

Determination of the binding mode of carbohydrates bound to galectin-1: combination of data from saturation transfer difference NMR and molecular modeling procedures.

Dissertation

zur Erlangung des Doktorgrades
des Department Chemie der Universität Hamburg

vorgelegt von

Jan-Christoph Westermann

aus Hamburg

Hamburg, 2006

UH



Universität Hamburg

1. Gutachter: Prof. Dr. Bernd Meyer
2. Gutachter: Prof. Dr. Volkmar Vill
Tag der Disputation: 29. September 2006

“Science may be described as the art of systematic oversimplification.”

– Sir Karl R. Popper

Die Vorliegende Arbeit entstand in der Zeit vom August 2001 bis Juli 2006 in der Arbeitsgruppe von Prof. Dr. Bernd Meyer am Institut für Organische Chemie des Departments Chemie der Universität Hamburg.

Herrn Prof. Dr. Bernd Meyer danke ich für die Überlassung des interessanten Themas, die wertvolle Unterstützung bei der Durchführung und die tolle Zeit in seiner Arbeitsgruppe

Table of contents

Table of contents	I
I. Abbreviations	III
II. Amino acids	IV
III. Carbohydrate monomers	IV
IV. Carbohydrate oligomers	V
V. Protein identification	V
1 Introduction	1
1.1 Drug discovery and development	1
1.2 Molecular modeling	3
1.2.1 Docking	3
1.3 NMR screening	6
1.4 Saturation transfer difference nuclear magnetic resonance (STD NMR)	8
1.4.1 Structural information using Overhauser effects and selective labeling on performing STD experiments: SOS-NMR	12
1.4.2 Complete relaxation and conformational exchange matrix (CORCEMA) analysis of STD effects	13
1.5 Galectins	13
2 Scope of the thesis	16
3 Results and discussion	17
3.1 Choice of a model system	17
3.2 Saturation of galectin-1	17
3.3 Epitope mapping of disaccharides	18
3.3.1 Methyl- β -D-lactoside	18
3.3.2 <i>N</i> -acetyl lactosamine (type II)	23
3.3.3 Lacto- <i>N</i> -biose (<i>N</i> -acetyl lactosamine type I)	25
3.3.4 Phenyl- β -D-galactoside	27
3.4 Estimation on the binding kinetics based on STD NMR results	29
3.5 Development of a docking procedure	29
3.5.1 Docking of lactose and lactose derivatives	34
3.5.2 Docking phenyl glycosides of β -D-galactose	35
3.5.3 Analysis of the conformation of docked <i>N</i> -acetyl lactosamine	39
3.5.4 Docking of lacto- <i>N</i> -biose	42
3.6 Docking constraints obtained from STD NMR experiments	44
3.6.1 Automated calculation of STD effects for docking results	45
3.6.2 Comparison of calculated and experimentally determined STD effects	47
3.6.3 Comparison of calculated STD for docked poses with results from STD NMR	51
3.6.4 Selection of docked LNB poses based on STD NMR	54
3.6.5 Binding mode of 3'-Sialyllactose	58
3.7 Influence of the histidine protonation state on the docking results	66

3.7.1	Docking of <i>N</i> -acetyl lactosamine with different histidine types in the receptor model _____	66
3.7.2	Docking of 2,3-sialyl lactose with different histidine types in the receptor model _____	68
3.7.3	Binding mode of lacto- <i>N</i> -tetraose _____	70
3.8	Molecular dynamics simulations of galectin-1 complexes _____	83
3.8.1	MD simulation of <i>N</i> -acetyl lactosamine and galectin-1 _____	84
3.8.2	Complexes of galectin-1 with lacto- <i>N</i> -tetraose _____	86
3.8.3	Constrained MD simulations of LNT in complex with galectin-1 _____	89
3.8.4	Simulation of the complex of sialyllactose and galectin-1 _____	95
4	Conclusions _____	100
5	Zusammenfassung _____	103
6	Experimental section _____	106
6.1	Chemicals _____	106
6.2	Toxicology _____	106
6.3	Buffers _____	106
6.3.1	TRIS buffer _____	106
6.3.2	PBS buffer _____	106
6.3.3	Preparation of galectin-1 _____	107
6.3.4	Preparation of ligand stock solutions _____	107
6.3.5	NMR samples _____	107
6.4	NMR recording and processing _____	107
6.5	NMR assignments of carbohydrates _____	108
6.5.1	Assignment for LNT _____	108
6.5.2	Assignment for 3'-sialyllactose _____	108
6.5.3	Assignment for ethyl- β -D-lactosamine (LacNAcOEt) _____	109
6.5.4	Assignment for lacto- <i>N</i> -biose (LNB) _____	109
6.5.5	Assignment for methyl- β -D-lactoside (LacOMe) _____	109
6.5.6	Assignment for phenyl- β -D- galactoside (PheGal) _____	110
6.6	Molecular modeling and computational chemistry _____	110
6.6.1	Preparation of X-ray structures _____	110
6.6.2	General setup for docking experiments _____	110
7	References _____	111
8	Appendix _____	116
8.1	Pulse programs _____	116
8.1.1	STD-HSQC _____	116
8.1.2	1D STD NMR spectra with separate <i>on</i> and <i>off resonance</i> FID's _____	118
8.1.3	1D STD NMR spectra with internal subtraction _____	124
8.2	Programs and scripts _____	127
8.3	Template files and configuration files _____	138

I. Abbreviations

1D, 2D, 3D	one, two and three dimensional
Ac	acetyl-
CDx	cluster of differentiation, x is the identification of the protein
CORCEMA	complete relaxation and conformational exchange matrix
COSY	correlation spectroscopy
CRD	carbohydrate recognition domain
DTT	dithiothreitol
ECM	extra cellular matrix
F1, F2	NMR domains
FID	free induction decay
<i>gg, gt</i>	<i>gauche-gauche, gauche-trans</i>
HMBC	hetero nuclear multi bond spectroscopy
HSQC	hetero nuclear single quantum spectroscopy
HTS	high throughput screening
ITC	isothermal titration calorimetry
K_D	dissociation constant
kDa	kilo dalton, 1000g/mol
K_i	inhibitory constant
k_{off} / k_{on}	dissociation rate / association rate
MD	molecular dynamics (simulation)
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect / enhancement
PBC	periodic boundary condition
PBS	phosphate buffered saline
Ph	phenyl-
RMSD	root mean square deviation
SAR	structure activity relation ship
S/N	signal to noise ratio
STD	saturation transfer difference
T1, T2	longitudinal and transversal relaxation time
<i>tg</i>	<i>trans-gauche</i>
TOCSY	total correlation spectroscopy
TPPI	time proportional phase increment
TRIS	tris(hydroxymethyl)aminomethane
trNOE	transferred nuclear Overhauser effect
T_{sat}	saturation time
UHTS	ultra high throughput screening
VDW	van der Waals
VS	virtual screening
WATERGATE	water suppression by gradient tailored excitation

II. Amino acids

Aminoacid	3-letter-code	1-letter-code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Asparagic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

III. Carbohydrate monomers

Abbreviation	Name
Fuc	Fucose
Gal / GalNAc	Galactose / 2-deoxy-2- <i>N</i> -acetyl-galactosamine
Glc / GlcNAc	Glucose / 2-deoxy-2- <i>N</i> -acetyl-glucosamine
Man / ManNAc	Mannose / 2-deoxy-2- <i>N</i> -acetyl-mannosamine
NeuNAc (Sia)	5-deoxy-5- <i>N</i> -acetyl-neuraminic acid (sialic acid)

IV. Carbohydrate oligomers

Abbreviation	Constituent monomers	Comment
DTG	Gal β 1-1 β Gal	di-thio- β -D-galactoside
Lac	Gal β 1-4Glc	Lactose
LacNAc	Gal β 1-4GlcNAc	Lactoseamine type II
LNB	Gal β 1-3GlcNAc	Lactoseamine type I, lacto- <i>N</i> -biose
LNT	Gal β 1-3GlcNAc β 1-3-Gal β 1-4Glc	Lacto- <i>N</i> -tetraose
SiaLac	NeuNAc α 2-3Gal β 1-4Glc	3'-Sialyllactose
SiaLacNAc	NeuNAc α 2-3Gal β 1-4GlcNAc	3'-Sialyllactosamine

V. Protein identification

Protein	Source	SWISSPROT	PDB	Protein name, Functional description
Galectin-1	human	P09382	1GZW	Galectin-1, Beta-galactoside-binding lectin L-14-I, Lactose-binding lectin 1, S-Lac lectin 1, Galaptin, 14 kDa lectin
Galectin-1	<i>bos Taurus</i>	P11116	1SLA,1SLB, 1SLC,1SLT	Galectin-1, Beta-galactoside-binding lectin L-14-I, Lactose-binding lectin 1, S-Lac lectin 1, Galaptin, 14 kDa lectin, HPL, HBL, MAPK activating protein MP12
Galectin-3	human	Q08380	1BY2, 1A3K	Galectin-3 binding protein precursor, Lectin galactoside-binding soluble 3 binding protein, Mac-2 binding protein, (Mac-2 BP, MAC2BP, Tumor-associated antigen 90K
Galectin-7	human	P47929	1BKZ, 2GAL, 3GAL, 4GAL, 5Gal	Galectin-7, Gal-7, HKL-14, PI7, p53-induced protein 1

1 Introduction

1.1 Drug discovery and development

The development of new compounds for medicinal application starts with the occurrence of a medicinal condition that requires treatment. These conditions may occur as a result of infections, inflammation, injuries or from ageing. The technological developments in the last one and a half centuries allowed the purposeful search for distinct new biologically active compounds. Prior to these developments medicinal active compounds resulted from purely accidental empirical data acquired over time¹.

Today's search for new drugs or new application for known drugs is based on the determination of the malignant event in the host organism at a molecular level. This usually results in several proteins and agents involved in these events. Identification of the molecular pathway of the event allows the selection of one or more possible points of application for a treatment, often referred to as target.

Once a suitable target has been identified the search for compounds that prevent malignant interactions is possible. This requires a test system that yields significant data. The significance of the test result is crucial since development of drugs is causing high costs. These costs would rise significantly if money is spent on development of a compound whose binding activity occurred as a result of a testing error (false positive). Late stage failure may also occur as a result of toxicity and poor pharmacokinetics².

After a testing system has been set up, compounds can be chosen for testing their biological activity against the target. Modern facilities screening for biological activities are capable of screening hundreds of thousands of compounds within days ((ultra) high throughput screening, (U)HTS)³. Progress in sensory equipment allows even further miniaturization and parallel testing so that millions of compounds can be screened within less than a week.

The sources of compounds initially screened are new chemical substances, compounds isolated from natural sources and compounds with known biological activity. From the initial compound library screened only a small fraction should exhibit biological activity. From these active compounds, also called hits, one molecule is chosen to be advanced by chemical modification. The choice of this molecule is based on the activity from the screening, the possibility for chemical modifications on the molecule and fulfillment of basic pharmacokinetic properties¹.

On the basis of the hit several molecules are synthesized to increase affinity and selectivity to the target. Additionally important parameters like solubility, toxicity, bioavailability and distribution of the compound in different compartments of the host organism are tested and optimized⁴.

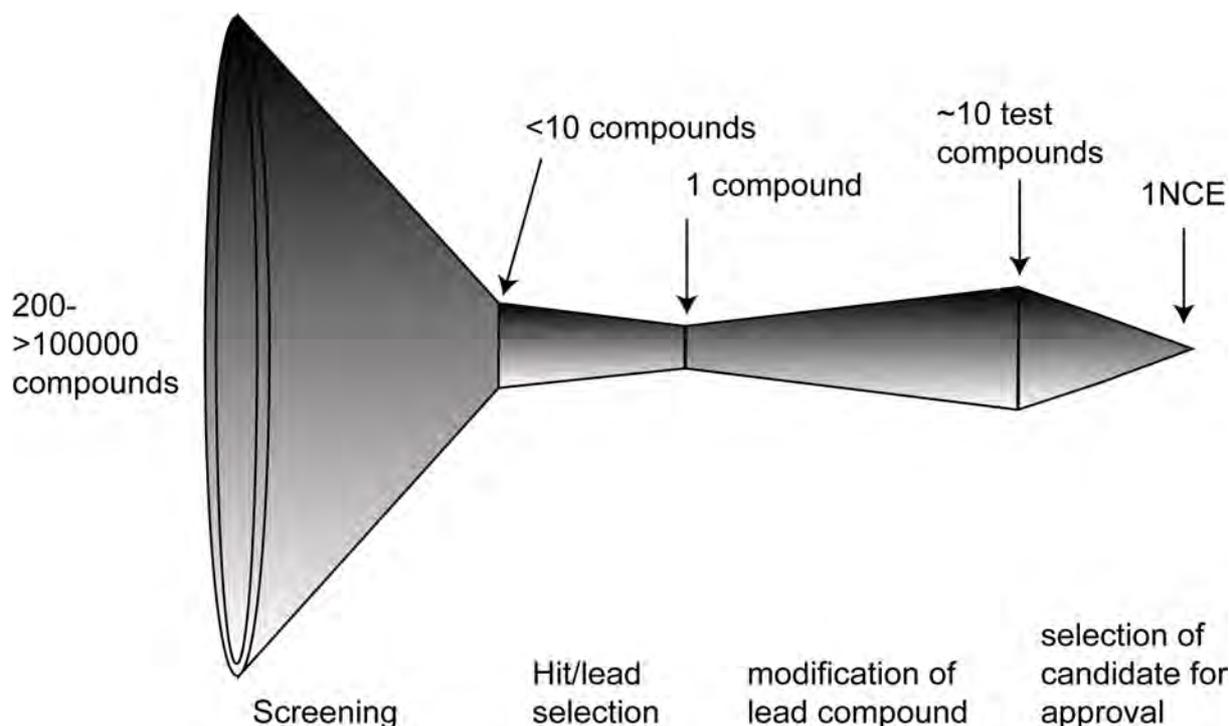


Figure 1.1: Representation of the drug development process: a large number of compounds are screened for activity, from the best hits a single compound is chosen for further development. From the resulting candidates one is developed to the final drug. The cumulative costs rise from left to right^{2,4,5}(NCE: new chemical entity)

The initial analytical method in the screening process is often based on the enzyme linked immunosorbent assay (ELISA). Further studies on the candidate compounds may require additional analytical methods to establish structure activity relationships (SAR) and quantitative SAR (QSAR). In addition to biological *in vitro* assays several physical testing methods exist to determine binding affinity of compounds to a target *in vitro* qualitatively and quantitatively. Examples of *in vitro* testing methods are isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and affinophoresis. Recently developed technologies also include various nuclear magnetic resonance (NMR) spectroscopy techniques for the characterization of binding events.

A valuable tool in the development of new compounds with biological activity is knowledge of the three dimensional structure of the complex of compound and target protein. The 3D structure of a protein can be solved using X-ray crystallography. This method does not impose limitations on the size of the protein studied but the proteins usually need to be soluble in water to yield crystals that can be investigated by X-ray methods. Many complexes of pharmacological interesting proteins with their natural ligands have been solved by X-ray diffraction. This is the basis for structure based design. Visualization of the complex allows the design of new compounds binding to the protein on the basis of the molecular interactions in the complex. Crystallization of proteins and protein-ligand complexes is not applicable to all targets or ligands. Especially integral membrane and trans-membrane proteins are not amenable to crystallization based methods.

