-^{15}N RDC on a ^{1}H frequency 800 MHz spectrometer and up to 0.5 ppm pseudo-contact shifts were measured [24].

These paramagnetism-based constraints are not only important for refining protein structure, but also provide benefits in studying protein-ligand systems. In addition to the orientational information from RDCs, the distance information from pseudo-contact shifts can be extracted to restrain the translation of the ligand relative to its receptors.

The magnitude of NMR observables can be drastically scaled down if the paramagnetic ion undergoes relative motion to the protein. RDCs and pseudo-contact shifts are proportional to the magnetic anisotropy. The lanthanide ions such as dysprosium or terbium usually have large enough magnetic anisotropy susceptibility, but this relative motion can average out the magnetic anisotropy. Efforts were therefore made to generate some universal tags with little flexibility. A comparison of the flexibility of EDTA-based tags and other peptide-based tags, such as lanthanide-binding peptides, is provided later in this chapter.

Using combinatorial screening, and beginning with a consensus sequence for calcium-binding motifs of EF-hand proteins, a short polypeptide (YIDTNNDGWYEGDELLA) with a particularly high affinity for lanthanide ions had been previously developed [31]. Lanthanide ions have very useful properties for luminescence studies, X-ray crystallography and NMR experiments involving paramagnetic effects. This makes fusion constructs between polypeptides like this, and proteins of interest, particularly appealing [32]. The properties of this original

peptide have been thoroughly investigated, including production of an X-ray structure [33]. Second generation tags also have been designed, including those with two lanthanide binding sites which give greatly improved magnitudes of alignment [34]. In our work, the original lanthanide binding tag was added to the C-terminus of the expression construct for our target protein. In order to minimize the flexibility of the tag relative to the protein and still maintain its native protein structure, we needed to minimize the length of the linker. Both N- and C-terminal tags were explored. The N-terminal tag was excluded from this study due to the formation of inclusion bodies on protein expression. The C-terminal tagged protein proved to be soluble and well-behaved. In the final construct, a C-terminal tag with no linker amino acid between the tag and the C-terminus of the protein was used.

We selected the carbohydrate binding domain of galectin-3 as our target protein. Galectin-3 is a mammalian protein that is found in the cytosol and the nucleus as well as extra-cellularly [35, 36]. In its extra-cellular role, galectin-3 has been implicated in modulation of inflammatory responses, migration of malignant cells, and the regulation of various growth factors. In its intracellular role, it has been implicated in various apoptotic events [37-39].

Galectin-3 is normally a two-domain protein with a C-terminal carbohydrate binding domain and an N-terminal oligomerization domain. For the reasons state above, we focused on the C-terminal domain. This domain has a molecular weight of approximately 15 kDa. It has a high-resolution crystal structure [40], and previous NMR studies have led to assignment of cross-peaks in its ^{15}N-^{1}H NMR spectrum [41]. Galactose-terminated cell surface oligosaccharides are believed to be its primary natural ligands. Here, we studied the binding of lactose as a small soluble analog of these ligands. Lactose binds to the protein weakly with a K_d of 0.2 mM and undergoes fast exchange on and off the protein binding site [42]. RDCs and pseudo-contact

shifts of both the protein (galectin-3-LBT) and the ligand were measured to allow a structure calculation of the ligand-protein complex.

4.2 Materials and Methods

4.2.1 Galectin-3C Preparation

In order to incorporate the EDTA-based tag to the protein, a reactive cysteine is needed. The same protein as that for specific alkylthiolation (addressed in chapter 2) was prepared here. To avoid precipitation and hydrolysis of the metal ions, a phosphate buffer or buffer with a pH>7 is not appropriate. The protein was stored in 75 mM phosphate buffer containing 5 mM DTT (pH=7.4). Before the reaction, the reductant was completely removed by buffer exchange to a MOPS buffer (50 mM MOPS, 50 mM NaCl, pH=6.8). The same MOPS buffer, after degassing by N_2, was used for NMR sample preparation.

4.2.2 Galectin-3-LBT Preparation.

The lanthanide binding tag was appended to the C-terminus by regular cloning techniques. The nucleotide sequence of the lanthanide binding tag is 5'-TATATTGATACCAATAATGATGGCTGGTATGAAGGCGATG AACTGCTGGCG-3'. The incorporation of this 51-base polynucleotide cannot be achieved by a single step PCR reaction. Instead, the polynucleotide was incorporated using three sequential polymerase chain reactions, each adding approximately one third of the sequence. No linker amino acid was introduced to prevent increasing the flexibility of the tag. One forward primer and three reverse primers were designed. These included 5'-AACGAGCGGCATATGCTGATTGTGCC-3' (forward primer), 5'-CATCATTATTGGTATCAATATATATCATGGTATATGAAGCAC-3' (first reverse primer), 5'-CATCGCCTTCATACCAGCCATCATTATTGGTATCAATATA-3' (second reverse primer)

and 5'-CGGCTCGAGTCACGCCAGCAGTTCATCGCCTTCATACCAGCC-3' (third reverse primer). A Xho1 restriction site was included in the third primer. A pET-3c vector containing the gene encoding galectin-3 CRD was used as the DNA template for the first PCR step. The PCR product was purified after each PCR step and was used for the subsequent PCR reaction. The complete PCR product after the third step and a pET-29a vector were double-digested by Nde1 and Xho1. The circular plasmid was obtained through ligation of the digested PCR product to the digested pET-29a vector using T4 DNA ligase (Biorad,CA). The correct sequence of the gene was verified by gene sequencing at IBL (UGA, GA).

The same protein expression protocol was used here for the expression of both unlabeled and ^{15}N-labeled protein as described in previous publications [10, 43]. Since the lanthanide ion will precipitate at pH>7, a pH 5.4 MES buffer which contained 50 mM MES, 50 mM NaCl ,5 mM EDTA and 2 mM DTT was used as the cell lysis and storage buffer. The protein was purified by binding the protein to a lactosyl-agarose affinity column and eluting with 300mM lactose. The resulting protein was pure, as tested through SDS-PAGE electrophoresis. The protein was concentrated and excess lactose was removed using an Amicon concentrator (Millipore, MA). Before the protein was mixed with lanthanide ions, EDTA was removed by a buffer exchange using 50mM MES buffer (50 mM MES, 50 mM NaCl, pH=5.4).

To ensure that the protein contained the tag and that the protein conformation had not changed significantly, mass spectrometry and ^{15}N-^{1}H HSQC spectroscopy were performed to check the molecular weight of the protein and the chemical shifts of all the residues. The molecular weight showed very good agreement with the theoretical value (17,533 Da). The HSQC spectrum was compared with the spectrum of the galectin-3 CRD and, except for a few belonging to the C-terminus, the cross-peaks did not show chemical shift changes. Extra

resonances also were observed, consistent in number with that expected for the lanthanide binding tag.

4.2.3 Incorporation of EDTA-Based Tag to Galectin-3C

The reductant-free protein sample galectin-3C was mixed with the reagent (EDTA-CA-MTS) in a stoichiometry of 1:2. The mixture was kept at 4^{o}C for 24 hr. The excess of reagent was removed through buffer exchange. The reaction efficiency was evaluated by collecting a ^{1}H-^{15}N HSQC spectrum. The spectrum showed no residual peaks. In other words, the reaction was complete so no further separation of reacted from unreacted protein was needed. A 0.5mM protein sample was made for the following NMR measurement.

4.2.4 NMR Experiments

To prevent the nonspecific binding of the metal ion with the protein, the protein was mixed with dysprosium chloride in a 1:1 ratio. The same concentration of diamagnetic protein sample was prepared by mixing the protein with lutetium chloride in a 1:1 ratio. NMR experiments were recorded at 25^{o}C at a ^{1}H frequency of 800 MHz or 600 MHz on Varian Inova NMR spectrometers. Both are equipped with triple resonance gradient probes; in the case of the 800 MHz spectrometer this is a cryogenic probe with enhanced sensitivity.

4.2.4.1 NMR Experiments for Galectin-3C

NMR experiments for EDTA-tagged galectin-3C were recorded at 25^{o}C in MOPS buffer (50 mM MOPS, 50 mM NaCl, pH 6.8) at a ^{1}H frequency of 800 MHz. Two parallel experiments were performed: one for a sample with 0.5 mM Dy^{3+} ions and another with 0.5mM Lu^{3+} ions. A sensitivity-enhanced coupled ^{1}H-^{15}N HSQC experiment with proton decoupling was used to measure the J-couplings for the lutetium sample and J+D for the dysprosium sample. The

differences of splittings in the ^{15}N dimension and the residual dipolar couplings were measured for comparison with that of galectin-3-LBT in the next step.

4.2.4.2 NMR Experiments for Galectin-3-LBT

NMR experiments for galectin-3-LBT were recorded at 25^{o}C in 50 mM MES buffer (pH 5.4) with 50 mM NaCl at a ^{1}H frequency of 80 0MHz or 600 MHz on Varian Inova NMR spectrometers. Although the level of protein alignment is proportional to B_0^2, paramagnetic relaxation enhancement also increases with B_0^2 resulting in a greater loss of observable protein peaks at higher fields. We therefore performed all protein observation experiments at 600 MHz whereas all ligand experiments were performed at 800 MHz. For ligands, both RDCs and PCSs are the average of free and bound states. For accurate measurement, a higher field is required to achieve both better S/N and larger alignment of bound ligands.

For protein RDC measurement, a ^{15}N-^{1}H HSQC-TROSY spectrum was recorded [44]. For protein PCS measurement, a normal ^{15}N-^{1}H HSQC spectrum was recorded at 600 MHz. 0.3 mM of ^{15}N labeled protein was mixed with 0.3 mM diamagnetic lutetium to obtain a diamagnetic reference spectrum. A sample with the protein at the same 0.3 mM concentration was prepared with 0.3 mM paramagnetic dysprosium to obtain PCSs. Residual dipolar couplings were measured as observed differences in offsets in the ^{15}N dimension between paramagnetic and diamagnetic samples. Pseudo-contact shifts were measured as differences in ^{1}H chemical shifts between diamagnetic and paramagnetic samples. The peak assignments were achieved by pairing the peaks of diamagnetic and paramagnetic samples along diagonal lines.

4.2.4.3 NMR Measurements of Lactose

For lactose ^{13}C-^{1}H RDCs, the diamagnetic sample was prepared by mixing 0.5 mM protein with 2.5 mM lactose and 0.5 mM lutetium ion in 50 mM deuterated MES buffer (50 mM

D-MES, 50 mM NaCl, pH=5.4). For the paramagnetic sample, 0.5 mM dysprosium was mixed with 0.5 mM protein and 2.5 mM lactose in 50 mM deuterated MES buffer. Coupled ^{13}C-^{1}H HSQC spectra were recorded at a ^{1}H frequency of 800 MHz using a cryogenic triple resonance probe. For the diamagnetic spectrum, 256 t1 points over 9500 Hz were acquired with 100 acquisitions each. For the paramagnetic spectrum, due to the severe line broadening by paramagnetic relaxation enhancement, 128 t1 points over 9500 Hz were acquired with 400 acquisitions each.

Constant-time COSY spectra also were recorded at an ^{1}H frequency of 800 MHz to obtain ^{1}H-^{1}H RDCs for the ligand [45]. The same diamagnetic sample was used for the diamagnetic reference spectrum. The paramagnetic sample was prepared by mixing 0.5 mM protein with 10 mM lactose and 0.5 mM dysprosium in 50 mM deuterated MES buffer. The mixing time was arrayed from 0.1 to 0.3 s with time increments of 0.05 s. For the lactose pseudo-contact shifts, the same samples as those used for RDC measurements were used except that 1mM tetramethyl ammonium chloride (TMAC) was added to both samples for chemical shift referencing. Regular ^{13}C-^{1}H HSQC spectra were recorded and PCSs were measured as the difference in ^{1}H chemical shifts in diamagnetic and paramagnetic samples.

4.2.5 Data Analysis

For proteins, the spectra were processed using NMRPipe [46]. For protein RDCs and PCSs, nonlinear line shape curve fitting was applied to each peak to fit them to a Gaussian curve and extract the peak center. The peak assignments were achieved by comparison with the deposited galectin-3 CRD spectrum [41]. Only the peaks having unambiguous assignments were selected for alignment tensor determination.

For ligand RDCs and PCSs, since the magnitudes are small, data were extracted using a Bayesian curve fitting program (XRAMBO) [47]. XRAMBO is a tool for the extraction of quantitative parameters from a set of experimental data. It is used for the extraction of accurate quantitative information from the data and providing realistic error estimates. A metropolis Monte-Carlo algorithm was used to directly sample points in parameter space. Free-induction decays were modeled as sums of exponentially-decaying sinusoids, each described by four parameters (intensity, decay constant, peak center position, and phase). The RDC was taken as the difference in splitting for diamagnetic and paramagnetic samples. XRAMBO extracts the most-likely estimate of peak centers and provides estimates of error as well. Thus, the splitting can be reliably estimated by subtracting of one peak center from another. A nonlinear regression of a model requires the starting values for all parameters being estimated. Correct estimates of the starting values can accelerate the convergence based on the least square estimation. XRAMBO requires the original free induction decay (FID) and an estimation of peak position as input. The original FID for several peaks were obtained by processing the 2D ^{1}H-^{13}C HSQC spectrum routinely, keeping the imaginary part and applying no weighting functions (refer to Appendix B). This was followed by summing a region of interest perpendicular to the ^{13}C dimension to get a 1D NMRPipe file using readROI and writeROI modules. The real and imaginary parts of the 1D projection were not correctly organized. The 1D projection was then reformatted using an M-based NMRPipe macro (see Appendix C) and inverse Fourier transformed to reconstruct the FID file. The initial peak positions, intensities and line widths can be estimated by NMRPipe auto-peak picking. These starting values usually give very good estimates for the non-linear modeling of the exponential-sinusoidal function. The Monte Carlo search was set to 16,000 iterations and the final parameter values and error estimation were

obtained after the rejection rate fell between 60%-70%. In practice, if the 60%-70% rejection rate is not satisfied after the first run, the output parameter estimates and FIDs can be used as the starting estimates for a second run. This procedure was repeated until the final rejection rate fell to between 60% and 70%.

For ^{1}H-^{1}H RDCs of the ligand, the same procedure as described by Tian, et. al. was used [45]. The splitting was extracted by fitting the ratio of cross-peak and auto-peak intensities as a function of the constant-time interval to a tangent function. This intensity-based method provides an accurate measurement so it can be used at a high ligand protein ratio.

4.2.6 Comparison of Measured RDCs

The flexibility of the EDTA-based tagged protein and lanthanide-binding peptide tagged protein was compared using RDCs, which are sensitive to the relative motion of the tag and the protein. A sufficiently large magnetic alignment is required to observe transferred residual dipolar coupling and pseudo-contact shifts in weakly-binding, fast-exchanging systems when the bound fraction is usually small. With paramagnetism, the magnetic alignment originates from the intrinsic magnetic susceptibility anisotropy. If the same lanthanide ion is incorporated in the same manner of coordination and a rigid body exists (i.e, there is no relative motion), the same range of RDCs can be observed.

4.2.7 Structure Calculation of Lactose-Bound Galectin-3-LBT

The order tensors were determined using the program REDCAT [48]. Since RDCs and PCSs share the same order tensors, if a diamagnetic reference is substracted [49] and if a rigid model is appropriate, both RDCs and PCSs can be used simultaneously to determine the alignment tensors. One only needs to specify the correct RDC_{max} and PCS_{max} constants in the input files (24350 Hz for ^{1}H-^{15}N RDCs, and 1.26E+10 Hz for ^{1}H PCSs at 600 MHz). However,

the appropriateness of a rigid model and, for PCSs, the actual position of the metal ion must first be determined. Hence, we initially used only RDCs to determine the order tensor elements (Sxx,Syy,Szz). The same tensors were then used to back-calculate PCSs as the ion was moved over grid in 2Å steps. The ability to find a position that fit both RDCs and PCSs well was taken as evidence of a rigid system and both sets were combined to get an appropriate estimate of the order tensor.

REDCAT calculates the order tensor elements (Sxx, Syy,Szz) using singular value decomposition based on the experimental residual dipolar couplings and X-ray crystal structure of the galectn-3 CRD. The position of the metal ion was then optimized using the internal variable [50] and PARArestraints modules [49] available for the XPLOR-NIH program. The estimated order parameters were specified as Da= 2.85 and Dr=-1.72. Da and Dr easily can be calculated from the determined elements of order tensor (Sxx, Syy, Szz) and the relationships are expressed by equations 4.1 and 4.2:

$$D_a = \frac{1}{2} D_{\max} S_{zz} \tag{4.1}$$

$$D_r = \frac{1}{2} D_{\max} \left(S_{xx} - S_{yy} \right) \tag{4.2}$$

(Here D_{max}=24350 for ^{1}H-^{15}N, 60400 for ^{1}H-^{13}C and 240200 for ^{1}H-^{1}H)

To perform this simulation, the protein and the lanthanide binding tag were grouped into two different clusters. The linker between the protein and the tag was set to be free for rotation and translation. PC6 was used as the molecular dynamics integration algorithm. The length of the flexible connection varied (from 5 to15 amino acids) as attempts were made to find the position of the metal ion giving the best fit of the PCSs and RDCs. One structure with the fewest PCS and RDC violations was selected for final ligand-protein complex determination.

These same procedures were applied when modeling the final ligand-bound protein complex structure (IVM and PARArestraints) [49, 50]. RDCs and PCSs of ligand and protein were used as orientational and distance restraints. Since two ^{1}H-^{1}H ligand RDCs, which could have variable inter-nuclear distances, were used for the final structure determination, XDIP was used to handle these ^{1}H-^{1}H couplings along with ^{1}H-^{13}C and ^{1}H-^{15}N couplings [51]. The protein was grouped into one cluster. The pyranose rings of galactose and glucose were grouped to maintain the most stable chair form during simulated annealing. Previous experiments [52, 53] had shown that when the lactose binds to different lectins, lactose adopts a conformation near the local energy minimum with little alteration in glycosidic bond angles. Hence, the relative rotation of galactose and glucose rings was restrained by setting the dihedral angles of the glycosidic bond to phi=50 and psi=-116 with a +/-5 degree variation. One intermolecular NOE also was included to accelerate the convergence in some of the calculations [10]. The calculated structure was validated by checking for normal van der Waals contacts and appropriate hydrogen bonds between the ligand and the protein.

4.3 Results

4.3.1 Protein RDCs and PCSs

Figure 4.2 presents a comparison of HSQC and TROSY overlays for samples of tagged galectin-3 with the diamagnetic metal Lu^{3+} (Figure 4.2.A) and the paramagnetic metal Dy^{3+} (Figure 4.2.B) at 600 MHz. HSQC cross-peaks (which are fully decoupled) are displaced from TROSY cross-peaks (which are not decoupled) by half of the sum of scalar and dipolar couplings in both dimensions. The difference in the offsets between diamagnetic and paramagnetic cases yields ½ the RDC coupling at each site. It is readily apparent that both positive and negative couplings can be observed and that the approximate range of coupling at

600 MHz is +/-6 Hz. These couplings are dependent on the square of the magnetic field and similar measurements at 900 MHz show a range of +/- 15 Hz. Interpolating to 800 MHz for comparison, this is approximately twice the range seen in a study of tagged ubiquitin [25], an indication that the mobility of the tag may be restricted. Cross-peaks for the non-tagged protein have been assigned previously and in most cases these assignments could be transferred unambiguously to the spectra in Figure 4.2.

Based on these assignments, it is clear that, for the paramagnetic sample, the cross-peaks from residues close to the metal center (15 Å or less) are severely broadened or completely disappear. This is due to the strong relaxation enhancement. These effects are also dependent on the square of the magnetic field and additional peaks are lost in spectra at 900 MHz. The dipolar interaction between the unpaired electrons of the lanthanide ion and the nuclei of the protein also causes changes in chemical shifts (see Figure 4.3). The chemical shift changes in ppm are independent of the type of the nucleus (^{1}H or ^{15}N) but depend on the distance of the nucleus to the metal ion, as well as on the Ln-nucleus vector orientation in the principal susceptibility frame [54]. For backbone ^{1}H and ^{15}N pairs at long distances from the metal, similar chemical shifts changes (in ppms) are expected in both dimensions. Comparison of spectra of complexes with paramagnetic ions (Dy^{3+}) to spectra of complexes with diamagnetic ions (Lu^{3+}) shows a characteristic diagonal shift in peak positions that helps pair the resonances of diamagnetic and paramagnetic samples. In principle, chemical shift anisotropy offsets, which occur with orientation of the protein, also lead to differences in chemical shifts [55].

However, at 600 MHz, for the complexes studied here, these chemical shift anisotropy offset are estimated to be less than 5 Hz and do not significantly contribute to most shifts. The pseudo-contact shifts however, are large enough to introduce a few additional ambiguities in the

assignments of shifted peaks. The ambiguities in assignments, accidental peak overlap, and peak broadening, reduced the number of RDCs and PCSs that could be measured to 37 and 34, respectively. The values of backbone ^{1}H-^{15}N RDCs and amide ^{1}H PCSs are shown in table 4.1.

Similarly, the sensitivity-enhanced ^{1}H-^{15}N coupled HSQC spectra were acquired for the measurement of RDCs of EDTA-tagged galectin-3C. Due to the paramagnetic relaxation effect, the peaks from the residues less than 15 angstroms away became broad or disappeared. But compared with galectin-3-LBT, more peaks can be observed. This is an indication of larger flexibility of the EDTA-based tag. To be more precise, RDCs of only those peaks having unambiguous assignments were recorded for order tensor determination. RDCs measured for EDTA-tagged galectin-3C are shown in table 4.2. From the determined order tensor of the receptor protein, the maximum transferred RDCs of the ligand can be estimated.

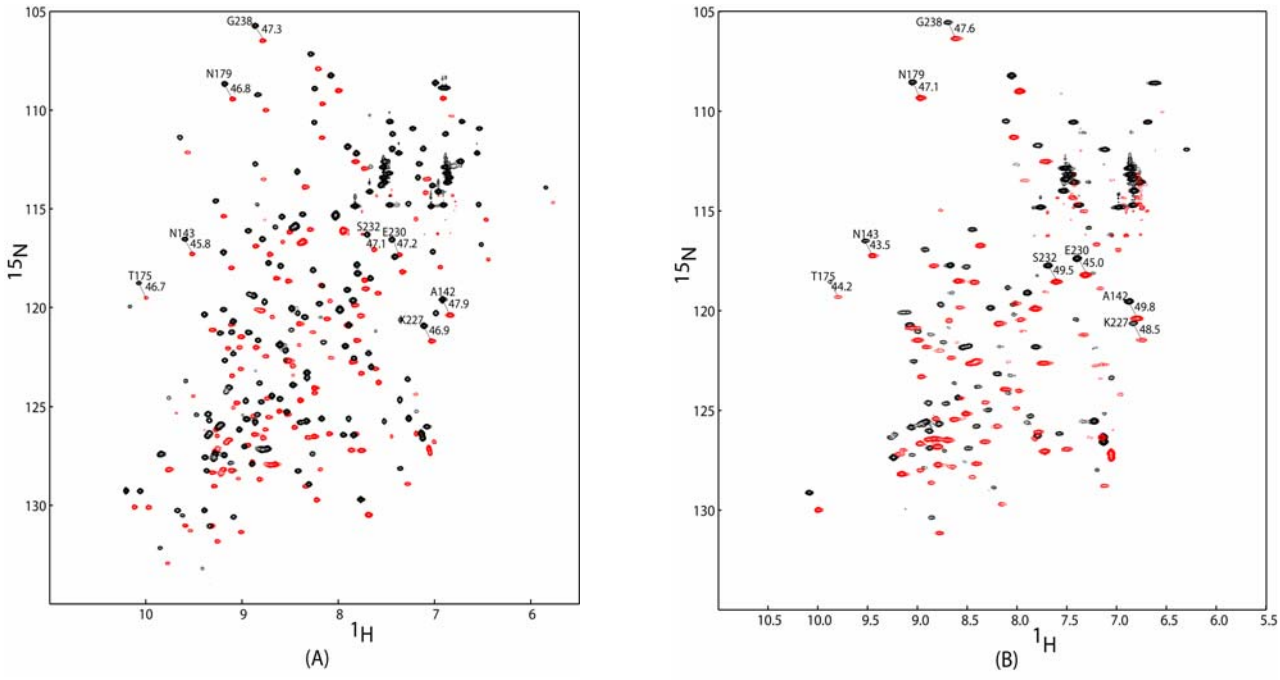

Figure 4.2 ^{1}H-^{15}N HSQC and TROSY overlays for 0.3 mM galectin-3-LBT at a ^{1}H frequency of 600 MHz with (A) 0.3 mM Lu^{3+} and (B) 0.3 mM Dy^{3+}.