NMR is capable to determine the structure of proteins in solution. This methodology is constrained to proteins of a certain size (<30 kD). The structure of a bound ligand can also be determined by NMR methods⁶.

Although modern screening technologies allow the screening of an increasing number of compounds the number of compounds approved for therapeutic application is retrogressive⁷. It has been proposed that this is a consequence of mismanagement and failure in the drug development process⁸. Furthermore, drug candidates submitted for approval may fail to meet admittance criteria.

1.2 Molecular modeling

Many areas of the drug discovery and development process can be supported by computer technology. This ranges from databases with information about compound properties over visualization of molecules and complexes for prediction of compound properties and generation of molecular structures and molecular complexes. These methods require considerable high computer performance.

In particular, the creation of complexes of ligands and a target protein are of interest. They allow the evaluation of ligand interactions and indicate possibilities for chemical modification without destroying present interactions. Calculated complexes of ligand and protein based on structural data on the protein are often referred to as pseudo crystals. The main method of obtaining pseudo crystals is placement (docking) of the ligand into the binding site of the protein. This docking can either be done manually or in an automated fashion. Docking is usually followed by scoring the ligand interaction with the protein. This allows a virtual screening of compounds, where several modifications of a starting ligand can be tested without the need of synthesis or biochemical assay. This can yield valuable information concerning the future modifications during the development process. Validation of predictions by *in vitro* tests is desirable. If the binding site of the protein is not known from crystal structure or NMR experiments it can be predicted⁹ and characterized¹⁰ from the 3D structure by computer methodologies.

Homology modeling allows the generation of 3D models of proteins where no X-ray structure is available. This methodology requires the 3D structure of a protein with a similar amino acid sequence as the target protein as a template. The higher the homology of target and template protein the more reliable the model generated becomes¹¹.

Computer programs can also be applied in the prediction of physicochemical and pharmacokinetic properties of compounds during the drug design process^{2;12}.

1.2.1 Docking

Screening molecules for their biological activity in modern pharmaceutical research requires considerable resources. High throughput screening methodologies allow the screening of more than 100000 compounds at a time. Preparation of testing

assays, storage of vast molecule repositories of sometimes over one million compounds are factors raising the operating expenses of pharmaceutical industry.

The process of docking aims at the creation of structural ensembles of ligand and receptor molecules also addressed as 'pseudo crystal structures'. Docking was designed to allow studies of binding processes without carrying out biochemical screening experiments. The basic idea is placing molecules of interest into predefined binding sites of proteins. This implies that the three dimensional structure of the protein of interest is known.

Docking programs offer the possibility to save several structural proposals from the docking procedure. These are usually the proposals that achieved the highest score with the respective scoring function. A single, specific proposal with its interactions, orientation and conformation is referred to as pose.

The first docking techniques relied on simple shape fitting of the small molecule into the receptor's binding site, ignoring ligand flexibility and allowing only rigid body movements^{13;14}.

While shape complementarities remain as basic feature in the search of new compounds as inhibitors, modern docking techniques employ more sophisticated methods today. This is feasible due to the rapid development in the performance of modern computers.

Further improvement of the performance of docking algorithms arose from the implementation of grids. A grid essentially is a lookup table storing data on the physical and chemical characteristics of the receptor at certain points. The first automated docking programs implementing grids were AutoDock¹⁵, based on the GRID program of Goodford¹⁶ and DOCK¹⁷.

DOCK is differing from most docking procedures in its approach to represent the receptor surface by a negative imprint of the groove or binding site using spheres¹⁸. The spheres describe potential interaction sites and can be used for matching ligand atoms while docking¹⁹.

Since the conformation of a bound ligand in the binding site usually is not known, treatment of ligand flexibility is a key factor in the docking procedure employed. Various methods have been employed in order to generate multiple conformers of the ligand, prior to docking or while being docked¹⁹. An additional approach to flexibility implemented in DOCK is an incremental construction algorithm^{20;21}. This technique is based on a fragment of the ligand, which is defined as anchor. The anchor fragment is placed in the binding site and the pose is optimized. Extensions to the fragment are attached incrementally and optimized individually. A similar approach is implemented in FlexX, where chemical interactions are screened rather than steric complementarities²². The potential inhibitor molecule is first docked into the predefined binding site in various orientations. For each of the general orientations of the molecule different conformations can be created. Alternatively

conformations can be created prior to the docking. Finally each of these structural proposals can be minimized.

Subsequent to the docking procedure is the evaluation of the structural proposal. This is often referred to as scoring since the score allows comparison of different binding modes or molecules. As mentioned above the first applications of docking methodologies relied on the satisfaction of spatial demands in the binding site of interest. With the broad availability of powerful computers more complex functions for the evaluations of intermolecular interactions occurring between ligand and receptor are accessible. The main contribution to the score usually comes from force field calculations of the non-bonded interactions in the complex. The minimum setup for the force field includes a term for van der Waals interactions and a term for electrostatic interactions, represented by the Coulomb potential (Equation 1.1). The force fields applied in docking programs usually include AMBER^{23,24}, OPLS-AA²⁵ and CHARMM²⁶.

$$\Delta E_{bind} = \sum_{i=1}^{Lig} \sum_{j=1}^{Rec} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} \right) + \frac{q_i q_j}{D r_{ij}}$$

Equation 1.1

The pair wise calculation of the atomic interactions is neglected at great distances. The cutoff for energy calculation depends on the setup and system studied but is generally set between 8 and 12 Å. As well as the selection of the parameters of the force field the determination of the dielectric constant D in the Coulomb term may be crucial. The dielectric constants determined for bulk compounds or liquids may not apply since usually only single molecules of water or ions from buffer/blood are between the surfaces of ligand and receptor or in their vicinity. Also setting D to 1 as for vacuum is not valid, since partial charges on the surface of ligand and receptor create an irregular pattern of polar interactions. Especially in cases of charge interactions between ligand and receptor (salt bridges) the choice of D proves crucial^{19,27}. Some scoring functions include a modified Lennard-Jones-potential accounting for hydrogen bonds. This proves useful if the representation of partial charges does not cover for these interactions.

The force field function does not account for entropy effects based on change of solvation and molecular flexibility upon binding^{19,28}. These effects can be taken into account by free energy perturbation or thermodynamic integration methods, but this is very time consuming. Using this approach binding free energy is calculated with errors within 1 kcal/mol²⁹.

The validity of a docking procedure is usually assessed by a benchmarking procedure. The main part of the benchmark is the recovery of the binding mode found in the crystal structure by binding mode prediction. Currently test sets exceeding 100 protein ligand structures are used to assess docking algorithm capabilities³⁰⁻³².

The capabilities of docking procedures can be employed in the drug development process. One application is screening virtual databases for active compounds. For successful fulfillment of this task the docking procedure must be able to identify active compounds from a set of chemically and structurally diverse compounds. Ideally the number of active compounds should be small compared to the contents of the database. The key benchmark in virtual database screening is the recovery of known active compounds from the library (hit rate). Higher hit rates result in less compounds to be screened experimentally¹⁹.

Once active compounds have been found by virtual and/or experimental screening, docking can be employed in the lead optimization process. Lead optimization or directed library design accounts for compounds of similar chemical structure but different activities towards the receptor. For this task the scoring function must be able to correctly rank chemically and structurally similar compounds¹⁹. The ability of a docking program to rank these compounds correctly can be tested if experimental binding data is available for some of the compounds in the test set. The availability of X-ray data of at least one compound of the library, bound to the receptor, is desirable. Based on these data and docking experiments the capabilities of the docking program can be assessed and the setup tuned to yield better results³³⁻³⁵.

Although docking programs usually perform satisfactory when screening for active compounds from chemically diverse libraries, they usually perform less accurate when ranking chemically similar compounds³⁶. The accuracy of the docking protocol depends on several factors, such as number of rotatable bonds, selection of base structure for the receptor if several are available, the molecular weight of the ligand docked and the choice of the docking program itself²⁷.

A common drawback of docking programs is the tendency to benchmark them versus test sets of pharmaceutically interesting receptors and drug like ligands. This poses some caveats when trying to predict the binding mode of highly flexible structures like oligosaccharides or oligopeptides. These ligands are of interest if the binding mode of endogenous ligands is unknown from other experimental procedures. The binding mode of the endogenous ligand to pharmaceutically interesting protein is valuable information in the drug design process.

1.3 NMR screening

The key element in screening methodologies is their reliability and versatility. Over the years several NMR based methods have been developed and established in the structure based drug design process. The versatility of NMR is based on its capability to yield results for molecules in solution or in solid state. NMR screening technologies not only provide information about the capability of a molecule to bind to a certain target protein even from mixtures, but also yield insight into structural features of these interactions. Suitable NMR experiments can be used to identify parts of the ligand in contact with the receptor or the structure in the bound state.

Several NMR methods also allow the determination of dissociation constants for the complex.

NMR methods also can be employed to determine parts of the protein in contact with the ligand, thus a binding site can be determined³⁷⁻³⁹. Upon binding the chemical environment of a nucleus close to the binding site is altered. Using specific isotope labeling in proteins ^{15}N - ^1H or ^{13}C - ^1H correlation spectra can be employed to detect parts of the receptor in contact with the ligand (structure activity relationships by NMR, SAR by NMR). Comparisons of spectra with and without ligand allow the detection of chemical shifts occurring. If the off-rate (k_{off}) of the complex is fast compared to the chemical shift timescale, correlation of the chemical shift change and ligand concentration provides a measure of the dissociation constant K_D ⁴⁰.

The transient transferred nuclear Overhauser effect (trNOE) can be used to determine binding activity of single ligands and ligand libraries in two ways. The intramolecular trNOE yields information about the conformation of the molecule in the bound state. The intermolecular trNOE is applied to determine the orientation of the ligand in the binding pocket. Size and sign of the NOE are depending on correlation time and spectrometer frequency. Correlation time itself is a measure for the molecular mobility. Small molecules bear a small correlation time and a positive NOE where the intensity is at a maximum in the range of up to several seconds mixing time. Large molecules like proteins have a lower correlation time and their NOE is of a negative sign. It builds up in a few hundred milliseconds. Molecules that interact with a macromolecule are subject to the low correlation time, their trNOE is negative. After dissociation the information is transported into solution and is detected. Thus a binding molecule in a compound library should bear a negative NOE in contrast to non binding molecules⁴¹.

By application of pulsed field gradients diffusion edited NMR experiments are capable of detecting the affinity of compounds to macromolecules by recording their altered diffusion behavior upon binding⁴².

The NOE-Pumping and reversed NOE-pumping experiments are based on the change of the relaxation behavior for protein and ligand upon binding. In the original NOE-pumping experiment the NOE is transferred from protein to ligand, then magnetization of the ligand is destroyed via a diffusion filter, followed by a NOE like mixing time. During this mixing time the ligand is a possible pathway of protein relaxation, causing ligand resonances to appear in correlation with mixing time. The reverse NOE pumping experiment employs a T_2 relaxation filter to destroy magnetization of the protein. In the adjacent NOE mixing time the protein serves as a relaxation partner for the bound ligand, causing an attenuation of ligand resonances compared to a reference spectrum where NOE mixing time and T_2 filter are interchanged^{43,44}.

Also based on the alteration of the relaxation behavior is the detection of binding events via broadening of ligand resonances. Due to additional relaxation via

interaction with the protein the ligands T_1 and T_2 relaxation times are reduced. This causes broadening of the ligands resonances. The broadening is detectable when comparing spectra of the ligand with and without protein⁴⁵.

The relaxation behavior may also be altered by introduction of spin labels to the protein. The spin label carries a radical electron. Unpaired electrons expose a massive gyromagnetic ratio which causes exceptionally small relaxation times⁴⁶.

Water molecules are often part of the interface in protein ligand complexes⁴⁷. Over their residence time of several hundred microseconds a negative NOE builds up. Free floating molecules in the solvent have positive NOEs. In the WaterLOGSY (Water-Ligand Observed via Gradient Spectroscopy) experiment a spectrum is recorded where irradiation is applied off resonance. This spectrum is subtracted from a spectrum with irradiation at the resonance of water (*on resonance*). In the difference spectrum binding compounds have positive signal while non binding compounds have negative signals⁴⁸.

The choice of the NMR screening experiment for a given target is depending on the kind of information that is to be retrieved and some knowledge about the system studied. Like any other screening method NMR screening methods are subject to certain limits in their environmental parameters. Especially the required amount of protein and the demand for selective or uniform isotope labeling in the protein may pose a hurdle upon availability of sufficient amounts of protein. The nature of the binding between ligand and protein may also cause problems with certain NMR methods, when the binding is too strong or weak for the method employed. Table 1.1 summarizes the constraints and features of the NMR screening methods here³⁸.

1.4 Saturation transfer difference nuclear magnetic resonance (STD NMR)

The saturation transfer difference NMR experiment allows qualitative and quantitative analysis of binding events between macromolecular receptors (e.g. proteins) and small molecule ligands^{49;50}.

Proteins exhibit noticeable line broadening in ^1H -NMR spectra. This leads to resonances in the high and low field region of the NMR spectrum (e.g. <0 and >10 ppm). Molecules with a molecular weight (MW) of more than 10 kD are subject to spin diffusion so that irradiation of selective resonances in these molecules rapidly spreads over the whole spectral width. This means, that the whole protein spectral width can be saturated by irradiation outside the spectral width of small molecules with sharp spectral lines.

While in contact with the protein a ligand is subject to the same NMR properties as the protein as a result of the slow tumbling of the complex. Saturation applied to the protein spreads to the ligand via dipolar interactions. This transfer of magnetization leads to attenuation (saturation) of signals from the ligand if a 90° high power pulse is applied after a certain saturation time and dissociation of the complex. This is exploited in the saturation transfer difference (STD) NMR

experiment, where the ligand spectrum resulting from saturation of the protein is deducted from a spectrum where no saturation was applied to the protein. In order to assure parity in experiment length and thermal conditions in the two spectra the second spectrum is acquired with a saturation using a selective pulse outside the spectral window of the protein (e.g. 20 ppm or more). The spectrum where no protein resonances are saturated is referred to as 'off resonance' and should equate a normal NMR spectrum without any saturation pulse. The spectrum where protein resonances are saturated is referred to as 'on resonance'. The resulting difference spectrum is referred to as 'STD spectrum'.