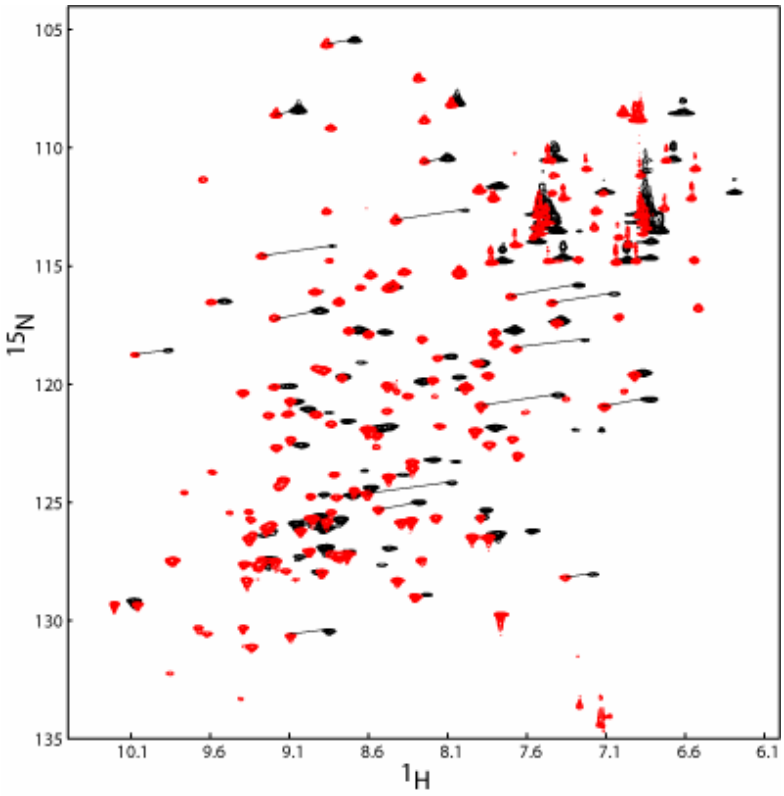

Figure 4.3 ^{1}H-^{15}N HSQC spectra at a ^{1}H frequency of 600 MHz for 0.3 mM galectin-3-LBT with 0.3 mM Lu^{3+} (red) or with 0.3 mM Dy^{3+} (black). Many peaks disappear for the protein sample with paramagnetic Dy^{3+}. Peak assignments were accomplished by drawing a diagonal line between diamagnetic and paramagnetic peaks as shown in the figure.

The observed ^{1}H-^{13}C RDC range of the ligand can be estimated using equation 4.3:

$$D_{obs} = D_{\max} S_{zz} \left(\frac{r_{H-N}}{r_{H-C}} \right)^{3} \left(\frac{\gamma_C}{\gamma_N} \right) P_{bound} \qquad (4.3)$$

where D_{max} is a constant (24350 for ^{1}H-^{15}N). r_{H-N} and r_{H-C} are the inter-nuclei vector length of ^{1}H-^{15}N and ^{1}H-^{13}C respectively, γ_c and γ_N are the gyromagnetic ratio of ^{13}C and ^{15}N, and P_{bound} is the fraction of bound-state ligand and can be calculated using known binding constants and the initial concentrations of both protein and ligand.

Table 4.1 ^{1}H-^{15}N RDCs and backbone amide ^{1}H PCSs of 0.3 mM galectin-3-LBT with 0.3 mM Dy^{3+} at a ^{1}H frequency 600 MHz spectrometer

ID	Residue	PCS (ppm)	RDC (Hz)
116	VAL	-0.13	-1.1
119	ASN	-0.28	4.6
139	LYS	-0.08	-2.0
141	ASN		-1.1
142	ALA	-0.05	-3.8
143	ASN	-0.08	4.6
144	ARG	-0.11	2.3
145	ILE	-0.17	3.1
151	ARG	-0.32	1.6
154	ASP	-0.23	1.2
155	VAL	-0.43	
157	PHE	-0.43	4.4
163	PHE	-0.08	-3.1
164	ASN	-0.07	1.7
166	ASN	-0.03	-2.4
167	ASN	-0.04	1.6
172	VAL	-0.18	3.6
175	THR	-0.21	4.3
176	LYS	-0.24	2.7
177	LEU	-0.16	0.1
179	ASN	-0.15	-0.5
180	ASN	-0.13	-0.7
182	GLY	-0.15	2.2
185	GLU	-0.15	5.7
187	GLN	-0.1	-3.4
189	VAL	0	0.6
196	LYS		2.3
220	PHE		0.0
222	ASN		-2.9
227	LYS	-0.29	-3.2
229	ASN	-0.44	2.0
230	GLU	-0.39	4.1
232	SER	-0.43	-4.5
233	LYS	-0.49	5.0
235	GLY	-0.45	
237	SER		-4.4
238	GLY	-0.18	-0.5
240	ILE	-0.12	-1.6
241	ASP	-0.08	0.9

Table 4.2 Backbone ^{1}H-^{15}N RDCs of 0.5 mM EDTA-tagged galectin-3 with 0.5 mM Dy^{3+} at a ^{1}H frequency 800 MHz spectrometer

Residue ID	J (Lu^{3+}) (Hz)	J+D(Dy^{3+}) (Hz)	RDC (Hz)	Back-Cal (Hz)
I134	92.7	93.2	0.5	0.5
L120	92.1	96.0	3.9	2.8
N143	89.8	90.1	0.3	0.7
G235	93.2	96.2	3.0	2.2
G238	93.9	92.1	-1.8	-1.5
G136	96.6	94.6	-2.0	-0.9
A142	94.4	95.2	0.8	0.2
V116	93.0	91.5	-1.5	-1.8
I145	92.6	93.8	1.2	1.8
T137	92.8	92.4	-0.4	-1.1
F198	93.0	92.9	-0.1	-0.3
E193	92.3	94.1	1.8	1.8
S237	94.0	94.7	0.7	-0.2
K139	95.6	95.3	-0.3	0.3
K227	94.0	94.1	0.1	0.0
D154	93.8	93.5	-0.3	-1.2
R151	90.4	92.1	1.7	1.6
K199	91.5	93.0	1.5	1.0
L219	92.5	94.7	2.2	2.5
S244	93.1	94.9	1.8	2.5
I240	91.9	91.2	-0.7	-0.7
G195	93.3	93.5	0.2	-0.3
E185	91.9	93.3	1.4	1.4

4.3.2 Flexibility Comparison of EDTA-Based Tag and LBT Tag

From the measured RDCs induced by the paramagnetic Dy^{3+} ions, the RDC range of EDTA-tagged galectin-3 is +/-3Hz in an 18.78 T magnet (800 MHz ^{1}H frequency spectrometer). The RDC range of LBT-tagged galectin-3 is +/-6 Hz in a 14.1 T magnet (600 MHz ^{1}H frequency spectrometer). The field-induced self-orientation is proportional to the square of the external magnetic field. The RDC range of LBT-tagged galectin-3 can be extrapolated to be +/-12 Hz in an 18.78 T magnet.

To be more precise, the order parameters of the protein were determined using REDCAT and the protein coordinates from the X-ray crystal structure. A direct comparison of the

determined Szz can confirm the flexibility of these two tags. The order parameters of EDTA-tagged galectin-3 are Sxx=1.9E-05, Syy=-8.8E-05, Szz=1.1E-04 at an 800 MHz magnetic field. RDCs of galectin-3-LBT were acquired at a spectrometer with a ^{1}H frequency 600 MHz. The order parameters of galectin-3-LBT at 600 MHz field are Sxx=-3.5E-05, Syy=-2.0E-04 and Szz=2.4E-04. Extrapolation to 800 MHz the order parameters of galectin-3-LBT are Sxx=-6.23E-05, Syy=-3.56E-04 and Szz=4.3E-04. Szz of galectin-3-LBT is about 4 times that of EDTA-tagged galectin-3. Assuming that cooridination of the metal ion to both tags is similar motion of EDTA-tag is of higher amplitude.

Galectin-3-LBT, after incorporating the lanthanide ion, induces a much larger order of alignment which is sufficient for the observation of the transferred residual dipolar couplings (estimated using equation 4.3) and pseudo-contact shifts. Only galectin-3-LBT was further investigated and the transferred residual dipolar couplings and pseudo-contact shifts were measured to determine the structure of lactose-bound galectin-3-LBT.

4.3.3 Analysis of Protein RDCs and PCSs.

The equations connecting RDCs and PCSs with the magnetic susceptibility tensor share a very similar form. This can be seen in equations 4.4 and 4.5 where θ and ϕ are polar angles of the interaction vectors in the principal susceptibility tensor frame, the $\Delta\chi$s are the axial and rhombic components of the susceptibility anisotropy tensor, and the *r*s are the lengths of the vectors between interacting pairs of nuclei and a nucleus and the paramagnetic center respectively [18, 56, 57]. Both equations (4.4 and 4.5) are written here to provide results in units of Hz.

$$RDC = -\frac{1}{120\pi^2}\frac{{B_0}^2}{kT}\frac{\gamma_i\gamma_j\hbar}{{r_{ij}}^3}[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi] \qquad (4.4)$$

$$PCS = \frac{B_0 \gamma_i}{24\pi^2 r_{im}^3}[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\phi] \quad (4.5)$$

The above equations assume one knows how to orient the principal frame of the susceptibility tensor in the molecular frame. This usually is not the case and the equations also can be written in an arbitrary molecular frame (usually the frame of an available PDB file). These forms are given in equations 4.6 and 4.7 where the θs are the angles of the interaction vector relative to the molecular frame axes.

$$PCS = \frac{B_0 \gamma_i}{8\pi^2 r_{im}^3}\sum_{ij}\chi_{ij}\cos\theta_i\cos\theta_j \quad (4.6)$$

$$RDC = -\frac{1}{40\pi^2}\frac{B_0^2}{kT}\frac{\gamma_i\gamma_j\hbar}{r_{ij}^3}\sum_{ij}\chi_{ij}\cos\theta_i\cos\theta_j \quad (4.7)$$

The equations can, in turn, be related to the level of order induced in the molecule as represented by the order tensor elements, S_{ij}, using the relationship in equation 4.8 [58].

$$S_{ij} = \chi_{ij}\frac{B_0^2}{20\mu_0 kT} \quad (4.8)$$

Althrough the susceptibility tensor elements (or order tensor elements) would appear to be the same in the two expressions above, there is a subtle difference. In both cases, the susceptibility tensor is an effective tensor, reduced in magnitude for NMR detectable sites in the protein by any internal motion between the tag and the protein. The reduction is, in principle, different in the two cases because of the difference in the distances and angles used, and this has led to a practice of separately fitting the two tensors [59]. Here we exploited the potential difference in the tensors to asses the nature of any motion. In cases where there appears to be no

motion, use of a common set of tensor elements results in a reduction in the number of parameters to fit by a factor of two. In cases where the LBT tag undergoes relative motion, the interpretation of pseudo-contact shifts becomes much more complicated due to the time averaging of $\left\langle \frac{3\cos^2\theta - 1}{2r^3} \right\rangle$ instead of $\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle$. The assumption of a rigid model can be tested by fitting the experimental data using a single structure.

To assess the validity of a rigid model, we used an order tensor derived purely from RDC data to back-calculate PCSs for models that differed only in placement of the lanthanide ion. The program REDCAT was used to determine the order tensor elements from RDCs [48]. At ^{1}H 600 MHz these were Szz=2.4E-04, Syy=-2.0E-04 and Sxx=-3.5E-05. Models used to back-calculate PCSs were generated by moving the ion over points on a 2Å grid, and at each point, equation 4.5, as modifided by substitution of order tensor elements, was used to back-calculate pseudo-contact shifts. Figure 4.4 shows correlation plots comparing experimental and calculated data for both PCSs (Figure 4.4.A) and RDCs (Figure 4.4.B). The fit of the PCS data is, in fact, reasonably good, confirming that a suitable model based on a single rigid structure was found.

The above calculation does not take into account constraints imposed by bond connections between the tag and protein or by van der Waals contacts between the tag and protein atoms. To introduce these constraints and improve the ion position, tools in the software package XPLOR-NIH were used. The position of the metal ion was adjusted using the internal variable [50] and PARArestraints module [49] as described in the Materials & Methods section of this chapter. Both RDC and PCS constraints were used in conjunction with a single set of pseudo-atoms representing the alignment frame.

By varying the number of flexible residues between the tag and the protein, a structure was identified with good molecular contacts, only one PCS violation larger than 0.1 ppm and no

RDC violations greater than 2 Hz. This was achieved by allowing a segment of 11 amino acids (from 244 to 254) to adopt a new conformation in the course of the simulation. The position of the lanthanide proved to be within 3 Å of that found by the grid search. The correlation plots for experimental PCSs plus RDCs and back-calculated PCSs plus RDCs show very good agreement (Figure 4.5 A & B), with Q-factors [60] of 0.23 and 0.27, respectively.

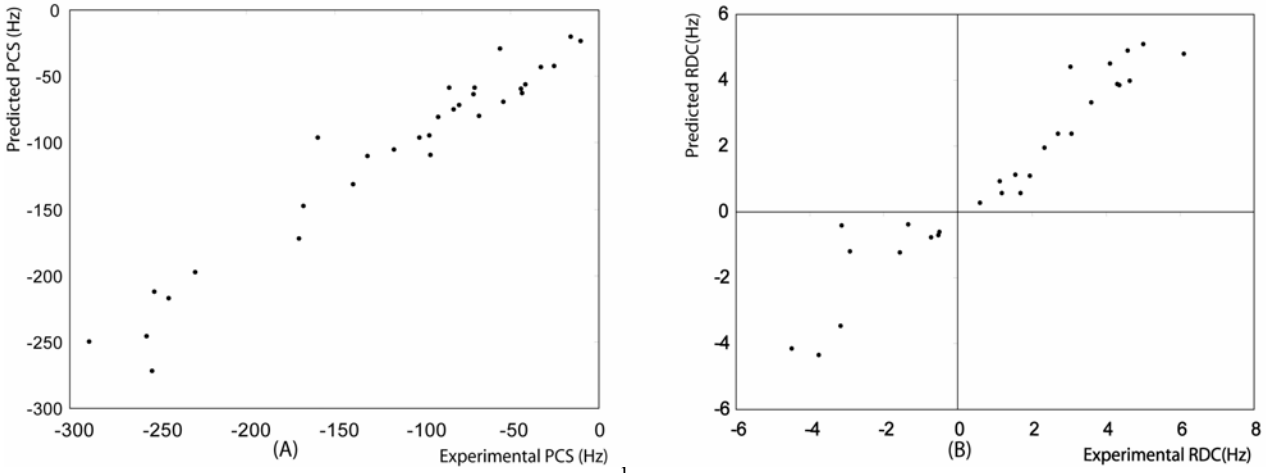

Figure 4.4 Experimental RDCs and PCSs at a ^{1}H frequency of 600 MHz versus back-calculated PCSs and RDCs using the structural coordinates determined by a 2 Å grid search

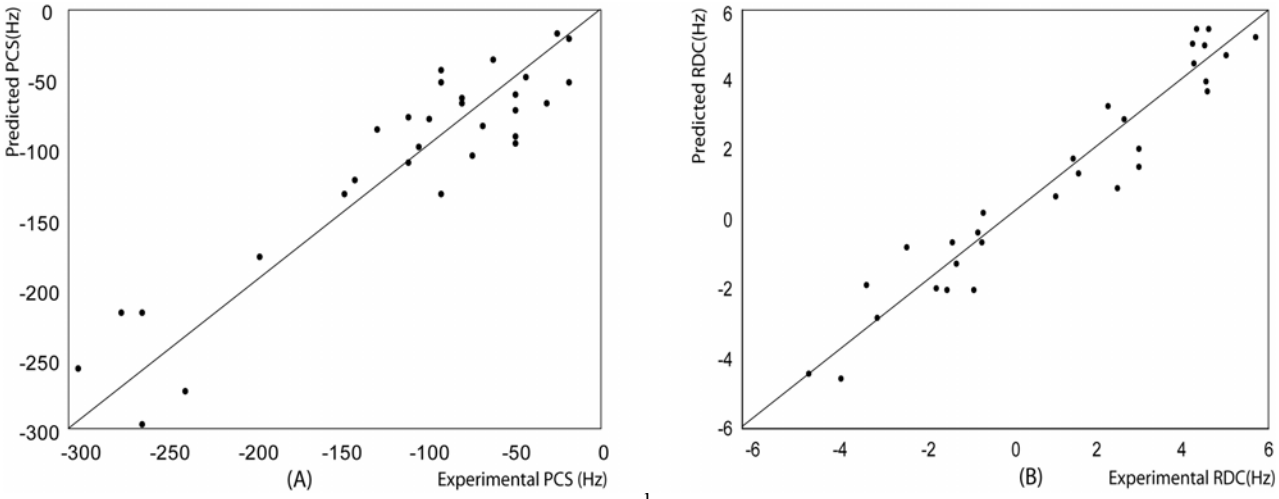

Figure 4.5 Experimental PCSs and RDCs at a ^{1}H frequency of 600 MHz versus back-calculated PCSs and RDCs using structural coordinates determined using Xplor-NIH.

The good agreement using only one structure suggests that the tag can be modeled as rigid with respect to the protein. From the calculated structure of the protein shown in Figure 4.6, the LBT tag is protruded but retains very good surface contact with the protein. This good contact restricts the relative motion of the tag.

4.3.4 Ligand RDC and PCS Measurement

RDCs and PCSs for the ligand were measured from ^{1}H-^{13}C HSQC spectra taken with material in natural abundance. An adequate signal-to-noise ratio is only achievable in these spectra if the ligand concentration is in excess over that of the protein. Under these circumstances, ligand RDCs and PCSs are heavily weighted by the fraction in the free state and measured values are much smaller than their bound values. A compromise between the signal-to-noise ratio and the size of the measured values was reached at a 5:1 ratio of ligand over protein.

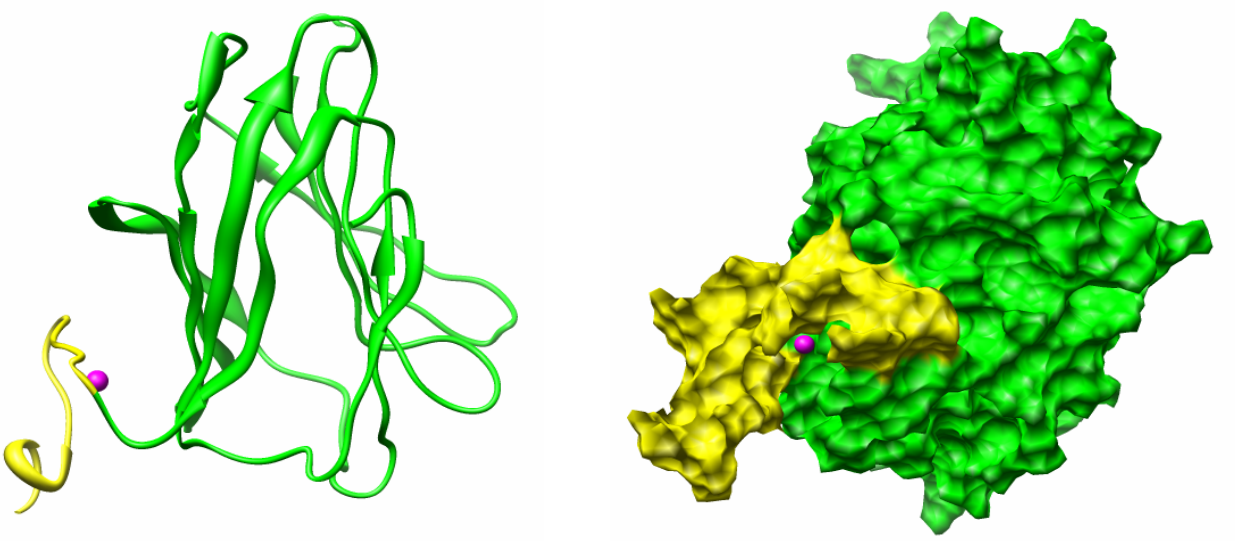

Figure 4.6 Galectin-3-LBT structure calculated from Xplor-NIH, based on the experimental RDCs and PCSs. The yellow fragment is the LBT peptide. The magenta is the metal ion Dy^{3+}.

^{1}H-^{13}C RDCs of the ligand were obtained through the measurement of the splitting difference in the ^{13}C dimension between the sample with diamagnetic Lu^{3+} and the sample with the paramagnetic Dy^{3+}. RDC measurements of C2-H2 of galactose and C5-H5 of glucose taken at 800 MHz are shown in Figure 4.7. To optimize the reliability of measured data, a Bayesian parameter estimation program (XRAMBO) [47] was used for the extraction of RDCs and PCSs. The RDCs and PCSs of lactose measured by XRAMBO are presented in Tables 4.3 & 4.4. Bound-lactose RDCs and PCSs were calculated using the following equations (4.9 and 4.10):

$$RDC_{bound}=RDC_{obs}/f_{bound} \quad (4.9)$$

$$PCS_{bound}=PCS_{obs}/f_{bound} \quad (4.10)$$

Here f_{bound} is the fraction of the bound-state ligand which can be determined from a K_d and total ligand and protein concentrations.

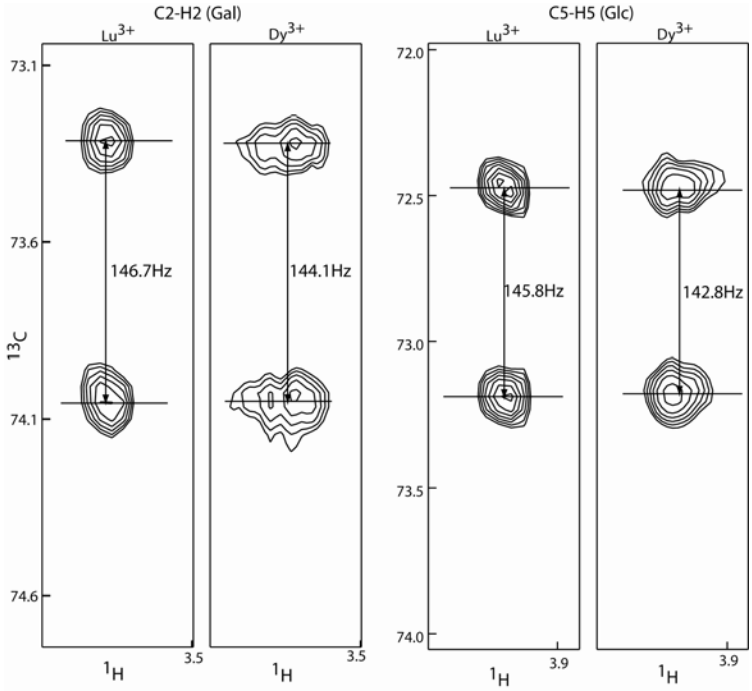

Figure 4.7 RDC measurements of C2-H2 of galactose and C5-H5 of glucose. RDCs were measured as the splitting difference in the ^{13}C dimension between the sample with diamagnetic Lu^{3+} and the sample with paramagnetic Dy^{3+}.

Table 4.3 RDCs for 2.5 mM Lactose with 0.5 mM galectin-3-LBT at a ^{1}H frequency of 800 MHz.