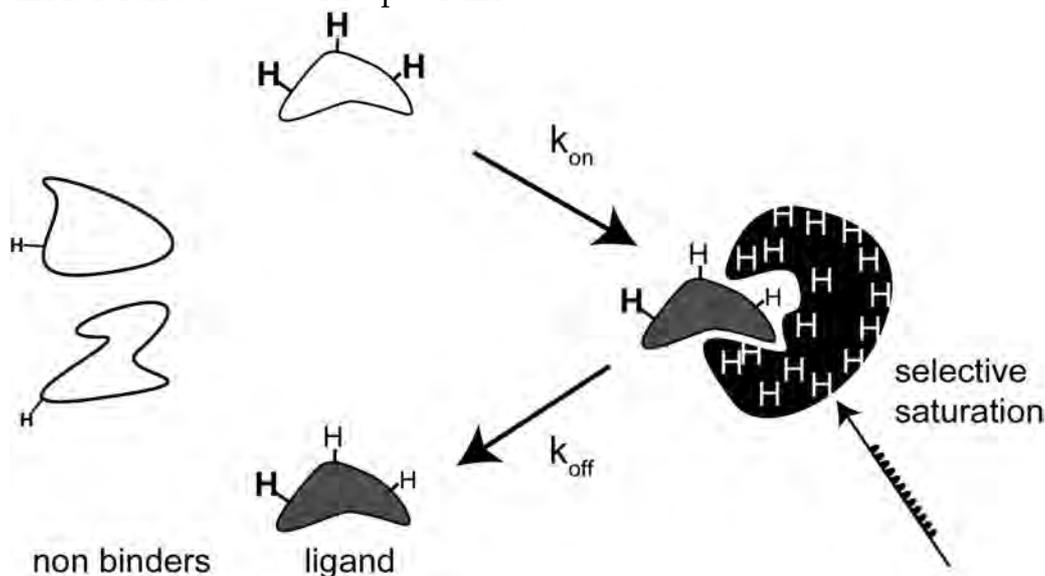


Figure 1.2: Scheme of STD NMR on molecular basis: ligands in contact to the protein are subject to magnetization transfer, causing attenuation of their signal intensity (depicted by smaller H's). Intensity of protons away from the protein or from non binding molecules does not experience saturation of their resonance signal.

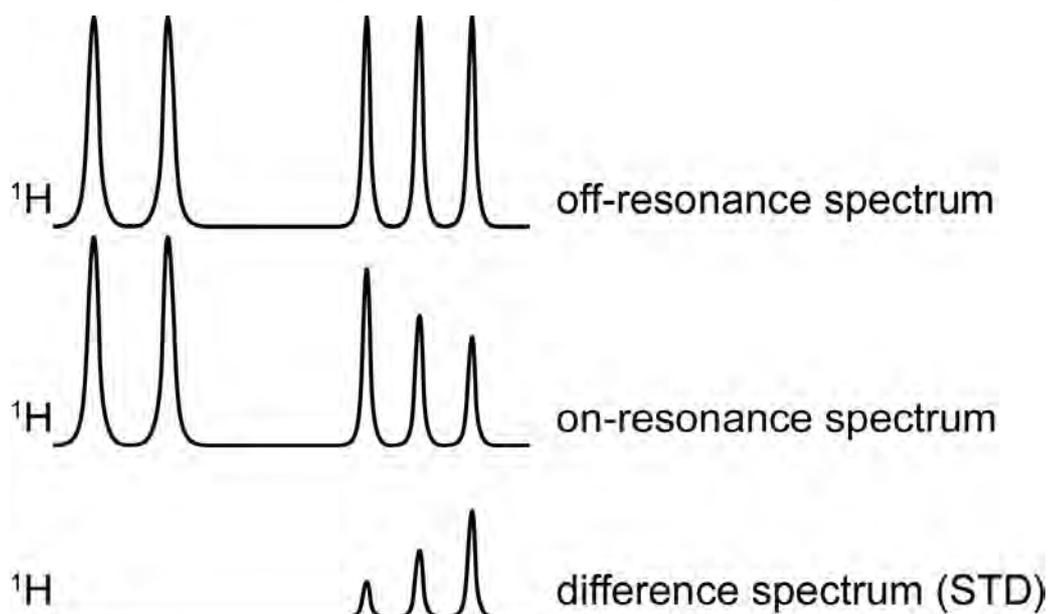


Figure 1.3: Schematic representation of off (top) and *on resonance* (middle) spectra and the resulting difference spectrum (bottom).

Atoms (i.e. protons) close to the protein should render a signal in the STD spectrum, while atoms or groups of atoms far away from the protein are not observable in the STD spectrum. Their intensity should not be changed in the *off resonance* spectrum when no saturation is transferred.

Due to the fact that saturation is transferred via dipolar interactions ligand protons in close contact with one or more protons in the protein should be subject to more saturation transfer. This generates signals of less intensity in the *on resonance* spectrum and observable signals in the STD spectrum.

This leads to generation of a binding epitope (i.e. pharmacophore) of the ligand at atom level resolution. In order to obtain a highly sensitive mapping of the binding epitope (group epitope mapping, GEM) a high excess of the ligand is desirable since the signal to noise ratio in the STD spectrum is enhanced⁵¹.

The basic STD pulse sequence contains a cascade of selective low power Gauss pulses for the saturation of the protein. A $T_{1\rho}$ filter can be applied to attenuate protein resonances, the HDO signal can be suppressed using a WATERGATE pulse sequence. 1D STD NMR spectra of complex or large molecules are difficult to interpret due to overlap of the resonances. 2D STD NMR like the STD TOCSY and the STD HSQC spectra can be used in these cases, but they are subject to less sensitivity and increased demand of experimental time.

To achieve a STD spectrum the binding event under surveillance needs to be fast in respect to the timescale of NMR experiment. If the T_2 relaxation time is short in comparison to the residence time of the ligand in the binding site uniform spread of saturation over the whole ligand molecule could occur.

The size of the STD effect is depending on the kinetic parameters of the complex (Equation 1.2).

$$K_D = \frac{[P] * [L]}{[PL]} = \frac{k_{off}}{k_{on}}$$

Equation 1.2

Low off-rates (k_{off}) denote a long residence time of ligand molecules in the binding pocket and low exchange of ligands in the solution. This causes a low turnover and low intensity in the difference spectrum. Huge on- and off-rates do not allow the ligand to stay in contact with the protein long enough for an effective transfer of saturation. Assuming a purely diffusion controlled mechanism for the association reaction forming the complex, k_{on} would be $10^7 \text{ s}^{-1}\text{M}^{-1}$. If K_D is known k_{off} can be estimated, it should exceed 1s^{-1} . The mean residence time t_r in the binding pocket is

$$t_r = \frac{\ln 2}{k_{off}}$$

Equation 1.3

STD NMR can be used to determine the K_D value of a ligand. This is done by comparison of the STD effects of one or more resonances from the ligand at various ligand concentrations versus a constant protein concentration.

$$STD = \frac{I_0 - I_{sat}}{I_0}$$

Equation 1.4

The STD effect ranges from 0 to 1, multiplication by 100 yields the absolute STD percentage. To compare different STD titration experiments the STD amplification factor is calculated.

$$STD - Amplification = \frac{I_0 - I_{sat}}{I_0} \cdot \frac{[L]_{total}}{[P]_{total}} = STD \cdot ligandexcess$$

Equation 1.5

In case of a specific binding event plotting the STD amplification factor versus the ligand concentration yields a curve that can be fitted after the one site binding model.

$$STD - Amplification = \frac{STD - Amplification_{max} \cdot c}{K_D + c}$$

Equation 1.6

STD NMR can be applied to any protein exceeding a mass of 10 kD and that provides a stable structure. The proteins may be soluble in the buffer system used but they may also be immobilized⁵⁰. Insoluble proteins like integral membrane proteins can be embedded into liposomes⁵² or their binding properties can be studied with cells expressing these proteins on their surface using intact cells and the saturation transfer double difference experiment⁵³.

STD NMR experiments offer high versatility in their requirements regarding protein amount and range of K_D . Apart from the lower limit of 10 kD of the protein to ensure spin diffusion there is no upper limit of the protein studied. Due to the nature of the experiment no information is gained on the epitope of the protein. STD NMR requires a minimum amount of 0.1 nmol unlabeled protein at a spectrometer frequency of 500 MHz. STD NMR offers ligand identification from compound libraries and information on the binding epitope. The binding affinity may be in the range of $K_D=100$ pM to $K_D=10$ mM³⁸.

Table 1.1: Features and limits of common NMR screening methods. Reverse NOE pumping (RNP) is usually not used for ligand epitope mapping, the epitope from Water-LOGSY represents the water contact surface. The amount of protein required refers to 500 MHz spectrometer frequency.

	SAR by NMR	Spin labeling	Diffusion editing	RNP	Water-LOGSY	STD NMR
Protein >30 kD	with TROSY	yes	yes	yes	yes	yes
Protein <10 kD	yes	yes	no	no	no	no
Protein isotope label	yes	no, but ligand spin label	no	no	no	no
Protein epitope	yes	no	no	no	no	no
Ligand epitope	no	no	no	yes	yes	yes
Required protein[nmol]	25	~1	100	~25	~25	0.1
K_D tight binding	no limit	100 pM	100 nM	1 nM	100 pM	100 pM
K_D weak binding	~1 mM	10 mM	1 mM	1 mM	10 mM	10 mM
Library screening	yes, but ligand is not identified	yes	yes	yes	yes	yes

1.4.1 Structural information using Overhauser effects and selective labeling on performing STD experiments: SOS-NMR

P.J. Hajduk *et al.* developed a method to create pseudo crystal structures from STD NMR data and docking proposals from DOCK⁵⁴. Distinct amino acids were selectively labeled with protons in an otherwise completely deuterated protein. Since for a specific protein the protonated amino acid is known, STD effect on the protons of the ligand could be assigned to a close contact to the respective amino acid in the previously determined binding pocket. Experiments with various amino acids labeled with protons allowed the determination of the orientation of the ligand in the binding site. The intensity of the SOS NMR signals was converted into distance restraints in order to select matching poses from an ensemble of structures proposed by DOCK. While this methodology yields information about the ligand epitope and the amino acids causing the respective effect it does have some drawbacks. First the expression of several selectively labeled proteins as well as the required perdeuterated protein is cost intensive. Second the lack of protons in the proteins reduces spin diffusion significantly, so that irradiation at several resonance frequencies was required. Hajduk *et al.* claimed that the STD spectra of the ligand, 2-(3'-pyridyl)-benzimidazole bound to the unlabeled FK506 binding protein yielded almost uniform intensities for all ligand protons. This may point to either unselective binding events or disadvantageous binding kinetics. Additionally, the proposed binding mode of the ligand in the binding site results in a molecule residing well inside the binding site with all protons facing the surrounding amino acids, which is a good explanation for uniform intensities, as the distances to the surface are similar.

1.4.2 Complete relaxation and conformational exchange matrix (CORCEMA) analysis of STD effects

V. Jayalakkshmi and N.R. Krishna adopted the CORCEMA analysis to obtain quantitative STD effects from STD NMR experiments incorporating relaxation processes and kinetic influences of the system surveyed⁵⁵. This procedure considers the influence of the complex and the ligand conformation on the relaxation rates of the ligand protons on the STD effects determined for these protons. The CORCEMA-ST protocol allows the prediction of STD effects for a given receptor ligand complex, if binding constant and correlation times are known. The CORCEMA method has been applied to the structural refinement of ligands bound to receptors determined by X-ray crystallography and structurally similar ligands⁵⁶⁻⁵⁸. For the construction of an accurate relaxation matrix the CORCEMA-ST protocol requires a high resolution structure of the protein.

1.5 Galectins

Galectins, also known as S-type or S-Lac lectins, are a family of metal independent β -galactoside specific lectins. They share a common amino acid sequence, especially in their carbohydrate recognition domain (CRD). Up to now 14 galectins have been identified in mammals and humans. Many more β -galactose specific lectins occur in plant, bacteria and vertebrates and invertebrates. Within one mammal species different galectins have a sequence identity from 20 to 40%, specific galectins (i.e. galectin-1) have an inter species (mammal) sequence identity from 80 to 90%. Galectins are abundant in the cytosol and extra cellular matrix.

Galectins are subdivided into 3 classes of galectins; prototype, chimera type and tandem repeat galectins⁵⁹⁻⁶¹. Prototype galectins consist of a single CRD and may occur in dimeric form depending on circumstances (galectin-1, -2, -5, -7, 10, -13, -14). Galectin-3 is the only known member of the chimera type. It has a single CRD, a N-terminal domain and an intervening glycine, proline and tyrosine rich domain, the latter consisting of repeats of 7 to 10 amino acids with the consensus sequence PGAYPG(X)₄ (X: any amino acid). The N-terminal domain of galectin-3 is often referred to as oligomerization domain. Like all galectins it lacks a signal peptide sequence. Tandem repeat galectins present two different CRDs linked by a repeating domain homologous to the repeat domain of galectin-3. The CRDs in one tandem galectin are different to each other and CRDs of other galectins (galectin-4, -6, -8, -9, -12)^{59;61;62}.

```

Query: 1 ACGLVASNLNLKPGECLRVRGEVAADAKSFLLNLGKDDNNLCLHFNPRFNAHGDVNTIVC
      ACGLVASNLNLKPGECLRVRGEVA DAKSF+LNLGKD NNLCLHFNPRFNAHGD NTIVC
Sbjct: 1 ACGLVASNLNLKPGECLRVRGEVAPDAKSFVLLNLGKDSNNLCLHFNPRFNAHGDANTIVC
Query: 61 NSKDAGAWGAEQRESAFPFPQPGSVVEVCISFNQTDLTIKLPDGYEFKFPNRLNLEAINYL
      NSKD GAWG EQRE+ FPFQPGSV EVC I+F+Q +LT+KLPDGYEFKFPNRLNLEAINY+
Sbjct: 61 NSKDGGAWGTEQREAVFPFPQPGSVAEVCITFDQANLTVKLPDGYEFKFPNRLNLEAINYM
Query:121 SAGGDFKIKCVAFE
      +A GDFKIKCVAF+
Sbjct:121 AADGDFKIKCVAFD

```

Figure1.4: Sequence alignment of galectin-1 from *bos taurus* (top, query)) and *homo sapiens* (bottom, subject). The sequence is identical to 86% and is homologue to 93%. (Alignment by BLAST-P on www.ebi.ac.uk)

Galectin expression is activated in particular developmental or physiological stages. Galectin-1 is abundant in skeletal, smooth and cardiac muscle cells, motor and sensory neurons, thymus, kidney and placenta. Galectin-1 has been shown to play a role in apoptosis in the context of cancer^{63,64} and HIV infected T-cells⁶⁵. Recently a monomeric form of galectin-1 (galectin-1 β) was described. It lacks 6 amino acids in the N-terminus and was shown to promote axonal regeneration⁶⁶. Galectin-3 is expressed in activated macrophages, basophils, mast cells, epithelial cells of the intestine, the kidney and some sensory neurons.

Galectin-1 and -3 have been shown to play a key role in pre-mRNA splicing, where each of the proteins can substitute deficiency of the other lectin, shown by specific CRD directed antibodies. Unspecific (in respect of the lectin) saturation of both CRDs with thio-digalactoside (TDG) led to loss of splicing activity in the spliceosome^{67,68}.

Expression and *in vivo* function of the remaining galectins is less understood.

Most galectins are expressed over a non classical (also termed non conventional) pathway. They are synthesized on free ribosomes in the cytoplasm and not packaged into vesicles prior to export. Inhibition of ER/Golgi-mediated protein synthesis does not affect expression of galectin-1 and -3. Galectin-1 and -3 have been shown to accumulate directly under the plasma membrane. Export seems to be achieved by formation of an exosome, a membrane bound vesicle. While reasons for accumulation and excretion remain unknown there is evidence, that targeting to the plasma membrane is the rate limiting step in their secretion⁶⁹.

Galectins have been shown to play vital parts of tissue response to inflammation, in cancer, trafficking and cell adhesion⁷⁰. Especially their feature of forming multivalent complexes makes them major facilitators of cell and tissue aggregation. It

has been shown that homodimeric galectin-1 can induce the formation of protein clusters on cell surfaces⁷¹.

Galectins share their preference for β -galactosides. Especially galactose derivatives like lactose (Lac) and *N*-acetyl-lactosamine (LacNAc) exhibit strong interactions with galectins. Glycoconjugates containing LacNAc residues are assumed to be natural ligands of galectins. Namely the *N*-type glycoproteins laminin, fibronectin⁷², CD43 and CD45 are endogenous ligands of galectin-1^{61;62;73}. Though many putative and definite ligands and interaction partners for galectins are known the distinct function of galectins in healthy or malign tissues as well as in inflammation remain unsolved.

2 Scope of the thesis

Automated docking procedures are a rich source of information on the binding process of potential medical agents to their respective receptors. Although the advance in the field of computer hardware technology and algorithms has provided a continuous source of improvement of the predictions, the quality of the results is fundamentally enhanced by access to experimental data for benchmarking and comparison.

NMR based screening procedures allow insight into the binding event at an atomic level in the liquid state. Careful choice of the NMR experiments allows the determination of biologically active compounds from compound libraries and the determination of groups involved in the binding event on behalf of the ligand and receptor.

In this study the results of STD NMR experiments provide experimental data to improve the quality of docking results. The docking will be performed using the software DOCK. The binding of *N*-acetyl lactosamine and derivatives thereof to bovine galectin-1 is chosen as a testing system.

The intensity of a given signal in a STD NMR spectrum is proportional to the proximity of the corresponding ligand proton to protons of the receptor binding site. Comparison of distances of ligand protons to protons of the receptor in structural ensembles created by DOCK with the results from STD NMR will allow evaluation of docking quality. Data on docking quality will be employed to improve the docking setup and selection of binding modes from structurally diverse docking results.

Once a method has been established, analysis of extensive data will have to be implemented.

Comparison of docking results to binding modes determined by X-ray crystallography allows the verification of the developed procedure.

3 Results and discussion

3.1 Choice of a model system

Galectin-1 was chosen as a model system in this study, because it is well studied. Various galectins have been crystallized with or without galactosides bound to their CRD⁷⁴⁻⁷⁹. Galectin-1 from bovine spleen is commercially available (Sigma). The protein as well as the ligands is highly soluble in water. This allows preparation of NMR samples in buffer solutions close to the ionic content of blood or the extra cellular matrix (ECM). Unmodified carbohydrates usually do not form aggregates such that the application of additives usually is needless. Since galectin-1 presents free thiol groups from six cysteine residues in its sequence, storage in a dithiothreitol (DTT) containing medium is required. A wide variety of β -galactoside based ligands for galectin-1 is smoothly covered by broad availability of β -galactose and *N*-acetyl-lactosamine based chemicals.

3.2 Saturation of galectin-1

The STD NMR experiment requires that the selective irradiation pulses saturate the resonances of the protein while they do not saturate the resonances of any ligands. M. Mayer *et. al.* were applying saturation pulses in the range of 0 ppm to -1 ppm or in the region of the resonances of aromatic protons⁴⁹. The latter option is not applicable if the ligand itself incorporates aromatic protons. Carbohydrates usually display resonances from 3 ppm to 5 ppm, deoxy-carbohydrates additionally display resonances from 2 ppm to 3 ppm and \sim 1 ppm in case of methyl groups. *N*-acetyl groups in carbohydrates display resonances between 1.9 ppm and 2.0 ppm. Ligands assayed in this study incorporated *N*-acetyl groups or aromatic residues. It was thus assumed that irradiation in the range from 0 ppm to -1 ppm should not interfere with the resonances of the respective ligands.

To assess whether magnetization is transferred to the protein and spreads over all protons STD spectra of the protein without ligand were recorded. In these spectra protein resonances were not suppressed. The reference *off resonance* spectrum of the galectin-1 sample contained sharp lines assigned to components of the buffer solution. None of these signals appeared in the STD spectra. This implicated that the impurities did not interact with galectin-1. The results indicated that galectin-1 resonances could be adequately saturated to transfer magnetization to a ligand in contact with the protein. Thus a main prerequisite for STD NMR experiments was met. For the ligands used in this study occurrence of artifacts when applying irradiation in the range from -1 ppm to 0 ppm was tested individually.

In this study the magnitude of the STD effect was determined in two ways. The first method compares the signal height of a resonance signal in a reference spectrum (*off resonance*) and difference spectrum. STD effects determined in this fashion are labeled with 'intensity'. STD spectra recorded with the pulse programs as shown in

chapter 8.1.2 allow the comparison of the difference spectrum with the actual off resonance spectrum of the same experiment (no internal subtraction as in chapter 8.1.3). STD NMR experiments using these pulse programs were also analyzed by integration of the region where resonances occurred. The same integration ranges were used in the *off resonance* and difference spectra. STD effects determined in this fashion are labeled 'integral'.

Table 3.1: Determination of the saturation transferred to galectin-1 in different shift regions at 0 ppm and -1 ppm at 500 MHz spectrometer frequency. Saturation transfer was determined by integration of the respective regions and intensity of the signal humps.

irradiation [ppm]	Integral	low field border	high field border	integral off res.	integral STD	sat. [%] (integr.)	sat. [%] (intens.)
0	1	7.498	6.004	0.707782	0.30835	43.6	48
	2	3.248	2.600	0.608538	0.22923	37.7	38
	3	1.800	1.301	0.822669	0.3501	42.6	45
	4	1.100	0.501	1	0.50818	50.8	49
-1	1	7.498	6.004	0.495841	0.090229	18.2	30
	2	3.248	2.600	0.333314	0	0	15
	3	1.800	1.301	0.241286	0.110631	45.9	45
	4	1.100	0.501	1	0.191692	19.2	20

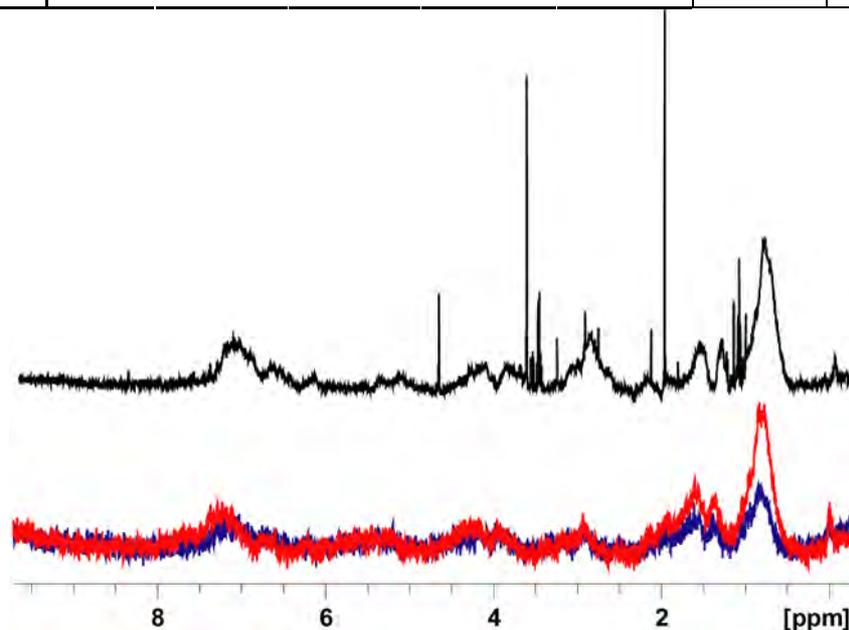


Figure 3.1: Saturation of bovine galectin-1 resonances was detectable at irradiation at 0 ppm (red) and -1 ppm (blue; black: *off resonance* spectrum). It was obvious that irradiation at frequencies further away from protein resonances decreased the efficiency of saturation. All spectra were acquired at 500 MHz spectrometer frequency, 300 K, pulse attenuation of 45 dB, 2 s saturation time, 1024 scans *on* and *off resonance* and receptor concentration of 18 μ M.

3.3 Epitope mapping of disaccharides

3.3.1 Methyl- β -D-lactoside

Lactose (Gal β (1-4)Glc) is a ligand for galectins. Compared to *N*-acetyl lactosamine type I or type II (1-3 and 1-4 linkage) lactose is reported to have a four fold weaker

affinity to galectin-1⁸⁰⁻⁸². β -D-O-methyl lactose ($\text{Gal}\beta(1-4)\text{Glc}\beta 1\text{-OMe}$, LacOMe) is commercially available. The fixed anomer allows unambiguous assignment of resonances in the 1D proton NMR spectrum. During acquisition of spectra for titration experiments it was noticed that the STD spectra were not analyzable due to phase errors that could not be corrected. These errors occurred with remarkable dominance in spectra with a spinlock pulse.

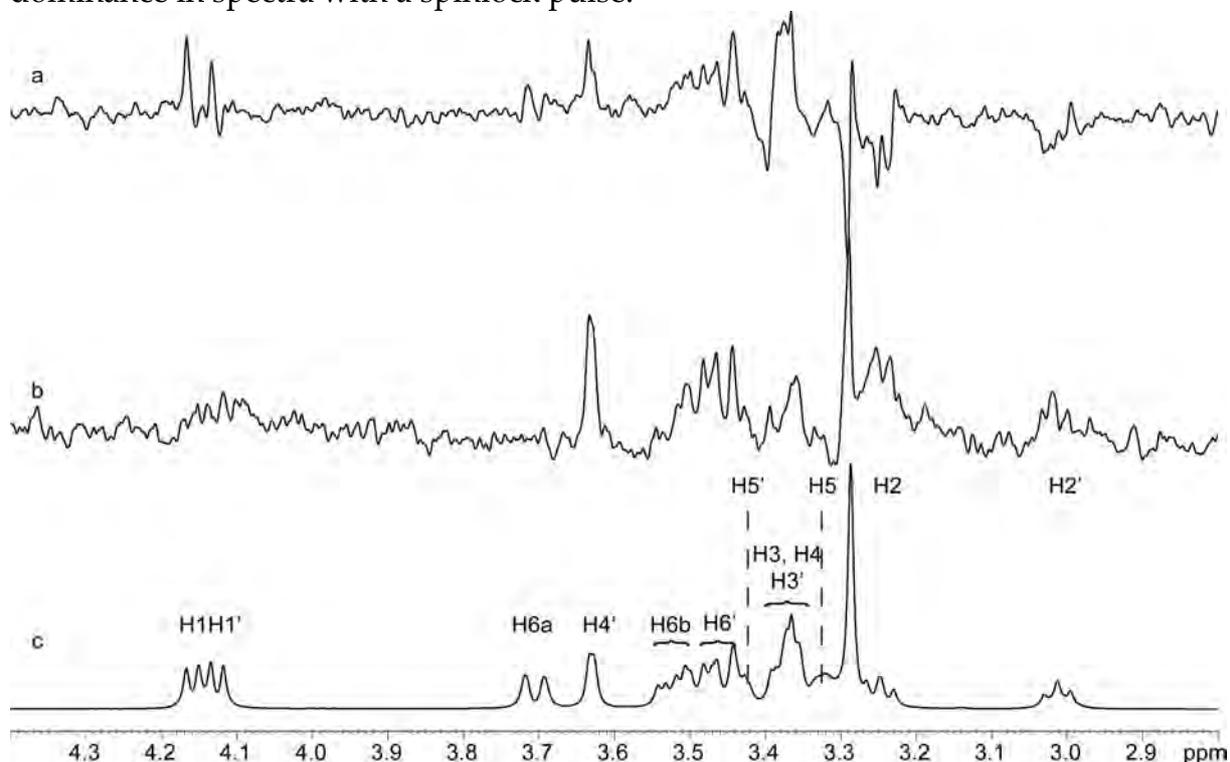


Figure 3.2: A: STD spectrum of β -D-Lactose with spinlock pulse program `std19slsp`; b: STD spectrum of the same sample without spinlock pulse (`std19sp`); c: reference spectrum with 30 ms spinlock. In all spectra the concentration of LacOMe was 909 μM (50 fold excess), $T=285\text{ K}$, the STD spectra were recorded with a pulse power of 55 dB, 1024 scans each on and off resonance and irradiation at 0 ppm on resonance and 40 kHz off resonance.

Investigation of experiments with different saturation times lead to the assumption that under the conditions applied in these experiments incomplete relaxation of the sample in between scans lead to the phase error. This was ascribed to the O-methyl group which was believed to comprise the highest relaxation time in the molecule. This may also explain the increasing error in vicinity of the O-methyl resonance (3.287 ppm). With increasing saturation time which allows further relaxation of the ligand molecule less phase error was detected.

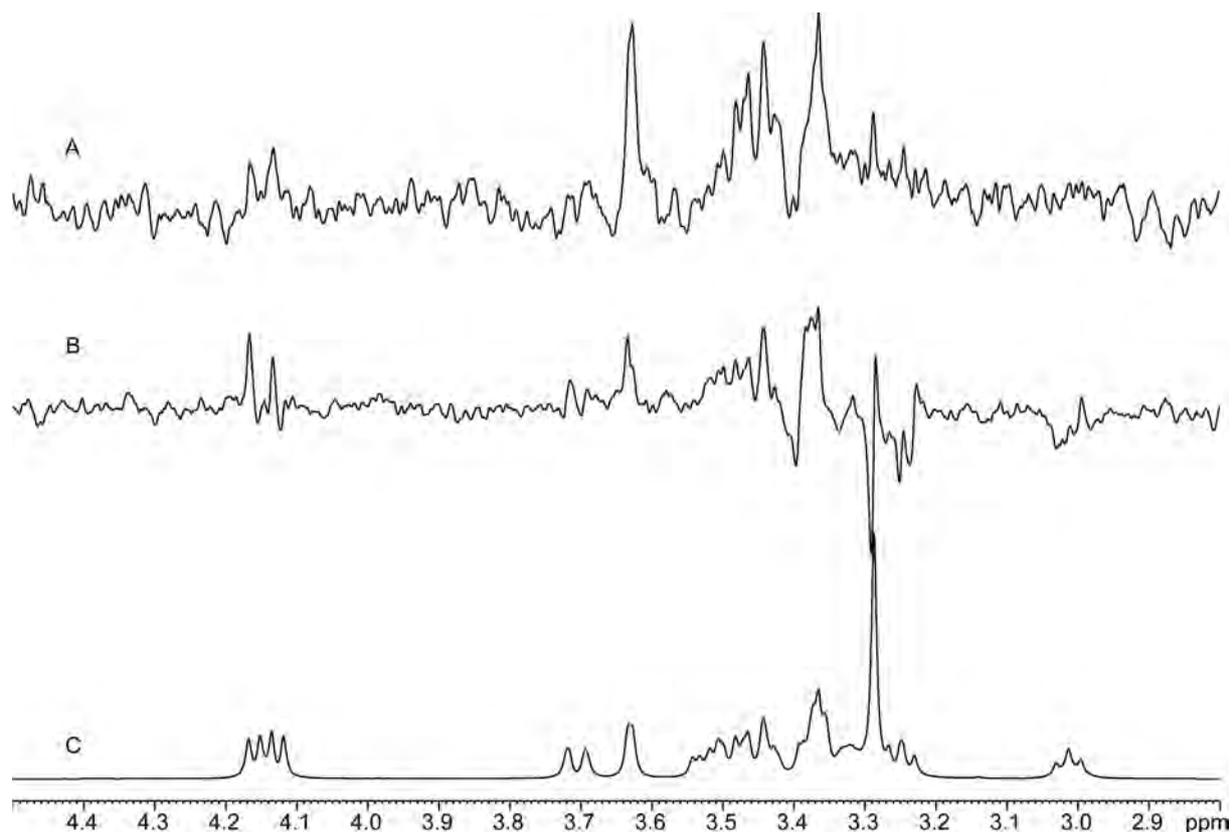


Figure 3.3: A: STD spectrum of LacOMe with $T_{\text{sat}}=2,5$ s, signal to noise ratio (S/N)=7.8; B: same sample with $T_{\text{sat}}=1,5$ s, S/N=11.0; C: reference spectrum, S/N=3756. All spectra were recorded with 30 ms spinlock pulse, at $T=285$ K, *on resonance* irradiation at 0 ppm and *off resonance* irradiation at 40 kHz. The STD spectra were recorded with 1024 scans in off and *on resonance* and the pulse program std19slsp. The reference spectrum was recorded with 128 scans. Ligand excess was 400 fold ($c(\text{LacOMe})=7$ mM).