		Lu^{3+}	Dy^{3+}	RDC_{obs}(Hz)	RDC_{bound}(Hz)	Back-calc(Hz)
Gal	C1-H1	163.8±0.1	160.8±0.6	-3±0.7	-15.0	-15.9
	C2-H2	146.7±0.1	144.1±0.4	-2.6±0.5	-13.0	-14.1
	C3-H3	139.7±0.1	137.8±0.3	-2.1±0.4	-10.5	-10.8
	C4-H4	147.7±0.1	149.3±0.6	1.6±0.7	8.0	10.6
	C5-H5	144.7±0.1	141.6±0.4	-3.1±0.5	-15.5	-15.1
	H1-H2	7.84±0.02	7.62±0.05	-0.22±0.07	-4.4	-3.1
Glc(α)	C1-H1	170.8±0.1	167.4±0.2	-3.4±0.3	-17.0	-18.2
	C2-H2	144±0.1	141.3±0.3	-2.7±0.4	-13.5	-10.4
	C3-H3	147.6±0.1	144.2±0.3	-3.4±0.4	-17.0	-11.6
	C5-H5	145.8±0.1	142.8±0.3	-3±0.4	-15.0	-10.4
	H1-H2	3.81±0.02	3.6±0.05	-0.21±0.07	-4.2	-2.4

Table 4.4 PCSs for 2.5 mM lactose with 0.5 mM galectin-3-LBT at a ^{1}H frequency of 800 MHz

		PCS_{obs}(Hz)	PCS_{bound}(Hz)	Back-cal(Hz)
Gal	H1	-10±3	-50±15	-49
	H2	-9±3	-45±15	-49
	H3	-9±3	-45±15	-43
	H4	-11±3	-55±15	-51
	H5	-11±3	-55±15	-59
Glc(α)	H1	-10±1	-50±5	-45
	H2	-11±1	-55±5	-55
	H3	-8±1	-40±5	-46
	H5	-10±1	-50±5	-43

4.3.5 Ligand Alignment Determination.

Coordinates for the heavy atoms in lactose were extracted from a crystal structure of a fungal galectin having lactose-bound (PDB ID:1ULC) [61]. Hydrogens were added using standard geometries, as provided in the "Protonate" tool of the software program AMBER. Using the calculated ^{1}H-^{13}C and ^{1}H-^{1}H RDCs for the bound ligand as shown in Table 4.3 and the extracted coordinates, principal order tensor elements (Sxx Syy Szz) were determined for the

ligand. REDCAT was used with error settings of +/-4 Hz. If measurements and projections based on binding constants are accurate, the order tensor elements determined for the ligand should be identical to those determined for the protein. The order tensor elements found for the ligand after scaling down from ^{1}H 800 MHz to ^{1}H 600 MHz were (Szz=2.6E-04, Syy=-2.4E-04 and Sxx=-2.2E-05). These are very close to those found for the protein (Szz=2.4E-04, Syy=-2.0E-04 and Sxx=-3.5E-05). The actual alignment frame orientations also should be shared between the protein and lactose. When the lactose and the protein coordinates are transformed to their principal alignment frames, the orientation of lactose relative to the protein can be examined. The transformation of lactose actually has a four fold ambiguity because of the insensitivity of RDCs to inversion of axes [62]. However, only one of the possible orientations has an orientation of the galactose ring of lactose that would allow contacts seen in the crystal structure. The actual orientation is 10-15° rotated from the galactose orientation found in the crystal structure.

4.3.6 Ligand-Bound Protein Complex Structure Calculation

Actual docking of the lactose into the galectin-3 binding site requires both orientational and translational constraints. Initially, we explored the ability of PCSs to add the required translational constraints. Because of broadening of the galactose proton resonances, the precision for the galactose resonances is worse than that for the glucose proton resonances. These are presented in Table 4.4. Using our optimized model for galectin-3 with its lanthanide tag, we carried out an XPLOR-NIH calculation in which the ligand and protein were treated as separate, rigid entities. The principal element of order tensor for the protein was used along with all measured RDCs and PCSs for both the protein and ligand. The set of ten minimum energy structures showed the lactose positions to be closely clustered (<1 Å). However, the position

relative to that of N-acetyllactosamine in a galectin-3 crystal structure (PDB ID: 1AK3) does deviate. When we superimposed the protein coordinates for a typical complex on the protein coordinates for 1AK3, the lactose position deviated from that of the N-acetyllactosamine by an RMSD for lactose heavy atoms of 4.3 Å. This is clearly not high accuracy positioning, but it does place the lactose well within the proper binding site.

We were able to measure one intermolecular NOE for the lactose-galectin-3 complex between the H3 of galactose and the amine proton of the W181 side chain, which gave an estimated distance of 2.5+/-0.5 Å [10]. We then added a constraint based on this NOE and repeated the XPLOR calculation. Figure 4.8 shows the final minimal energy structure of the lactose-bound galectin-3 CRD. In comparison with the crystal structure of N-acetyl lactosamine bound galectin-3 CRD, lactose showed a small degree (~10-15^{o}) of rotation, but no van der Waals contacts or hydrogen bond violations were observed. The RMSD (0.46 Å) of the bound ligand was calculated by superimposing 10 structures with minimal energy using VMD [63]. The RMSD of the lactose relative to the X-ray structure for the average of these 10 structures is 1.91 Å when the protein coordinates were superimposed.

4.4 Discussion

In the above studies, we have shown that the incorporation of an EDTA-based tag or lanthanide-binding peptide to align a protein in a high magnetic field can provide substantial RDC measurement. The relative motion of the tag with respect to the protein greatly averages out the measurable RDCs. Several single bonds existing between the EDTA moiety and the protein allow the free motion of the tag. Although RDCs of the EDTA-tagged galectin-3C can be accurately measured for protein structure refinement, the transferred NMR observables of the ligand are too small to measure.

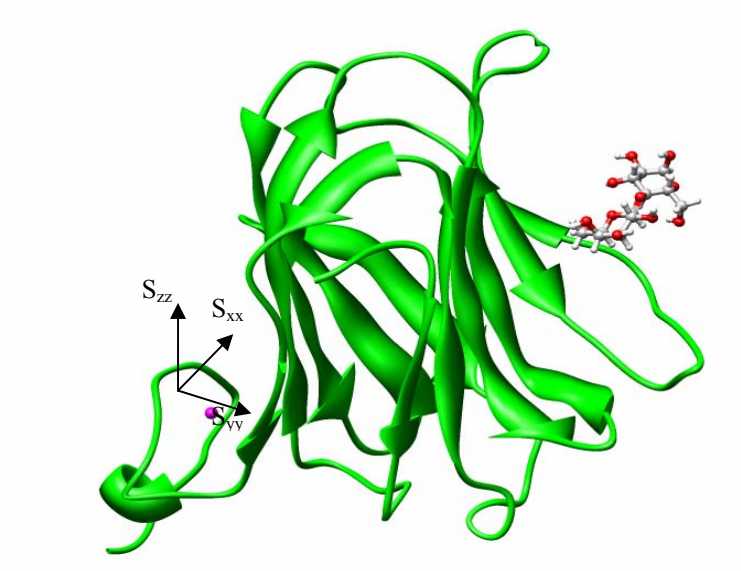

Figure 4.8 Final ligand-protein complex structure determined by XPLOR-NIH using ligand and protein RDCs, PCSs and a single ligand-protein NOE.

In contrast, the incorporation of a lanthanide-binding peptide provides adequate RDCs and PCS constraints to position and orient a ligand bound to the tagged protein. Agreement of the ligand position and orientation is acceptable, despite rather small PCSs and the resulting low precision of distance constraints. The use of PCSs is very important since they provide distance constraints that translationally constrain the ligand in much the same way that protein-ligand NOEs would. However, unlike NOEs, which are proportional to $1/r^6$, PCSs are proportional to $1/r^3$. This, along with a much larger paramagnetic magnetic moment, allows distance constraints of up to 35 Å to be measured accurately. In our case, the glucose residue for which we could measure PCSs is more than 30 Å away from the metal center (based on the calculated ligand bound protein complex) and, despite the small values of PCSs, the constraints were useful. With the inclusion of just a single NOE in the galactose ring, an excellent structure was obtained.

PCSs, as an alternative source of distance information, are particularly important for carbohydrate-protein interactions as the number of observable NOE constraints is often small.

One caveat for our model is the assumption that there is no internal motion between the lanthanide-binding tag and the protein. We attempted to minimize such motions by including no linkage amino acids between the protein and tag. This may not always produce viable constructs or a system that fits a rigid model. There are examples in the literature where motion of tags clearly exists [25] and there is one other example where this motion appears to have been minimized [23]. Some authors have also devised ways to deal with the averaging of some of the observables that result [64].

It is important to note that, even though our data on the protein fit a rigid model well, this does not necessarily imply that there was no motion. There are, in fact, very simple motional models for which averaging produces distance-dependent effects that are analytically identical to those of a rigid model. One involves the motion of the tag over the surface of a sphere having the lanthanide location for the rigid model at the center of the sphere. This can be understood by drawing an analogy between the magnetic field coming from the induced dipole at the lanthanide and the electric field from an electric dipole (A. Redfield, personal communication). The electric dipole can be represented as two point charges and it is readily accepted that the field from a charge distributed uniformly over the surface of a sphere is identical to that from an equivalent point charge at the center. This accidental degeneracy of rigid and motional models was also pointed out for the case of rapid averaging of NOEs a number of years ago [65]. For pseudo-contact shifts, there is the additional complication of averaging the orientation of the susceptibility tensor, and whether such averaging would have the same effects on PCSs and

RDCs. A qualitative argument can be made that these would have the same effect as long as the distance to the site of observation were larger than the radius of the sphere.

One of the main advantages of the lanthanide tag approach is the ability to collect data without the use of hydrophobic alignment media. One of the long-term objectives of applications like this is the investigation of the bound geometry of drug candidates. Very frequently, the binding affinities of drugs are enhanced by the incorporation of hydrophobic moieties [42, 66]. These modified compounds tend to associate directly with many of the hydrophobic media used for RDC alignment. Such associations make data from the free ligand dominate average measurements made on exchanging systems and prevent analysis of bound-state geometries.

One of the continuing limitations of approaches that depend heavily on orientational data is the requirement that sufficient data be collected to completely define the geometry of the system. For RDCs, one must determine five parameters (order, asymmetry, and three angular terms) for each rigid unit under consideration. A relatively large number of interaction vectors giving rise to RDCs must therefore be non-colinear. In carbohydrates, finding and measuring a sufficient number is often difficult because, for example, ^{13}C-^{1}H vectors in β-linked glucose rings are nearly parallel. PCSs, of course, help here as well. However, another potential advantage of lanthanide tags is that different lanthanides can give distinctly different orientational alignments [18, 23, 67]. We have presented only data on Dy^{3+} here. We also collected data using Tb^{3+}, but the alignment tensor was not sufficiently different to warrant inclusion of this information in our current analysis.

References

1. Greer, J., et al., *Application of the 3-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design.* Journal of Medicinal Chemistry, 1994. **37**(8): p. 1035-1054.

2. Henry, C.M., *Structure-based drug design.* Chemical & Engineering News, 2001. **79**(23): p. 69-+.

3. Klyosov, A.A., Z.J. Witczak, and D. Platt, *Carbohydrate Drug Design.* Series: ACS Symposium Series number No. 932, 2006.

4. Barchi, J.J., *Emerging roles of carbohydrates and glycomimetics in anticancer drug design.* Current Pharmaceutical Design, 2000. **6**(4): p. 485-501.

5. Post, C.B., *Exchange-transferred NOE spectroscopy and bound ligand structure determination.* Current Opinion in Structural Biology, 2003. **13**(5): p. 581-588.

6. Moore, J.M., *NMR techniques for characterization of ligand binding: Utility for lead generation and optimization in drug discovery.* Biopolymers, 1999. **51**(3): p. 221-243.

7. Poveda, A., et al., *Solution conformation and dynamics of a fungal cell wall polysaccharide isolated from Microsporum gypseum.* Glycoconjugate Journal, 1998. **15**(3): p. 309-321.

8. Mayer, M. and B. Meyer, *Characterization of ligand binding by saturation transfer difference NMR spectroscopy.* Angewandte Chemie-International Edition, 1999. **38**(12): p. 1784-1788.

9. Meyer, B. and T. Peters, *NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angewandte Chemie-International Edition, 2003. **42**(8): p. 864-890.

10. Zhuang, T.D., H. Leffler, and J.H. Prestegard, *Enhancement of bound-state residual dipolar couplings: Conformational analysis of lactose bound to Galectin-3.* Protein Science, 2006. **15**(7): p. 1780-1790.