It was decided to abandon the titration experiments and to concentrate on the determination of the binding epitope of β -D-LacOMe on contact with galectin-1. For the determination of the binding epitope the spectrum acquired with spin lock pulse, a saturation time of 2.5 s and 400 fold ligand excess was used. STD effects were determined comparing signal intensities of STD spectra with the respective reference WATERGATE spectrum considering the scan ratio of the spectra.

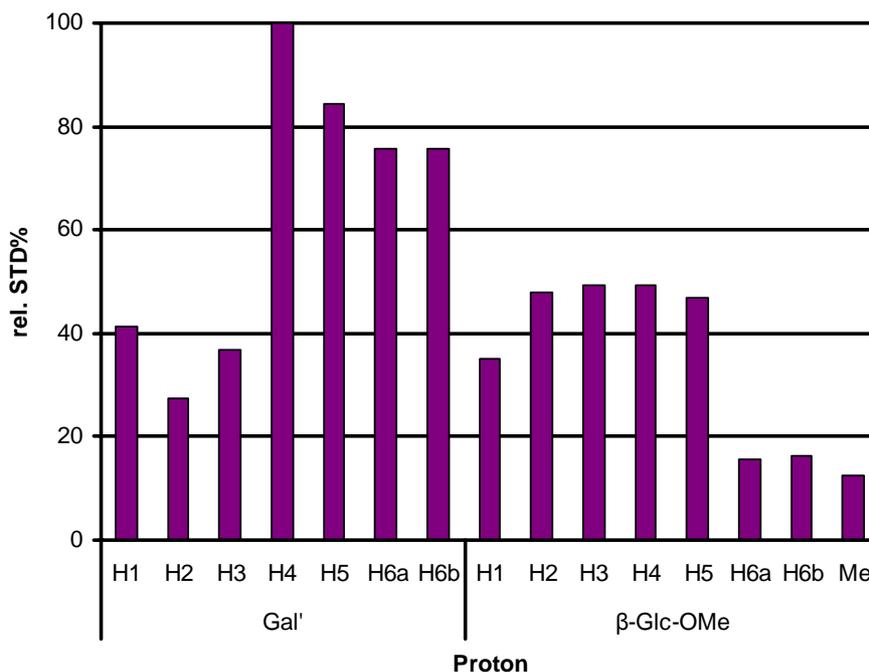


Figure 3.4: Epitope of LacOMe, represented by the relative STD percent: The resonances for H3, H4 and H6a' and H6b' could not be assigned distinctly. H4' to H6' of the β -Gal residue exhibit the strongest effects detected. The noise level in the STD spectrum was at 17%.

The STD NMR spectra revealed intensities for all resonances of the ligand. Inspections of several resonances in the STD spectrum exposed intensities in the range of the background noise level and thus could be discarded. From this a modified binding epitope was deduced. The resonances of H6a and H6b along with H2' did not stick out of the background noise, their relative STD percentage was determined to be around 16% for H6a/b and around 27% for H2'. P.J. Hajduk et al. applied a cutoff of 15% relative STD in their implementation of SOS STD NMR⁵⁴. In this case the cutoff is almost 30% relative STD intensity. This is due to the disadvantageous signal to noise ratio in the STD spectrum. The reason for the signal to noise ratio may be slow kinetics of the binding event or the long saturation time causing artifacts. The *O*-methyl group of β -D-LacOMe also caused a signal with intensity below this boundary (12.5%) but the signal clearly sticks out of the background noise level and thus was not discarded. The magnitude of the *O*-methyl resonance in the STD spectrum was three times higher than the magnitude of the noise level at 3.0 ppm (H2 resonance) and only half of the magnitude of the proton in 6'-position.

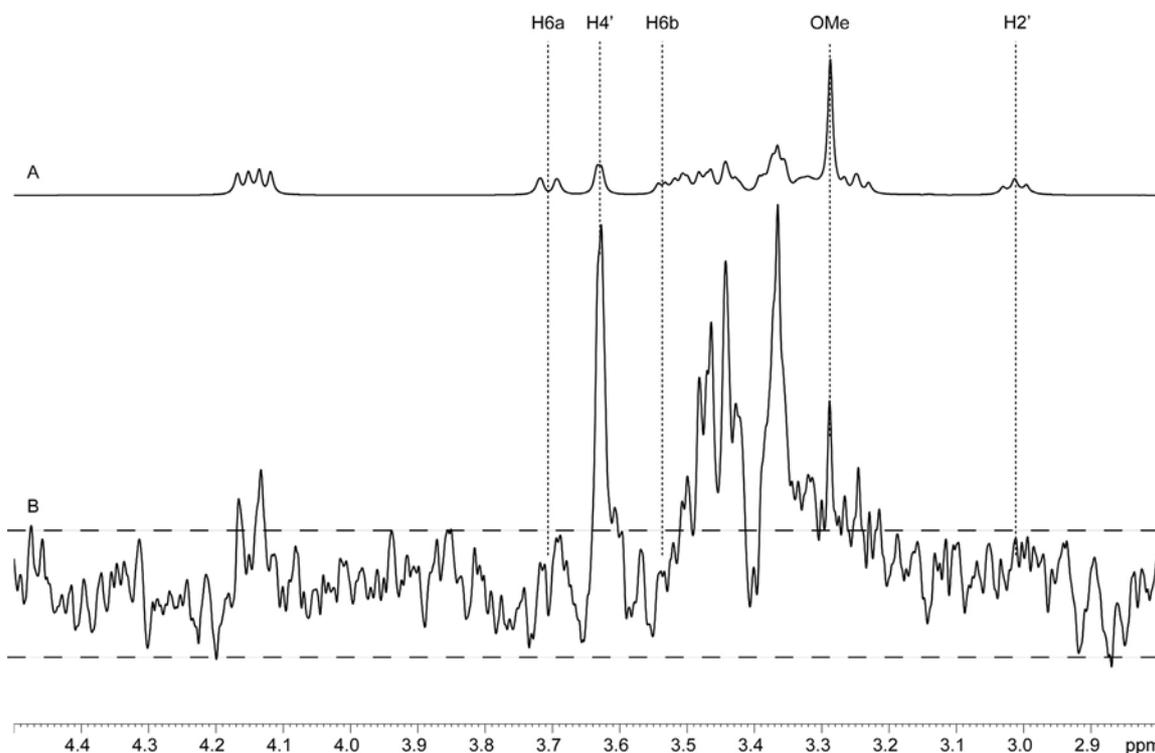


Figure 3.5: Spectra of LacOMe as before. Panel a: reference spectrum; b: STD spectrum with dashed lines indicating the boundaries of background noise level. Dotted lines mark the assigned resonances. H6a, H6b and H2' exposed intensities below the upper boundary and where thus discarded from the binding epitope.

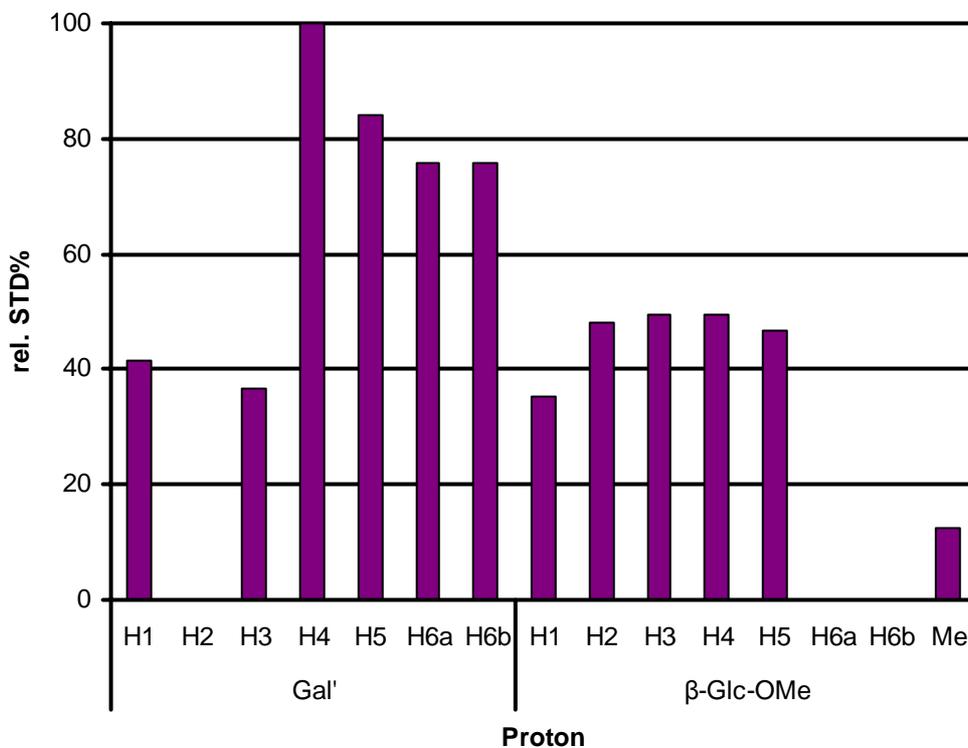


Figure 3.6: Epitope of LacOMe determined by STD NMR: Signals with intensity below background noise level were discarded.

3.3.2 *N*-acetyl lactosamine (type II)

N-acetyl lactosamine (type II, Gal β 1-4GlcNAc, LacNAc) is often taken as a reference molecule in galectin binding studies and is assumed to be the principle binding motif on glycoproteins binding *in vivo*.

The titration experiment of galectin-1 and LacNAc was carried out using an anomeric mixture of commercially available LacNAc. STD NMR titration experiments were carried out using 10 nmol galectin-1 in 530 μ L sample volume ($c(\text{galectin-1})=19 \mu\text{M}$). LacNAc was added ranging from 5 fold excess ($c(\text{LacNAc})=95 \mu\text{M}$) to 70 fold excess ($c(\text{LacNAc})=1.33 \text{ mM}$). STD NMR spectra were acquired at 300 K and the intensity of the methyl group from the *N*-acetyl glucosamine residue was determined by integration and taken for the determination the binding affinity. From the STD NMR experiments a dissociation constant of $K_D=55\pm 92 \mu\text{M}$ was determined. F.P. Schwarz et al. determined the dissociation constant of LacNAc to galectin-1 from bovine spleen to be $K_D=45 \mu\text{M}$ ($K_A=2.22\cdot 10^4 \text{ M}^{-1}$) by isothermal titration calorimetry (ITC)⁸³. If the differences in the experimental setup for the determination of the dissociation constant are taken into account it can be assumed that the deviation is negligible and the results were equivalent.

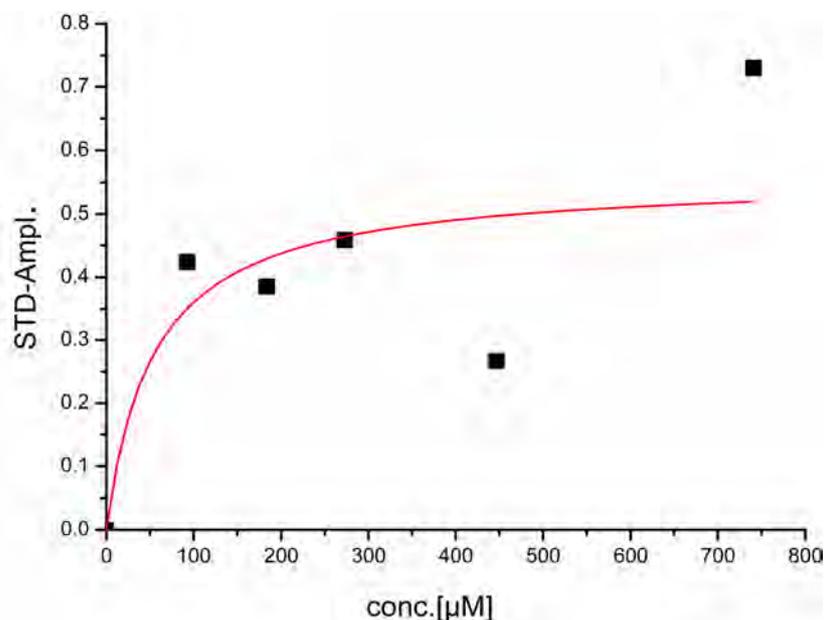


Figure 3.7: STD titration curve acquired for LacNAc. The results for the STD amplification factor for the individual concentrations are shown as squares. In the curve fitted after the one site binding model is shown.

The resulting STD spectra proved insufficient for epitope mapping. For epitope mapping of LacNAc the sample was composed of commercially available β -D-O-ethyl-lactosamine (LacNAcOEt) and galectin-1.

Determination of the STD effect was achieved by integration of the signals in the *off resonance* and STD spectra. Additionally the STD effect was determined by comparison of the signal intensities in the *off resonance* and difference spectra. Integration yielded intensities for H1, H1', H3, H2' and the protons from the

methylene groups, but in the STD spectrum the respective signals did not come out of the background noise. Artifacts were found for the signals of the methyl resonances from the N-acetyl residue and the ethyl group in the blank test. Artifacts were determined to be less than 3% by comparison of intensities. Other signals did not stick out of the background noise in the STD spectrum of the blank test.