11. Shimizu, H., A. Donohue-Rolfe, and S.W. Homans, *Derivation of the bound-state conformation of a ligand in a weakly aligned ligand-protein complex.* Journal of the American Chemical Society, 1999. **121**(24): p. 5815-5816.

12. Seidel, R.D., T.D. Zhuang, and J.H. Prestegard, *Bound-state residual dipolar couplings for rapidly exchanging ligands of His-tagged proteins.* Journal of the American Chemical Society, 2007. **129**(15): p. 4834-4839.

13. Jain, N.U., S. Noble, and J.H. Prestegard, *Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings.* Journal of Molecular Biology, 2003. **328**(2): p. 451-462.

14. Chen, Z.J. and B. Reif, *Measurements of residual dipolar couplings in peptide inhibitors weakly aligned by transient binding to peptide amyloid fibrils.* Journal of Biomolecular Nmr, 2004. **29**(4): p. 525-530.

15. Gaponenko, V., et al., *Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long-range restraints.* Journal of Biomolecular Nmr, 2004. **28**(3): p. 205-212.

16. Allegrozzi, M., et al., *Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 angstrom from the metal ion.* Journal of the American Chemical Society, 2000. **122**(17): p. 4154-4161.

17. Hughes, I.D., et al., *Lanthanide contraction and magnetism in the heavy rare earth elements.* Nature, 2007. **446**(7136): p. 650-653.

18. Bertini, I., et al., *Magnetic susceptibility tenser anisotropies for a lanthanide ion series in a fixed protein matrix.* Journal of the American Chemical Society, 2001. **123**(18): p. 4181-4188.

19. Lee, L. and B.D. Sykes, *Strategies for the Use of Lanthanide Nmr Shift Probes in the Determination of Protein-Structure in Solution - Application to the Ef Calcium-Binding Site of Carp Parvalbumin.* Biophysical Journal, 1980. **32**(1): p. 193-210.

20. Bertini, I., et al., *NMR spectroscopy of paramagnetic metalloproteins.* Chembiochem, 2005. **6**(9): p. 1536-1549.

21. Bertini, I., et al., *Paramagnetism-based versus classical constraints: An analysis of the solution structure of Ca Ln calbindin D-9k.* Journal of Biomolecular Nmr, 2001. **21**(2): p. 85-98.

22. Ascone, I., R. Fourme, and S.S. Hasnain, *Introductory overview: X-ray absorption spectroscopy and structural genomics.* Journal of Synchrotron Radiation, 2003. **10**: p. 1-3.

23. Su, X.C., et al., *Site-specific labelling of proteins with a rigid lanthanide-binding tag.* Chembiochem, 2006. **7**(10): p. 1599-1604.

24. Ikegami, T., et al., *Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions.* Journal of Biomolecular Nmr, 2004. **29**(3): p. 339-349.

25. Wohnert, J., et al., *Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings.* Journal of the American Chemical Society, 2003. **125**(44): p. 13338-13339.

26. Ma, C. and S.J. Opella, *Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy.* Journal of Magnetic Resonance, 2000. **146**(2): p. 381-384.

27. Franz, K.J., M. Nitz, and B. Imperiali, *Lanthanide-binding tags as versatile protein coexpression probes.* Chembiochem, 2003. **4**(4): p. 265-271.

28. Haberz, P., et al., *Two new chiral EDTA-based metal chelates for weak alignment of proteins in solution.* Organic Letters, 2006. **8**(7): p. 1275-1278.

29. Leonov, A., et al., *Convenient synthesis of multifunctional EDTA-based chiral metal chelates substituted with an S-mesylcysteine.* Chemistry-a European Journal, 2005. **11**(11): p. 3342-3348.

30. Rodriguez-Castaneda, F., et al., *Paramagnetic tagging of diamagnetic proteins for solution NMR.* Magnetic Resonance in Chemistry, 2006. **44**: p. S10-S16.

31. Nitz, M., et al., *A powerful combinatorial screen to identify high-affinity terbium(III)-binding peptides.* Chembiochem, 2003. **4**(4): p. 272-276.

32. Martin, L.L., et al., *Rapid combinatorial screening of peptide libraries for the selection of lanthanide-binding tags (LBTs).* Qsar & Combinatorial Science, 2005. **24**(10): p. 1149-1157.

33. Nitz, M., et al., *Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide.* Angewandte Chemie-International Edition, 2004. **43**(28): p. 3682-3685.

34. Martin, L.J., et al., *Double-lanthanide-binding tags: Design, photophysical properties, and NMR applications.* Journal of the American Chemical Society, 2007. **129**(22): p. 7106-7113.

35. Barondes, S.H., *Soluble Lectins - a New Class of Extracellular Proteins.* Science, 1984. **223**(4642): p. 1259-1264.

36. Cooper, D.N.W. and S.H. Barondes, *Evidence for Export of a Muscle Lectin from Cytosol to Extracellular-Matrix and for a Novel Secretory Mechanism.* Journal of Cell Biology, 1990. **110**(5): p. 1681-1691.

37. Akahani, S., et al., *Galectin-3: A novel antiapoptotic molecule with a functional BH1 (NWGR) domain of Bcl-2 family.* Cancer Research, 1997. **57**(23): p. 5272-5276.

38. Yang, R.Y., D.K. Hsu, and F.T. Liu, *Expression of galectin-3 modulates T-cell growth and apoptosis.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(13): p. 6737-6742.

39. Nakahara, S., N. Oka, and A. Raz, *On the role of galectin-3 in cancer apoptosis.* Apoptosis, 2005. **10**(2): p. 267-275.

40. Seetharaman, J., et al., *X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution.* Journal of Biological Chemistry, 1998. **273**(21): p. 13047-13052.

41. Umemoto, K. and H. Leffler, *Letter to the Editor: Assignment of H-1, N-15 and C-13 resonances of the carbohydrate recognition domain of human galectin-3.* Journal of Biomolecular Nmr, 2001. **20**(1): p. 91-92.

42. Tejler, J., H. Leffler, and U.J. Nilsson, *Synthesis of O-galactosyl aldoximes as potent LacNAc-mimetic galectin-3 inhibitors.* Bioorganic & Medicinal Chemistry Letters, 2005. **15**(9): p. 2343-2345.

43. Leffler, H., F.R. Masiarz, and S.H. Barondes, *Soluble Lactose-Binding Vertebrate Lectins - a Growing Family.* Biochemistry, 1989. **28**(23): p. 9222-9229.

44. Kontaxis, G., G.M. Clore, and A. Bax, *Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times.* Journal of Magnetic Resonance, 2000. **143**(1): p. 184-196.

45. Tian, F., P.J. Bolon, and J.H. Prestegard, *Intensity-based measurement of homonuclear residual dipolar couplings from CT-COSY.* Journal of the American Chemical Society, 1999. **121**(33): p. 7712-7713.

46. Delaglio, F., et al., *Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes.* Journal of Biomolecular Nmr, 1995. **6**(3): p. 277-293.

47. Andrec, M. and J.H. Prestegard, *Metropolis Monte Carlo implementation of Bayesian time-domain parameter estimation: Application to coupling constant estimation from antiphase multiplets.* Journal of Magnetic Resonance, 1998. **130**(2): p. 217-232.

48. Valafar, H. and J.H. Prestegard, *REDCAT: a residual dipolar coupling analysis tool.* Journal of Magnetic Resonance, 2004. **167**(2): p. 228-241.

49. Banci, L., et al., *Paramagnetism-based restraints for Xplor-NIH.* Journal of Biomolecular Nmr, 2004. **28**(3): p. 249-261.

50. Schwieters, C.D. and G.M. Clore, *Internal coordinates for molecular dynamics and minimization in structure determination and refinement.* Journal of Magnetic Resonance, 2001. **152**(2): p. 288-302.

51. Tjandra, N., J. Marquardt, and G.M. Clore, *Direct refinement against proton-proton dipolar couplings in NMR structure determination of macromolecules.* Journal of Magnetic Resonance, 2000. **142**(2): p. 393-396.

52. Alonso-Plaza, J.M., et al., *NMR investigations of protein-carbohydrate interactions: insights into the topology of the bound conformation of a lactose isomer and beta-galactosyl xyloses to mistletoe lectin and galectin-1.* Biochimica Et Biophysica Acta-General Subjects, 2001. **1568**(3): p. 225-236.

53. Cambillau, C., *3D structure. 1. The structural features of protein-carbohydrate interactions revealed by X-ray crystallography in: J. Montreuil, J.F.G. Vliegenthart, H. Schachter (Eds.),* . Glycoproteins. 1995. 29-65.

54. Jensen, M.R., et al., *On the use of pseudocontact shifts in the structure determination of metalloproteins.* Magnetic Resonance in Chemistry, 2006. **44**(3): p. 294-301.

55. John, M., et al., *Weak alignment of paramagnetic proteins warrants correction for residual CSA effects in measurements of pseudocontact shifts.* Journal of the American Chemical Society, 2005. **127**(49): p. 17190-17191.

56. Barbieri, R., C. Luchinat, and G. Parigi, *Backbone-only protein solution structures with a combination of classical and paramagnetism-based constraints: A method that can be scaled to large molecules.* Chemphyschem, 2004. **5**(6): p. 797-806.

57. Kurland, R.J. and B.R. McGarvey, *Isotropic NMR shifts in transient metal complex:The Calculation of the Fermi contact and Pseudocontact Terms.* Journal of Magnetic Resonance, 1970. **2**(3): p. 286-301.

58. Prestegard, J.H., C.M. Bougault, and A.I. Kishore, *Residual dipolar couplings in structure determination of biomolecules.* Chemical Reviews, 2004. **104**(8): p. 3519-3540.

59. Keizers, P.H.J., et al., *Increased paramagnetic effect of a lanthanide protein probe by two-point attachment.* Journal of the American Chemical Society, 2007. **129**(30): p. 9292-+.

60. Bax, A., *Weak alignment offers new NMR opportunities to study protein structure and dynamics.* Protein Science, 2003. **12**(1): p. 1-16.

61. Walser, P.J., et al., *Structure and functional analysis of the fungal galectin CGL2.* Structure, 2004. **12**(4): p. 689-702.

62. Al-Hashimi, H.M., et al., *Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings.* Journal of Magnetic Resonance, 2000. **143**(2): p. 402-406.

63. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**(1): p. 33-&.

64. Iwahara, J., C.D. Schwieters, and G.M. Clore, *Ensemble approach for NMR structure refinement against H-1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule.* Journal of the American Chemical Society, 2004. **126**(18): p. 5879-5896.

65. Lemaster, D.M., et al., *Protein Dynamics and Distance Determination by Noe Measurements.* Febs Letters, 1988. **236**(1): p. 71-76.

66. Sorme, P., et al., *Low micromolar inhibitors of galectin-3 based on 3 '-derivatization of N-acetyllactosamine.* Chembiochem, 2002. **3**(2-3): p. 183-189.

67. Pintacuda, G., et al., *NMR structure determination of protein-ligand complexes by lanthanide labeling.* Accounts of Chemical Research, 2007. **40**(3): p. 206-212.

Chapter 5

Conclusions and future directions

Ligand-protein interactions depend upon the specificity and stability of the ligand binding. Understanding the mechanisms underlying this type of molecular recognition is critical for successful drug design [1]. It is clear that this understanding can be used to improve binding affinities, but there are other issues as well. A drug candidate is normally selected after a series of lead compound optimizations to also improve its solubility, bioavailability, metabolic profile and half-life. The availability of a three-dimensional structure of the lead-bound protein can help with these optimizations as well. Knowing the shape and chemical properties of the lead molecule and the protein active site can help in identifying allowable structural modifications in achieving both enhanced binding and these additional properties.

Structure determinations based on NOEs and transferred NOEs are reliable reporters on bound ligand geometry and are the basic NMR methods used in structure-based drug molecule design. However, when there are no or very few intermolecular NOEs, and when information about protein binding site location is needed, other methods must be implemented. In this thesis, three alternate approaches for accurately determining the structure of lactose-bound galectin-3 were described. These approaches do not rely only on measuring a number of intermolecular NOEs to generate distance constraints. Other constraints, mainly from RDCs, are measured to find the relative orientation of the ligand with respect to its protein receptors, and additional long-range structure restraints are derived from pseudo-contact shifts which restrain the intermolecular distance similar to intermolecular NOEs.