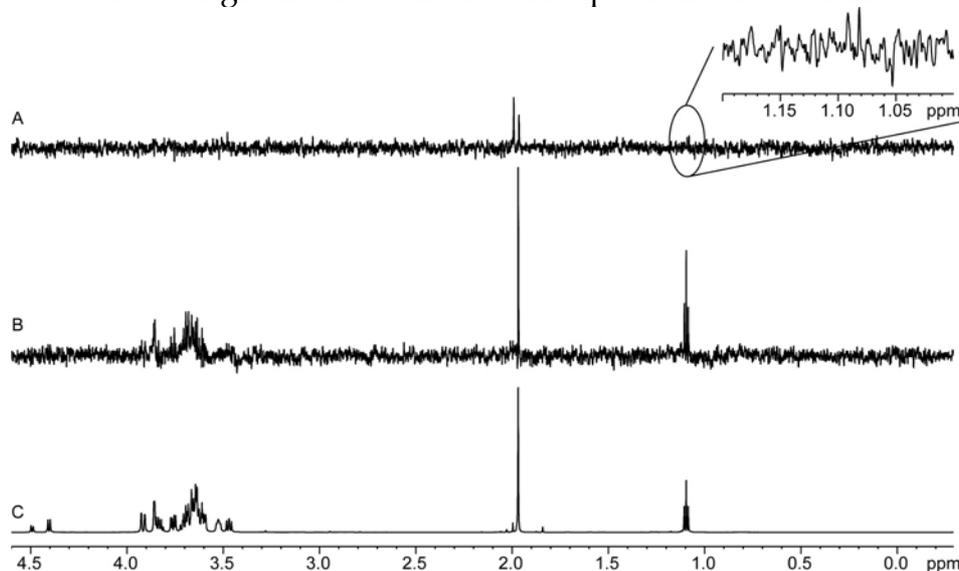


Figure 3.8: A: STD spectrum from blank test of LacNAcOEt ($c=1$ mM), the inset shows enlargement of the region where the signal of the methyl group from the ethyl residue (1.085 ppm). B: STD spectrum from LacNAcOEt ($c=1.6$ mM; 100 fold excess). C: *off resonance* spectrum, irradiation at 20 kHz. All spectra were acquired at 700 MHz with irradiation at -525 Hz *on resonance* and 20 kHz *off resonance*, 300 K, 43 dB pulse power for saturation pulses and 2 s saturation time.

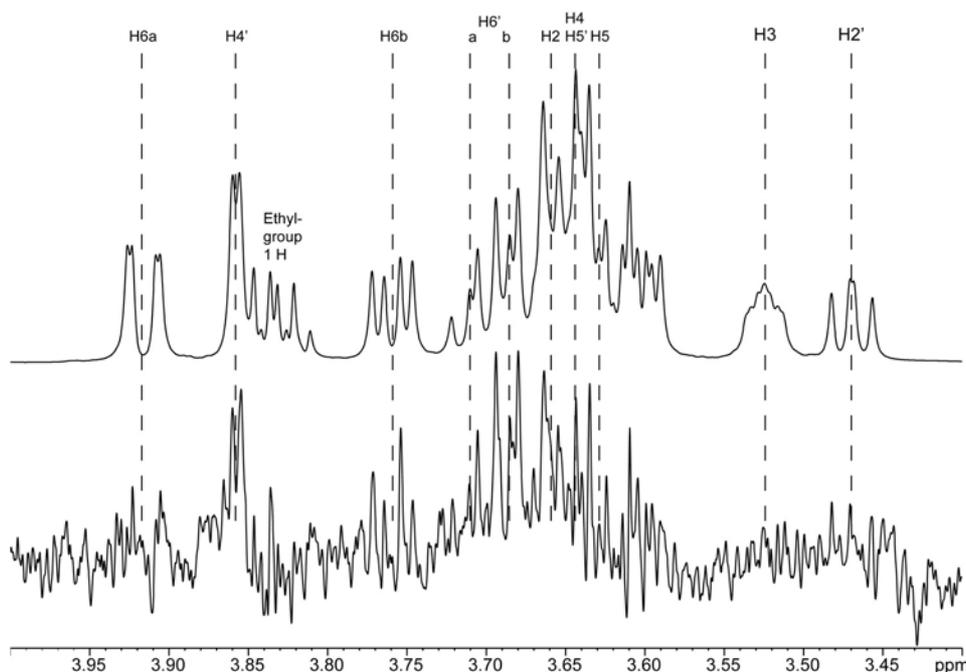


Figure 3.9: Section of the STD spectrum (bottom) and *off resonance* spectrum (top) of LacNAcOEt from 3.4 to 4 ppm: For H3 and H2' intensities were determined from integration but discarded since the signals did not come out of the background noise. The signal of the second proton of CH₂ from the ethyl group is located at ~3.65 ppm (not mentioned in assignment).

The comparison of the results gained by these two methods is shown in Figure 3.10. It became clear that the methods were yielding qualitatively equivalent results.

Based on the distance dependence of the STD effect from the X-ray structure analysis it was expected that the highest effects should be present on the H6 protons of galactose. The crystal structures solved for bovine galectin-1 in contact with lactosamine (1slt)⁷⁸ and a biantennary carbohydrate with a lactosamine residue on the non reducing termini (1sla)⁷⁷ both showed that the OH-group in 6-position of the Gal residue is deeply buried in the binding site. In the STD spectra H6a' and H6b' exhibited the strongest intensities for the epitope supporting this expectation. The strong effect determined for the methyl group of the ethyl residue (H2) hinted to inaccurate artifact handling.

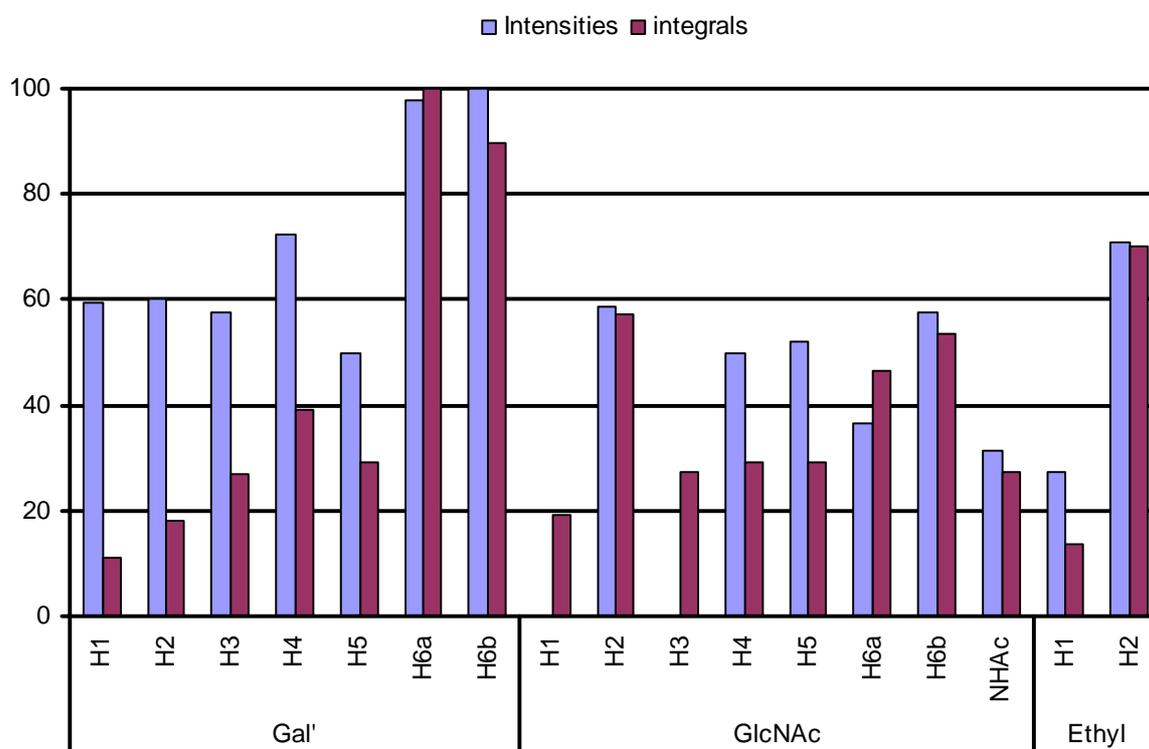


Figure 3.10: The comparison of relative STD percentages determined by integration and signal intensity reveals the lesser sensitivity of the integration method. H1' to H3' were determined to be almost equal in intensity while the integration method reveals differences for individual protons. H1 and H3 did not rise out of the noise level, H4 and H5' could not be discriminated.

3.3.3 Lacto-*N*-biose (*N*-acetyl lactosamine type I)

Several studies have shown that Lacto-*N*-biose (Gal β 1-3GlcNAc, LNB) is binding to galectins and affinity to galectin-1 is similar to that of LacNAc. N. Ahmad et al. report the affinity with $K_D = 111 \mu\text{M}$ ($K_a = 0.9 \text{M}^{-1} \cdot 10^{-4}$) to bovine heart galectin-1, determined by ITC⁸¹.

STD spectra of LNB were recorded using an anomeric mixture of α - and β -LNB since no derivative with locked anomeric position was available. The epitope derived from STD NMR resembles that of lactosamine type II with the highest intensities

detected on H6' and H5'. The intensities of the protons of the GlcNAc residue were almost zero except for the acetyl group, H6 and H3 in the α -configuration. With the exception of H2' and H3' the intensities of the protons of the Gal residue were of equivalent magnitude when determined by integration for both anomers. Regarding the intensities determined by integration for the β -anomer it was obvious that the epitope resembled that of the type II lactose derivatives. H6' and H5' exhibit the strongest intensities and H2' is the weakest signal in the STD NMR spectrum.

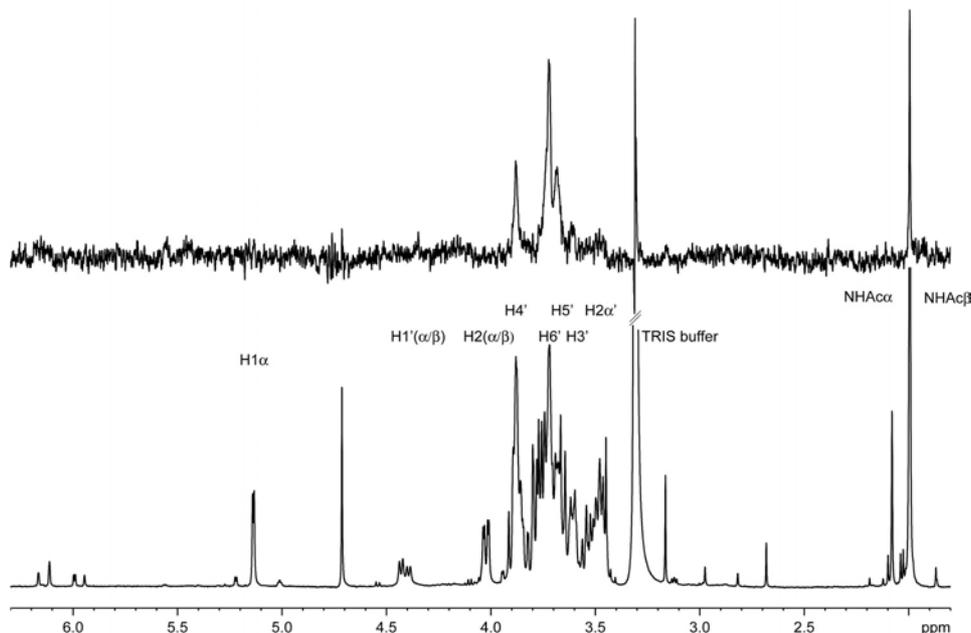


Figure 3.11: 1D STD NMR spectrum of the anomeric mixture of LNB. Acquired at 300 K, irradiation at 0 ppm *on resonance* and 40 kHz *off resonance*, 2048 scans 1.5 s saturation time and 45 dB pulse power for the saturation pulses.

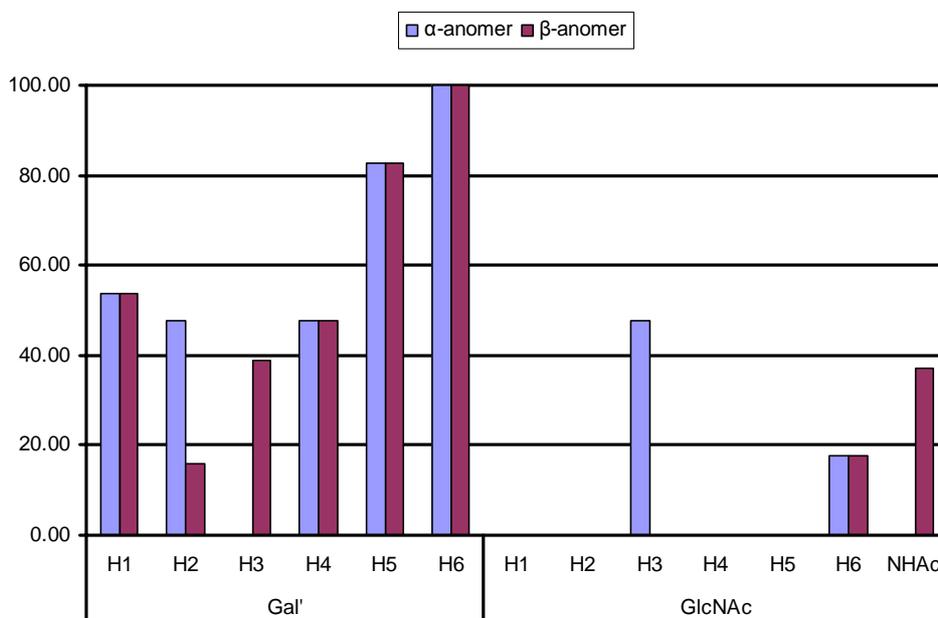


Figure 3.12: Epitope for both anomers of LNB. For most resonances of the Gal residue no difference in their chemical shift is seen, thus the same STD intensities arise. The binding pattern resembles that of lactose and N-acetyl lactosamine for the H5' and H6' positions.

3.3.4 Phenyl- β -D-galactoside

β -D-phenyl galactoside (PheGal) is an unnatural galactose derivative. It was chosen as a reference compound for disaccharides.

The binding affinity of PheGal to galectin-1 was determined by STD NMR titrations as described for LacNAc. The sample contained 10nmol of galectin-1 in 540 μ L sample volume resulting in 18.5 μ M concentration of the receptor. PheGal was added ranging from 5 fold excess ($c(\text{PheGal})=92.5 \mu\text{M}$) to 50 fold excess ($c(\text{PheGal})=925 \mu\text{M}$). For the determination of the STD amplification factor the integral of the resonances of the aromatic protons from the phenyl ring was used. The titration experiment of PheGal with galectin-1 yielded a selective binding curve with $K_D=313 \mu\text{M}$.

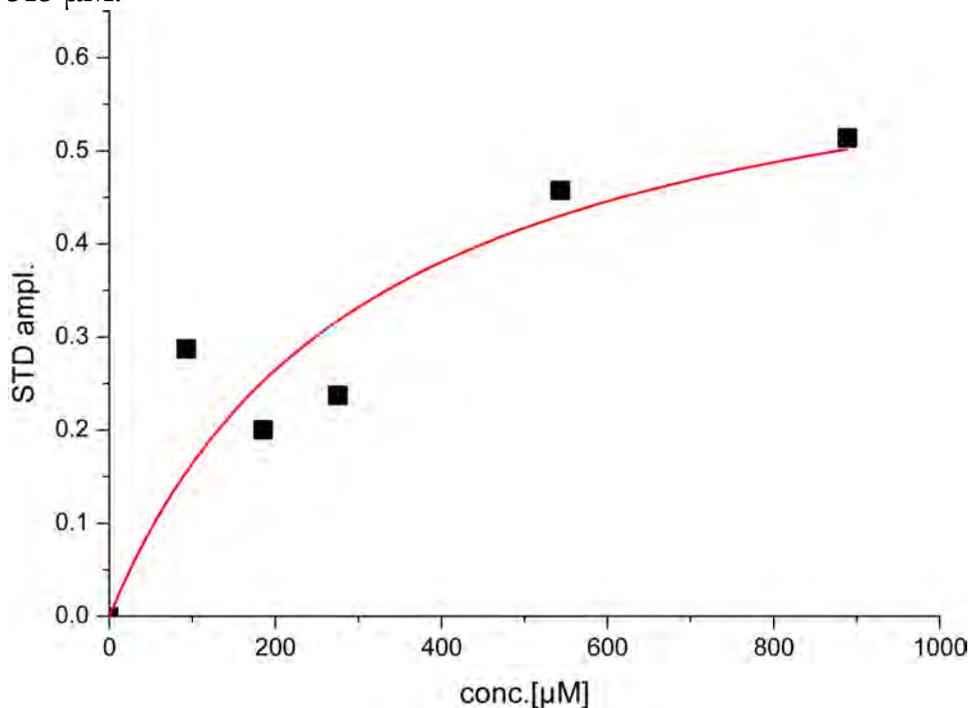


Figure 3.13: STD titration curve acquired for PheGal. The STD amplification factors are presented as squares, the fitting curve for the one site binding model as red line.

The last sample of the titration experiment was used for the epitope mapping. The sample contained 500 nmol PheGal, equivalent to a 50 fold excess to the protein. The STD spectrum had considerable background noise but clearly depicted a selective STD effect. At the temperature of 300 K the signals from H6, H5 and H2 significantly overlapped. The strongest STD effect was determined for the aromatic protons in *meta* and *para* position of the phenyl ring. These signals also overlapped. The H6 position did not exhibit the strongest STD effect for the carbohydrate protons. This could be accounted for by the overlap with other signals. If it is assumed that H2 had a minor or no contact with the protein the resulting STD intensity would apply only for H5 and H6. H6 and H5 always had significant intensity in the disaccharides described above.

It became obvious, that H3 and H1 signal intensities were almost at the level of the background noise. Thus signal intensities in the STD spectrum were checked again

by comparison of their signal intensity with the intensity of the *off resonance* spectrum.

This led to a different binding epitope. The relative STD values of H6, H5 and H4 compared to each other were similar to those noted for these protons in the Gal residue in disaccharides. H1, H2 and H3 exposed an even higher relative STD value than determined by integration. The strongest intensity of all protons was detected for the phenyl protons in *ortho* position.

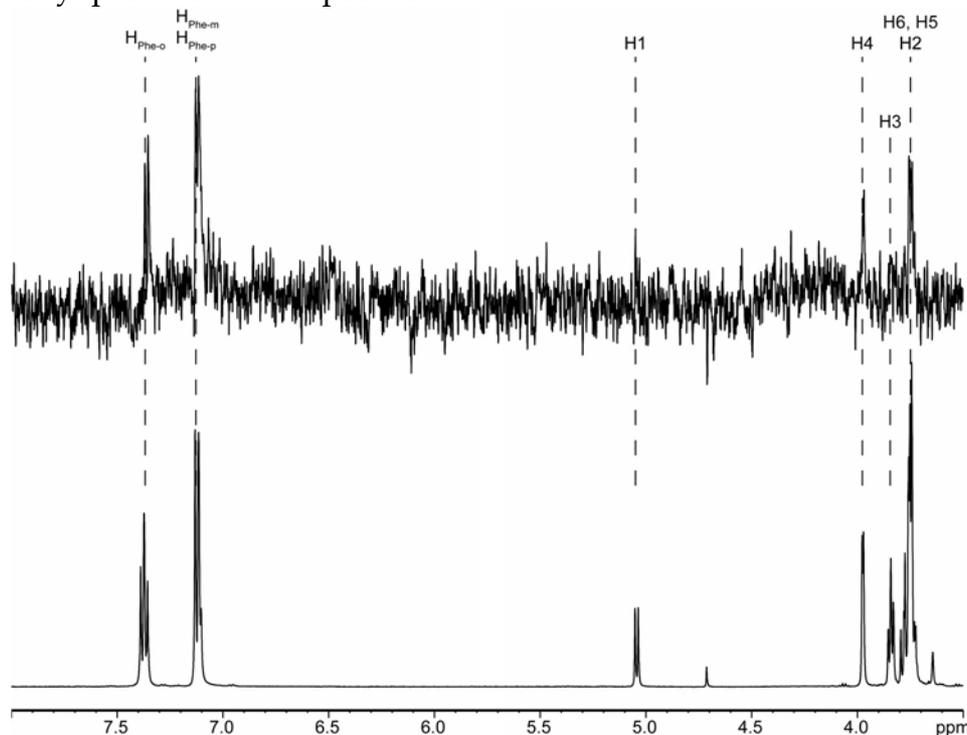


Figure 3.14: Top row: STD spectrum of PheGal at 50 fold excess, 2048 scans, 300 K, 1.5 s saturation time, 45 dB pulse power and irradiation at 0 Hz on and 40 kHz *off resonance*. Bottom row: *off resonance* spectrum for reference. Line broadening of 1 Hz was applied. This allowed distinguishing between several signals in the hump at 3.75 ppm.

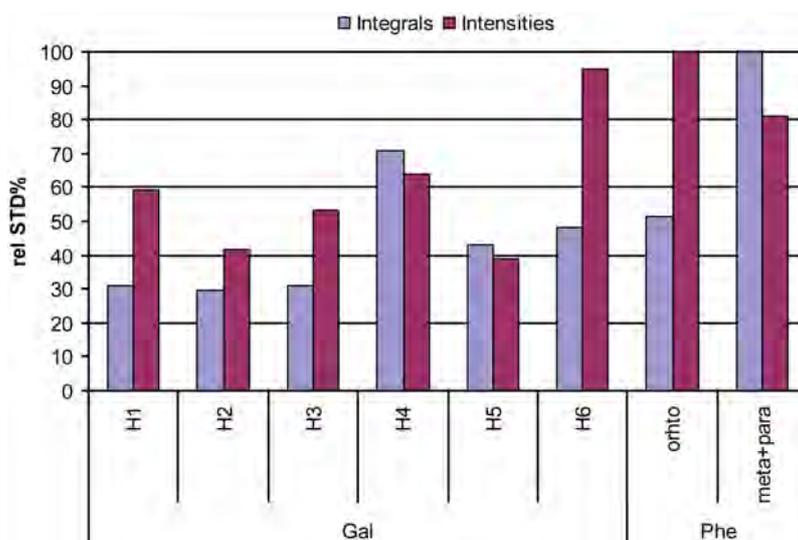


Figure 3.15: Comparison of the relative STD percent determined by integrals and intensities. H6 of the Gal residue represents the most intensive signal in the carbohydrate. The protons in *ortho* position of the phenyl ring are the strongest signal.

For lactose and *N*-acetyl lactosamine in complex with galectins studied by crystallography hydrogen bonds are often found to form between protein and carbohydrate. Both, direct hydrogen bonds and bonds mediated by water molecules occur. These hydrogen bonds contribute significantly to the binding affinity and cause a structural lock for the bound ligand. The phenyl ring in PheGal can not form hydrogen bonds, thus the affinity is lower. Additionally it can be assumed that in the bound state the phenyl ring is not locked by hydrogen bonds and is capable to rotate along the glycosidic linkage. The protons in *ortho* position are closest to the bound Gal residue and thus to the protein and therefore gain the highest saturation transfer from the protein.

3.4 Estimation on the binding kinetics based on STD NMR results

All STD NMR spectra acquired using galectin-1 and different carbohydrates in this study showed low intensities of the signals in the difference spectra. This resulted in low signal to noise ratios (S/N). The possibility to acquire STD NMR spectra and the quality of the STD NMR spectra is related to the binding kinetics of the system studied. From Equation 1.2 it can be derived that k_{on} and k_{off} are key factors for the intensities detected in the difference spectrum.

Low off-rates cause a low turn over. Thus, within saturation time only few ligand molecules reach the binding site of the protein and therefore, do not achieve strong saturation. This causes low intensities in the STD spectrum because the concentration of free ligand molecules with saturated resonances in the sample is low.

Very high on- and off-rates result in an insufficient saturation transfer. The ligand molecules are in contact with the protein binding site only for a short time.

W. Dettmann *et al.* determined the off rate of bovine galectin-1 to immobilized lactose and asialofetuin, which exposes terminal *N*-acetyl lactosamine residues, by surface plasmon resonance. Dettmann *et al.* determined $k_{off}=0.5 \cdot 10^{-3} \text{ s}^{-1}$ for lactose and $k_{off}=1.1 \cdot 10^{-3} \text{ s}^{-1}$ for asialofetuin. With a K_D value of $155 \mu\text{M}$ for lactose^{83;84} k_{on} calculated to be $3.32 \text{ s}^{-1}\text{M}^{-1}$ using Equation 1.3. Taking the k_{off} value of asialofetuin as reference off rate for *N*-acetyl lactosamine and $K_D=55 \mu\text{M}$ as determined in this study (Figure 3.7) the respective on-rate is $k_{on}=20.0 \text{ s}^{-1}\text{M}^{-1}$.

The results are in accordance with estimations made from the buildup of magnetization of resonance from different saturation times (data not shown). Especially the off-rate is too low to allow strong signals and high S/N ratios in the difference spectrum.

3.5 Development of a docking procedure

The classical docking process requires structural data of the receptor. Galectin-1 from *bos taurus* and *homo sapiens* has been crystallized in complex with carbohydrates^{74;77;78;85}. For the docking studies in this thesis the galectin from *bos taurus* from crystal structures 1sla⁷⁷ and 1slt⁷⁸ were taken. 1sla is a crystal structure

comprising of bovine galectin-1 in complex with a biantennary oligosaccharide with *N*-acetyl lactosamine residues at the end of the antenna. 1slt is a crystal structure comprising of bovine galectin-1 in complex with *N*-acetyl lactosamine. Both crystal structures contain the protein as a homodimer. The high homology to human galectin-1 allows a comparative discussion (Figure 1.4).

Docking experiments with the receptor from 1slt required first correction to the published X-ray structure by repairing side chains and by modifying of the oxidized cysteine residues. In 1slt the thiol group of Cys2 is disordered and thiol groups in Cys16, Cys88 and Cys120 occur in oxidized state. D. Liao *et al.* assumed that cysteine oxidation did not alter carbohydrate binding properties⁷⁸. In this study several samples without DTT for protection against oxidation lost their binding affinity. Thus binding could not be detected by STD NMR. The protein coordinates from 1slt required modifications to allow proper handling of the cysteine residues. Further modification was required regarding the side chains of amino acids Ala25, Asp26, Gln93, Thr94 and Glu115. All these residues are located outside the binding site. The protein from the X-ray structure 1sla did not require any modifications prior to docking. Without modifications of the published data docking results acquired with the protein structure were incomprehensible. This may be ascribed to the incapability of GRID to properly deal with these side chains while creating the scoring grid for the receptor binding site. Docking spheres were generated using the program SPHGEN of the DOCK software suite. Sphere clusters were checked to occupy the CRD. Spheres outside the CRD were deleted.

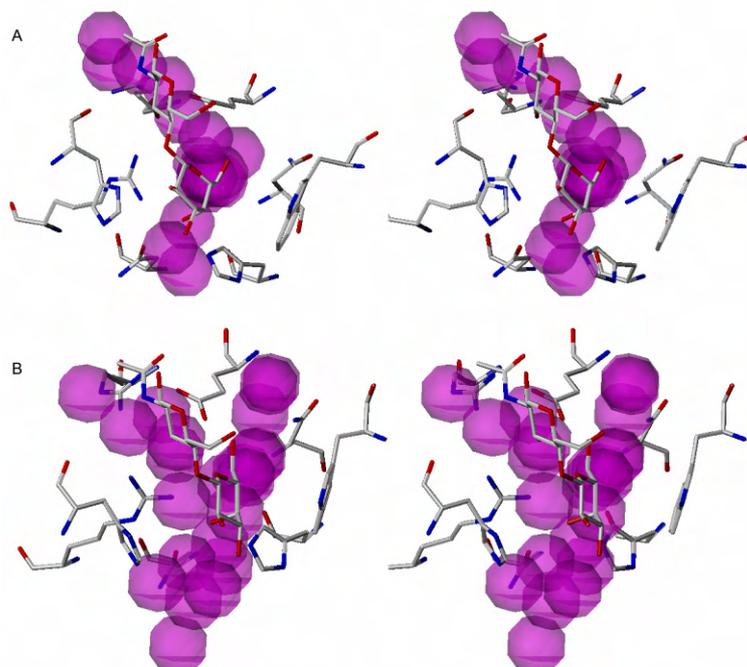


Figure 3.16 A: Spheres for docking ligands into the binding sites of 1sla. B: Spheres for docking ligands into the binding sites of 1slt. Spheres are colored in magenta and rendered in transparent space fill. Shown are the main residues of the binding site, His44, Asn46, Arg48, His52, Asn61, Trp68, Glu71 and Arg73, and the heavy atoms of the respective *N*-acetyl lactosamine residues.

First attempts were made using the protein structure from 1sla and standard parameters for most variables in the setup of Dock as suggested in the manual. Other than suggested in the manual, ligands were treated as flexible.

The results for docking a flexible LacNAc disaccharide into the binding site of 1sla were unsatisfactory. DOCK was unable to generate a complex of the ligand extracted from the crystal structure 1sla and the protein from 1sla that resembled the X-ray structure.

Since the ligand structure docked to galectin-1 was taken from the same X-ray structure induced fit effects could be ruled out. Induced fit effects have been shown to dramatically affect the ability of docking programs to reproduce the binding mode from the X-ray structure⁸⁶. Following the procedures applied by Erikson *et al.* to assess the impact of ligand flexibility and change in receptor structure in different complexes⁸⁶ it was tried to dock a different LacNAc disaccharide to galectin-1 from 1sla. This was done in order to check whether the failure to dock the disaccharide taken from the X-ray structure 1sla to galectin-1 was a result of errors in the coordinates of receptor or ligand, the conformation of the ligand or the setup of the docking procedure. Several X-ray structures of galectins in complex with *N*-acetyl lactosamine have been published. The X-ray structure 1slt represents bovine galectin-1 in complex with *N*-acetyl lactosamine⁷⁸. The disaccharide ligand from 1slt was docked to galectin-1 from 1sla with the same parameters as before. The results obtained with this ligand were not in agreement with the binding mode of LacNAc in the crystal structures. Since no indications were found that errors were present in the coordinates of ligand and receptor it was decided that the setup for the docking procedure was inadequate to dock a flexible disaccharide ligand into a given binding site.