Weakly-binding fast-exchanging systems, which are often found in small carbohydrate-lectin interactions, complicate structure determinations due to the dominant contributions of the free-state ligand to the observables. The NMR observables of the ligand, such as chemical shifts and RDCs, are the average of the free-state and bound-state properties, weighted by the fraction of the free- and bound-states. To extract reliable bound-state data, we suggest enhancing bound-state observables ("Methods" section, chapters 2 and 3) or minimizing free-state observables (chapter 4). With these methods, information about the bound-state ligands dominates the observed average.

Most drug molecules are either hydrophobic or carry a hydrophobic moiety which increases their affinity and permeability across the cell membrane. The methods described in chapters 2 and 3 have limited applications for hydrophobic ligands, while the paramagnetism-based method described in chapter 4 is applicable to both hydrophobic and hydrophilic ligands. Our lab and other research groups are continuing to explore the possibility of improving the application of paramagnetism-based constraints induced through the incorporation of a rigid lanthanide binding tag. As discussed in chapter 4, some of the most useful features include the fact that the ligand pseudo-contact shifts can be measured up to 35 Å away from the paramagnetic center. They are however quite small at this distance, as are the RDCs from field induced alignment of these centers. In our application we supplemented data with one additional intermolecular NOE to improve definition of the ligand binding geometry. However, improvements directed at increasing the size of PCSs and RDCs would also alleviate this problem.

5.1 Application and Development of New Methods

Pseudo-contact shifts drop as the cube of the distance between the nucleus of interest and the paramagnetic metal center. While this is steep, it is not as steep as the inverse sixth power of distance seen in NOEs and PREs. In our case, the ligand was more than 30 Å away from the metal ion so the pseudo-contact shifts were already very small. However, one way to increase the pseudo-contact shifts is to shorten the distance by moving the site of tag attachment. The amino acid sequence of the protein can be mutated to introduce a surface cysteine. This is similar to what we did in chapter 2 to allow introduction of a terminal propyl chain [2]. This method has been exploited by Dr. Otting's group. In their studies, the lanthanide-binding peptide tag was attached to the N-terminal DNA-binding domain of an *Escherichia coli* arginine repressor through a disulfide bond. The resulting protein contained only a single cysteine residue which encouraged specific attachment. The tag was determined to be rigid. Up to 2.9 ppm pseudo-contact shifts induced by Tm^{3+} were observed. Even for nuclei 30Å away from the metal ion, the pseudo-contact shifts were as large as 0.2 ppm [3]. This method also would work well for the galectin-3 CRD since there is only one inaccessible cysteine. A site mutation of serine into cysteine allows for the specific attachment of the lanthanide-binding tag to the protein. This fine-tuning of the distance could potentially increase the measurement of pseudo-contact shifts from several Hz to several hundreds of Hz for fast-exchanging ligands.

Another means of potentially improving the magnitude of the measurable effects is by making the attachment of the tag more rigid. Internal motions average effective susceptibility anisotropies of the tag and reduce PCSs and RDCs. They also can make PCSs more difficult to interpret. In our research, the lanthanide-binding peptide was attached to the C-terminus of

galectin-3 CRD by a single short linker. Although the tag proved to be fairly restricted, this will not always necessarily be the case.

Interestingly, a two-site attachment of the lanthanide-binding tag has been developed in Ubbink's group [4]. In their work, a caged lanthanide NMR probe (cLaNP) was synthesized based on a metal chelate 1,4,7,10-tetraazacyclododecane-1-[*N*-oxidopyridine-2-yl)methyl]-4,7,10-triacetic acid (H_3DO3A-pyNox) . The two functional methanethiolsulfonate moieties specifically reacted with two cysteines on the protein surface. The two-site attachment greatly reduces the internal motion of the tag and allows up to 6 Hz RDCs at a 14.1T magnetic field if Yb^{3+} is used [5]. It is still small compared with the RDCs of metalloproteins which show little internal motion of the metal ion. However, compared with a single-site attachment, the order of alignment was increased by a factor of 2.5. The abundance of σ bonds between the cysteine side chains and the metal chelate apparently allows a significant degree of freedom. This might not be the case for a lanthanide-binding peptide if the tag has multiple contact sites with the protein.

The lanthanide-binding peptide usually attaches to the C- or N-terminal of the protein providing one attachment site. The flexibility of the tag can be reduced by introducing a second contact point between the tag and the protein. This could be a cysteine-cysteine disulfide bond if appropriate cysteine residues are introduced. However, it could also rely on other types of interaction. As shown in our case, the lanthanide-binding tag has a very good hydrophobic surface contact with the protein which helps reduce the internal motion of the tag. Hydrogen bonding, and ionic interactions, could also contribute a second interaction point and improve the rigidity of the attached tag. The lanthanide-binding tag itself is very hydrophobic. Carefully designing a hydrophobic patch in other proteins close to the lanthanide-binding tag might increase the hydrophobic interactions between the protein and the tag. Tag to protein ion pair

interactions could also be introduced. In principle, this is similar to reducing the internal motion by forming a bidentate complex, the technique used by Ubbink's group.

A different method for utilizing a double-lanthanide-binding tag (dLTB) has been developed by Imperiali et al [6, 7]. In her lab, a short peptide was designed to bind two lanthanide ions with strong binding affinity. The initial purpose of this double-lanthanide-binding tag was to solve the phase problem for the structure determination of ubiquitin by single-wavelength anomalous diffraction [7]. The single-lanthanide-binding tag failed in the structure determination due to the relative motion of the tag with respect to the protein. The double-lanthanide-binding tag, co-expressed with the protein, proved to be ordered enough for the phasing in structure determination. The sequence of this double-lanthanide-binding tag is GPGYIDTNNDGWIEGDELYIDTNNDGWIEGDELLA which is, essentially, the single-lanthanide-binding tag repeated twice. The double-lanthanide-binding tag had little impact on the protein structure and functions after being co-expressed. Theoretically, the binding of two lanthanide ions could induce up to a 2-fold increase in magnetic anisotropy, provided that the magnetic susceptibility tensors of these two lanthanide ions share an alignment frame. In the NMR study of this double-lanthanide-binding tag, substantial enhancement of the protein alignment was observed due to minimal motion of the tag relative to the protein. However, the interpretation of NMR data from dLTB-tagged proteins can be quite complicated relative to that for a single lanthanide binding tag. RDCs, which are distance-independent, can be interpreted as usual except that the determined order tensors are the average of two metal ions. Other paramagnetism-based constraints, such as pseudo-contact shifts, paramagnetic relaxation enhancement, and cross correlation between the dipole-dipole and Curie relaxation, are distance-

dependent. If averaged order tensors are used, the distance will be some effective distance between the nucleus of interest and the origin of the averaged order tensors.

5.2 New Galectin-3 Inhibitor Studies

5.2.1 Lactose Derivatives

The availability of both X-ray and NMR structures of lactose-bound galectin-3 provides the basis for the development of galectin-3 inhibitors. Several effective galactose and lactose derivatives have been synthesized at Dr. Nilsson's group at Lund University [8-11]. Lactose, as a natural ligand, binds weakly to the galectin-3 carbohydrate recognition domain. The binding site of galectin-3 CRD is open-ended and groove-like. The galactose residue is not deeply buried in the binding site so a large interaction group (e.g. benzyl) can be accommodated. In principle, the binding affinity increases when additional intermolecular interactions are introduced if the pre-existing interaction network is maintained. That was the case for all of the lactose derivatives synthesized by Dr. Nilsson's group. The binding affinity increased dramatically when the 3-position of the galactose moiety was substituted by an aromatic amide. One theory is that additional CH_2-π interactions arise at the interface of the aromatic ring with the side chains of one or two arginine residues in the vicinity of the binding site. Arginine is highly conserved in the carbohydrate recognition domain within the galectin family. From the perspective of NMR, we are interested in several issues related to these interactions: how the benzyl ring is oriented to maximize the arginine-arene interaction, whether the ligand undergoes significant conformation change, and whether the original interaction network is maintained after the binding of the lactose derivatives.

To understand this arginine-arene interaction from a structural point of view, we collaborated with Dr. Leffer at Lund University and obtained one of the N-acetyl lactosamine

derivatives for an NMR study. This molecule has a 50 μM binding affinity with galecitn-3 CRD, about four times stronger than lactose. The chemical structure of this compound is shown in Figure 5.1. Although the binding affinity is moderate, our preliminary results indicate that it is in the slow exchange regime. The bound-ligand experiences the exchange broadening. An excess of ligands does not help in this case because the detectable free ligands do not carry bound-state information. Ideally, we anticipate being able to promote fast exchange of this ligand through changes in experimental conditions. However methods applicable to slow exchanging system can also be explored.

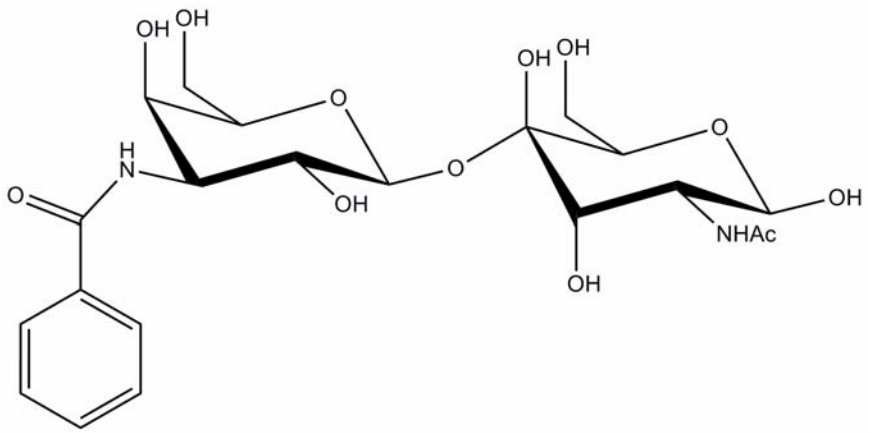

Figure 5.1 N-acetyl lactosamine derivative. The binding affinity is significantly increased due to the additional interaction between the aromatic benzyl group and the protein.

5.2.2 Multivalent Ligands for Galectin-3

All of our studies described above were based on the simple disaccharide lactose. We anticipate that future application will be to larger molecules and ones of more immediate biological interest. Carbohydrates usually appear in more complex forms, that are either larger or are glycocojugates such as glycolipids or glycoproteins. The structural details of the

interactions between galectin-3 and these more complex carbohydrates are hardly known. A simple divalent ligand Gal-α(1,3)Gal-β(1,4)GlcNAc-R, a xenoantigen of xenotransplantation is a great target to begin the study of the interaction between Galectin-3 and multivalent oligosaccharides.

Xenotransplantation to humans is considered to be successful only after the two stages of rejection, hyperacute and delayed xenograft rejection (DXR), are overcome [12]. The mechanisms of hyperacute rejection have been well studied but the mechanism of DXR remains unclear. A lectin-dependent recognition pathway is believed to be involved in the monocyte and NK cell infiltration which is one of the four major interrelated pathways for the development of DXR [13, 14]. It has been demonstrated that monocytes can bind to the major xenoantigen Gal-α(1,3)Gal-β(1,4)GlcNAc-R . Galectin-3 mediates this interaction. This xenoantigen carries a beta-galactoside ((Gal-β(1–3/4)GlcNAc) that can be specifically recognized by galectin-3. The fact that galectin-3 is often found in mammalian cells including monocytes and macrophages also supports this theory.

The structural details of the interaction between α-Gal and galectin-3 remain unknown. Hypotheses have been proposed: galectin-3 binds primarily to N-acetyl lactosamine while the terminal galactose might find its binding sites in the carbohydrate recognition domain [15] or make contacts through the folding of the flexible N-terminal domain [16].

NMR offers a good opportunity to gain insights into this interaction on an atomic level. The availability of a xenoantigen Gal-α(1,3)Gal-β(1,4)GlcNAc-R mimic (Figure 5.2) provided by our collaborator, Dr. Geert-Jan Boons, allowed a preliminary NMR study of galectin-3 interaction with this trisaccharide. Preliminary NMR studies show that this trisaccharide strongly binds with the galectin-3 carbohydrate recognition domain. Future structural studies of

this trisaccharide-galectin-3 interaction could provide a model for the interaction of other multivalent ligands with galectin-3.

In conclusion, the results obtained in this thesis showed that, even in the case of sparse intermolecular NOEs, the relative orientation or geometry of bound-state carbohydrates was well restricted through the inclusion of either residual dipolar couplings or paramagnetism-based constraints. The accuracy of the bound-state data was improved through alterations in protein constructs that allowed data from the bound-state ligand to dominate measured average parameters. We tested our methods on a weakly-binding, fast-exchanging galectin-3-lactose system. Yet, we believe that these same methods will be applicable to more complex and more biologically relevant fast-exchanging systems. The observations on these new systems are expected to contribute to lead optimization in structure-based drug discovery.

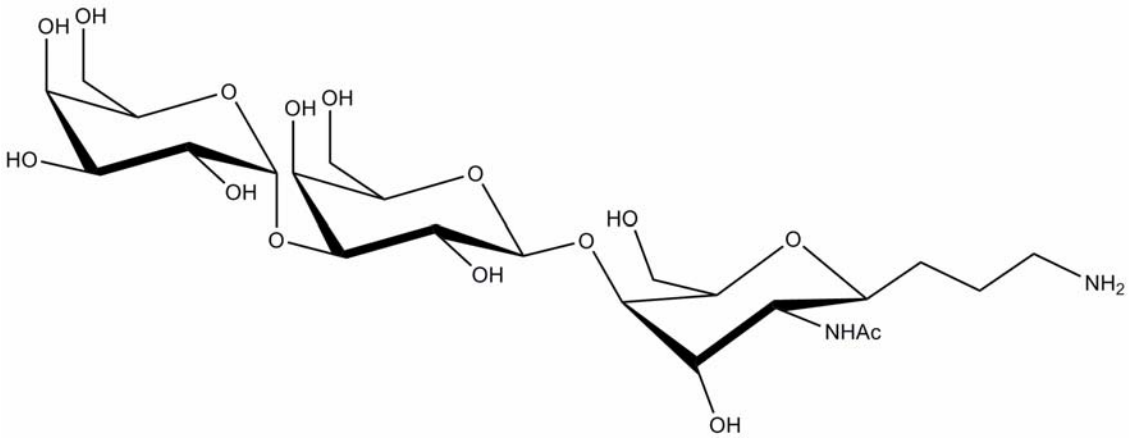

Figure 5.2 Chemical structure of a xenoantigen Gal-α(1,3)Gal-β(1,4)GlcNAc-R mimic. It consists of two galactose residues and an N-acetyl glucosamine residue. The reducing end of the glucose residue is protected by a propyl amine.

References:

1. Hans-Joachim Böhm and G. Schneider, *Protein-Ligand Interactions: From Molecular Recognition to Drug Design*, ed. Raimund Mannhold, Hugo Kubinyi , and G. Folkers. 2003: Wiley. 262.

2. Zhuang, T.D., H. Leffler, and J.H. Prestegard, *Enhancement of bound-state residual dipolar couplings: Conformational analysis of lactose bound to Galectin-3.* Protein Science, 2006. **15**(7): p. 1780-1790.

3. Su, X.C., et al., *Site-specific labelling of proteins with a rigid lanthanide-binding tag.* Chembiochem, 2006. **7**(10): p. 1599-1604.

4. Keizers, P.H.J., et al., *Increased paramagnetic effect of a lanthanide protein probe by two-point attachment.* Journal of the American Chemical Society, 2007. **129**(30): p. 9292-+.

5. Polasek, M., et al., *Lanthanide(III) complexes of a pyridine N-oxide analogue of DOTA: exclusive M isomer formation induced by a six-membered chelate ring.* Chemical Communications, 2004(22): p. 2602-2603.

6. Martin, L.J., et al., *Double-lanthanide-binding tags: Design, photophysical properties, and NMR applications.* Journal of the American Chemical Society, 2007. **129**(22): p. 7106-7113.

7. Silvaggi, N.R., et al., *Double-lanthanide-binding tags for macromolecular crystallographic structure determination.* Journal of the American Chemical Society, 2007. **129**(22): p. 7114-7120.

8. Cumpstey, I., et al., *Studies of arginine-arene interactions through synthesis and evaluation of a series of galectin-binding aromatic lactose esters.* Chembiochem, 2007. **8**(12): p. 1389-1398.

9. Sorme, P., et al., *Low micromolar inhibitors of galectin-3 based on 3 '-derivatization of N-acetyllactosamine.* Chembiochem, 2002. **3**(2-3): p. 183-189.

10. Sorme, P., et al., *Structural and thermodynamic studies on cation - II interactions in lectin-ligand complexes: High-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction.* Journal of the American Chemical Society, 2005. **127**(6): p. 1737-1743.

11. Cumpstey, I., et al., *C-2-symmetrical thiodigalactoside bis-benzamido derivatives as high-affinity inhibitors of galectin-3: efficient lectin inhibition through double arginine-arene interactions.* Angewandte Chemie-International Edition, 2005. **44**(32): p. 5110-5112.

12. McGregor, C.G.A., et al., *Cardiac xenotransplantation: Progress toward the clinic.* Transplantation, 2004. **78**(11): p. 1569-1575.

13. Hancock, W.W., *Delayed xenograft rejection.* World Journal of Surgery, 1997. **21**(9): p. 917-923.

14. Jin, R.Y., et al., *Human monocytes recognize porcine endothelium via the 14 interaction of Galectin 3 and alpha-GAL.* Journal of Immunology, 2006. **177**(2): p. 1289-1295.

15. Seetharaman, J., et al., *X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution.* Journal of Biological Chemistry, 1998. **273**(21): p. 13047-13052.

16. Barboni, E.A.M., et al., *Molecular modeling and mutagenesis studies of the N-terminal domains of galectin-3: evidence for participation with the C-terminal carbohydrate recognition domain in oligosaccharide binding.* Glycobiology, 2000. **10**(11): p. 1201-1208.

Appendices

Appendix A: An example of AIR restraints definition

```
! HADDOCK AIR restraints for 1st partner
assign ( resid 144  and segid A)
      (
       ( resid 2  and segid B)
    or
       ( resid 3  and segid B)
      )  2.0 2.0 0.0
!
assign ( resid 162  and segid A)
      (
       ( resid 2  and segid B)
    or
       ( resid 3  and segid B)
      )  2.0 2.0 0.0
!
assign ( resid 165  and segid A)
      (
       ( resid 2  and segid B)
    or
       ( resid 3  and segid B)
      )  2.0 2.0 0.0
!
assign ( resid 181  and segid A)
      (
       ( resid 2  and segid B)
    or
       ( resid 3  and segid B)
      )  2.0 2.0 0.0
!
assign ( resid 184  and segid A)
      (
       ( resid 2  and segid B)
    or
       ( resid 3  and segid B)
      )  2.0 2.0 0.0
!
assign ( resid 186  and segid A)
      (
```

```
       ( resid 2  and segid B)
     or
       ( resid 3  and segid B)
      )  2.0 2.0 0.0
!
! HADDOCK AIR restraints for 2nd partner
!
assign ( resid 2  and segid B)
      (
       ( resid 144  and segid A)
     or
       ( resid 162  and segid A)
     or
       ( resid 165  and segid A)
     or
       ( resid 181  and segid A)
     or
       ( resid 184  and segid A)
     or
       ( resid 186  and segid A)
     or
       ( resid 180  and segid A)
     or
       ( resid 182  and segid A)
     or
       ( resid 183  and segid A)
     or
       ( resid 187  and segid A)
     or
       ( resid 166  and segid A)
     or
       ( resid 143  and segid A)
     or
       ( resid 164  and segid A)
      )  2.0 2.0 0.0
!
assign ( resid 3  and segid B)
      (
       ( resid 144  and segid A)
     or
       ( resid 162  and segid A)
     or
       ( resid 165  and segid A)
     or
       ( resid 181  and segid A)
     or
```

```
   ( resid 184  and segid A)
or
   ( resid 186  and segid A)
or
   ( resid 180  and segid A)
or
   ( resid 182  and segid A)
or
   ( resid 183  and segid A)
or
   ( resid 187  and segid A)
or
   ( resid 166  and segid A)
or
   ( resid 143  and segid A)
or
   ( resid 164  and segid A)
  )  2.0 2.0 0.0
```

Appendix B: An nmrPipe processing script for maintaining both the real and imaginary data

```
#!/bin/csh
nmrPipe  -in test.fid              \
| nmrPipe -fn SOL   \
| nmrPipe  -fn SP -off 0.5 -end 0.99 -pow 2 -c 1.0  \
| nmrPipe -fn FT  -ai          \
| nmrPipe -fn PS -p0 111  -p1 0     \
| nmrPipe  -fn EXT -x1 3.0ppm -xn 5.50ppm -verb 2 -sw  \
| nmrPipe -fn TP     -hyper        \
| nmrPipe -fn FT  -ai               \
| nmrPipe -fn PS -p0 -118 -p1 0       \
| nmrPipe -fn TP    -hyper     \
| nmrPipe -out temp_withim.ft2  -ov
```

Apendix C: An M-based nmrPipe macro for the reconstruction of 1D projection

```
/* this macro is intended to extract a portion of a 2D spectrum including imaginary data,
following a 1D projection using nmrPipe. The data after the projection should be reshuffled
using this macro to get the correct format. */
   FDATASIZE = 512;
   FDQUADFLAG =106;
   NDQUADFLAG = 1007;
   NDSW = 1003;
   NDAPOD = 1002;
   NDORIG = 1004;
   NDCAR = 1013;
   NDX1 = 1022;
   NDXN =1023;
   NDCAR = 1013;

   char  thisName[256];
   char  thatName[256];
   char  otherName[256];
   float fdata[FDATASIZE];
   float fdata2[FDATASIZE];

/* Extracting header parameters from 1D dataset without imaginary */
   (void) sprintf (thisName,hdrName, 1);
   inFD  = dOpen( thisName, "r" );
   (void)  dReadHdr( inFD, fdata, 0 );
   (void) dClose( inFD );

   old_x_size = getParm( fdata, NDSIZE, CUR_XDIM);
   (void) setParm(fdata, NDQUADFLAG, 0, CUR_XDIM);
   (void) setParm(fdata, NDSIZE, old_x_size, CUR_XDIM);
```

```
    (void) setParm( fdata, FDQUADFLAG, 0, CUR_XDIM);
    (void) sprintf( thatName, outName,1);
    outFD = dOpen( thatName, "w" );
    (void) dWrite( outFD, fdata, wordLen*FDATASIZE );

/* Extracting header parameters from the 1D dataset with imaginary */
    (void) sprintf (otherName, inName, 1);
    in2FD  = dOpen( otherName, "r" );
    (void) dReadHdr( in2FD, fdata2, 0 );
    old_xsize = getParm(fdata2, NDSIZE, CUR_XDIM);
    new_xsize=old_xsize/2;
    (void) dClose( in2FD);
    planein_size = wordLen*old_xsize;
    skip_size=wordLen*FDATASIZE;

    float planein[planein_size];
    float fidout[wordLen*old_xsize];
    float xdimft_real[wordLen*old_xsize];
    float xdimft_im[wordLen*old_xsize];
    float xpoint[wordLen];
    float xpoint_im[wordLen];
    float temp[wordLen*new_xsize];
    float temp_im[wordLen*new_xsize];

/* Reading in the big ft2 file while skipping the header and beginning lines */
    in2FD = dOpen( otherName, "r");
    (void) dSeek( in2FD, skip_size);
    (void) dReadB( in2FD, planein, planein_size);
    (void) dClose( in2FD);
    offsetin = 0;
    offsetout = 0;
```

```
    /* Reading the real component */
    for( j = 0; j < new_xsize; j ++)
                {
                offset_x = 2*j;
                vvCopyOff( temp, planein, 1, j, offset_x);
                 }
    (void) vvCopyOff( fidout,temp, new_xsize, offsetout , 0);

    /* Reading the imaginary component */
    for( k = 0; k < new_xsize; k ++)
                {
                offset_2 = 2*k+1;
                vvCopyOff( temp_im, planein, 1, k, offset_2);
                }
        offsetout2 = offsetout + new_xsize;
        (void) vvCopyOff( fidout, temp_im, new_xsize, offsetout2 , 0);
   (void) dWriteB( outFD, fidout, wordLen*old_xsize);
   (void) dClose( outFD);
printf("Done. \n");
```