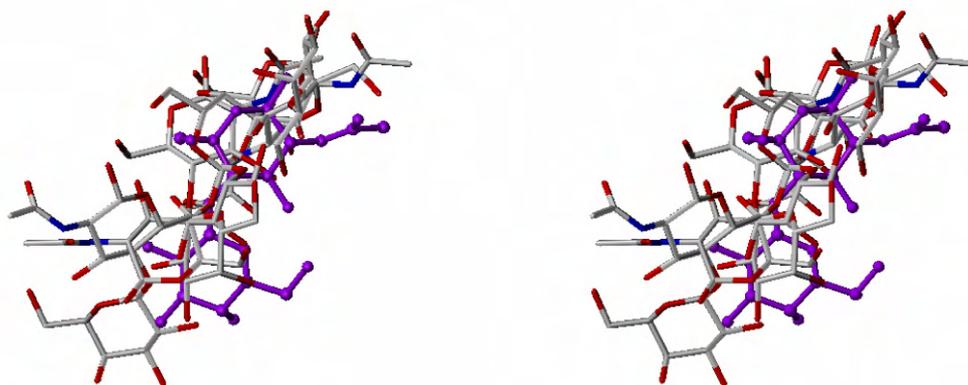


Figure 3.17 Shown are the five top scoring positions of 20 poses from a docking run using default parameters for flexible scoring. For comparison the position of the ligand from the X-ray structure is given in violet and ball and stick rendering, protons are not shown. It is obvious that none of these poses is in good agreement with the data from crystal structure. None of the 20 poses was in agreement with the crystal structure (not shown for clarity of display). This result was achieved by using protein chain A and a ligand from 1sla.

Since the default setup of DOCK failed to produce reasonable results it was decided to test the performance of DOCK using custom features. DOCK allows the definition of an anchor fragment⁸⁷. An anchor fragment must be a rigid segment of the ligand molecule, i.e. a ring structure. DOCK does not change the structure of rings in ligand molecules. Once the anchor segment has been defined the molecule is automatically divided into non overlapping segments. The anchor is processed first in the docking process. The anchor can be oriented and scored separately from the remaining molecule. The orientation of the anchor can be optimized by energy minimization prior to further processing. After placement of the anchor segment remaining parts of the molecule are reattached to the anchor starting with the largest fragment. During this phase conformational search is performed by optimizing the orientation of the reattached segments. The final ligand structure can be minimized again.

Based on the established specificity for galactose residues of galectin-1 and the finding of the STD NMR results in this thesis (chapter 3.3.2) it was decided to define the pyranosyl ring of galactose as the anchor fragment.

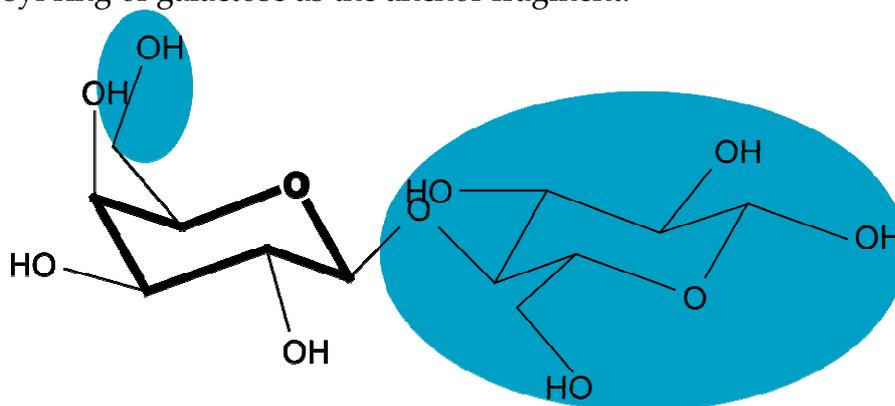


Figure 3.18: Schematic representation of the ligands anchor, highlighted in bold and the fragments attached after placing the anchor fragment in the binding site (blue background). Since DOCK treats ring systems as rigid entities the conformation and orientation of atoms bound directly to the anchor fragment is predetermined already (O1'-O4' and C6' of Gal residue). Hydrogen atoms of the hydroxyl groups are treated flexible.

After initial placing of the molecule, prior to the minimization of the complex, ligand atom may occupy space already occupied by atoms of the receptor. DOCK checks for these overlaps of ligand and receptor atoms and discards poses exceeding a given number of overlaps, also referred to as 'bumps'. During optimization and minimization the docked ligand can recover from some clashes. In this study a maximum occurrence of two bumps per pose was allowed.

With these adjustments the results from docking experiments improved drastically. Several ligands were subsequently docked to the protein structure taken from 1sla.

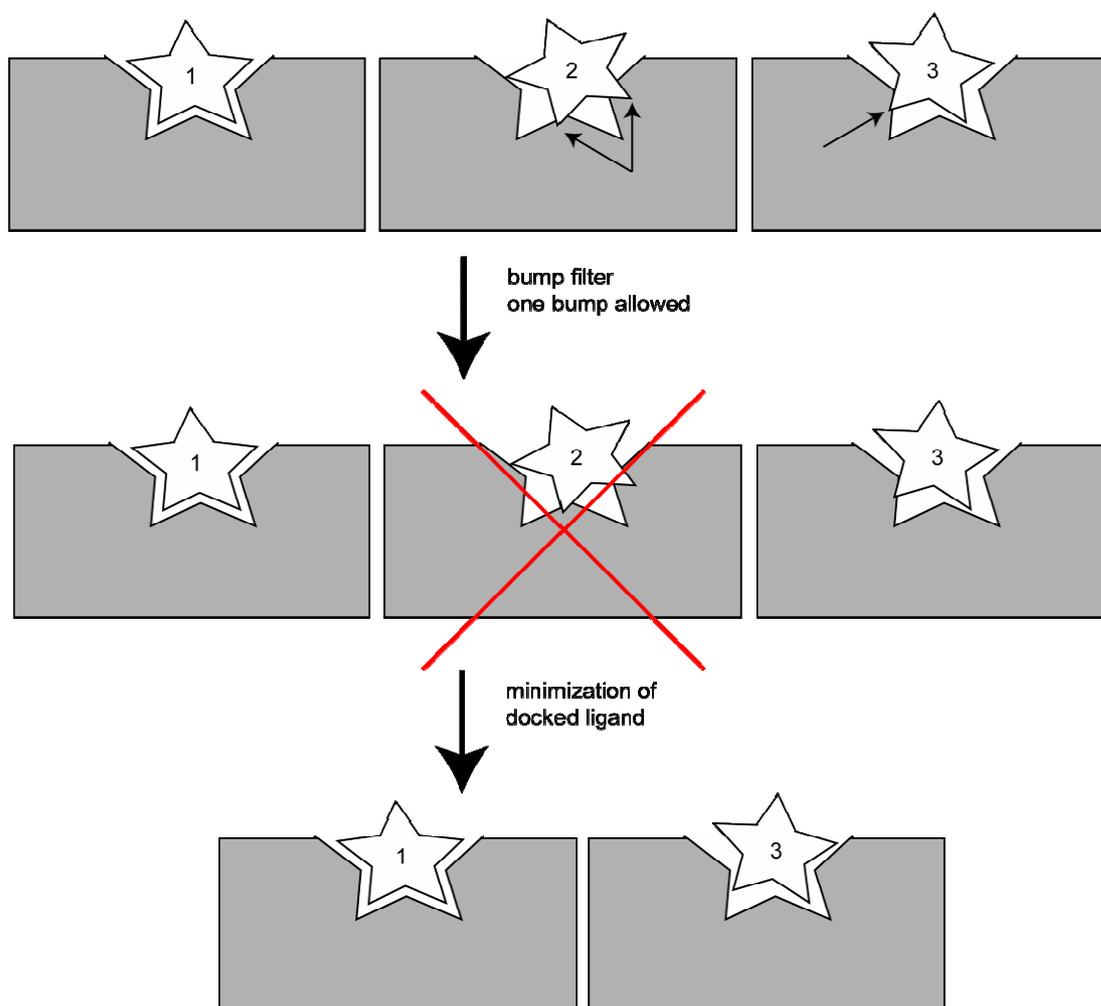


Figure 3.19: Functionality of the bump filter implemented in DOCK. The top row schematically depicts three docking poses directly after initial placement of the ligand. Pose 2 has two bumps, pose 3 one bump (indicated by arrows). Pose 1 does not have a bump. If only one bump is allowed pose 2 is discarded (middle row). Final minimization of the ligand 'recovers' the ligand by either moving the whole molecule or flexible parts of the ligand (bottom row).

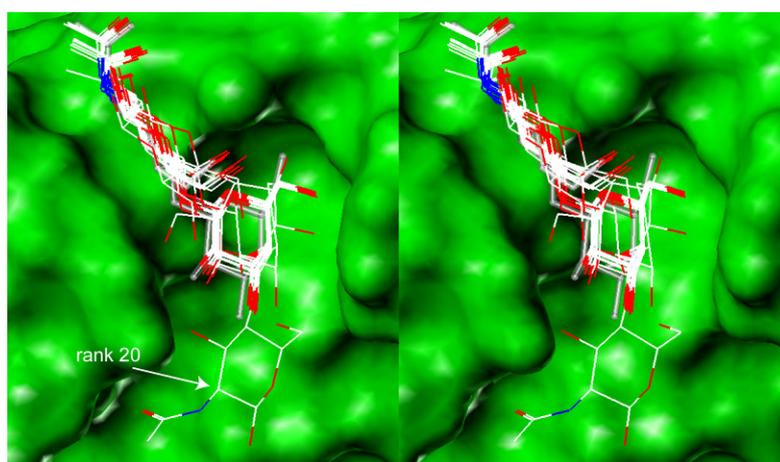


Figure 3.20 After docking LacNAc (taken from 1sla) to galectin-1 (1sla) only the last of the 20 best poses is docked in a different orientation from the cluster and the template from the crystal structure. Protons were omitted for clarity.

3.5.1 Docking of lactose and lactose derivatives

The docking procedure that was capable of docking the LacNAc disaccharide taken from 1sla into the binding site of galectin-1 from 1sla was applied to dock various disaccharides into the binding site of galectin-1. The LacNAc disaccharide ligands used were extracted from the crystal structures 1slt and 1a3k⁷⁶. Additionally, a model of ethyl- β -D-N-acetyl lactosamine (LacNAcOEt) was built for which experimental data on the binding epitope were determined by STD NMR (Figure 3.10). The LacNAc ligand from 1slt was modified to obtain a model of LacOMe where experimental data from STD NMR were also available. These experiments were made to assure that docking of the LacNAc ligand was not an artifact because the same ligand occupied the binding site in the structure that was used for the docking procedure.

For all type II (Gal β 1-4Glc) based lactose derivatives the majority of the top 20 scoring results were placed in agreement with the pose of the original ligand. The best cluster of results was achieved from the ligand structure retrieved from 1slt. The resulting cluster did not exhibit a single outlying pose, which is better than for the original ligand. This led to the conclusion that the docking protocol was capable of dealing with various disaccharide ligands. The result for ethyl- β -D-lactosamine (LacNAcOEt) was less consistent when compared to the other lactosamine ligands. The different placement of the ligand in comparison to the reference may be a result of the additional rotatable bonds resulting from the ethyl group. J.A. Erickson *et al.* found that the number of rotatable bonds in a ligand is affecting the accuracy of poses proposed from flexible docking experiments as the conformational sampling may become incomplete²⁷.

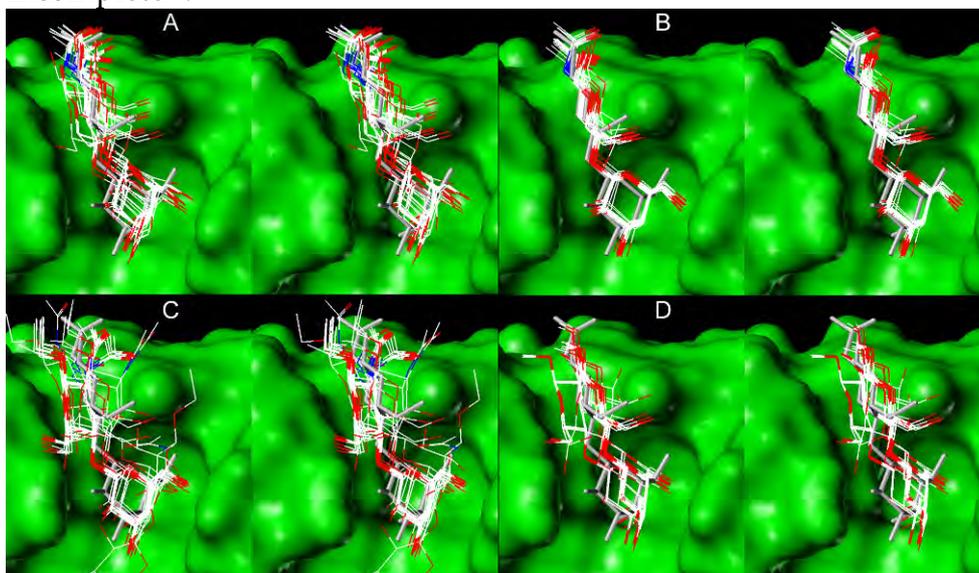


Figure 3.21 Shown are the docking results of several lactose (type II) derivatives docked to galectin-1 from X-ray structure 1sla (green surface). A: LacNAc ligand extracted from X-ray structure 1a3k, B: LacNAc ligand extracted from X-ray structure 1slt, C: LacNAcOEt, D: LacOMe. For comparison the heavy atoms of the LacNAc residue in 1sla are shown in white and capped stick rendering in all panels, all panels are in crossed eye stereo view, hydrogen atoms were omitted for clarity.

3.5.2 Docking phenyl glycosides of β -D-galactose

The phenyl ring of phenyl galactosides is a six-membered ring that may occupy the space of the Glc/GlcNAc residue in lactose or *N*-acetyl lactosamine. As shown in Figure 3.23 some of these ligands were docked in a fashion similar to that of *N*-acetyl lactosamine. In contrast to the pyranosyl ring of glucose in the lactose derivatives, the phenyl ring is planar. This leads to a worse fit in the binding site.

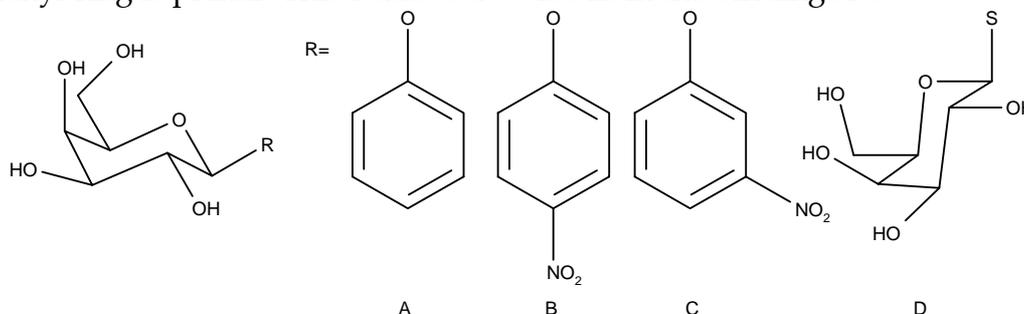


Figure 3.22: Ligands docked to galectin-1. A: PheGal, B: *p*-nitro-PheGal, C: *o*-nitro-PheGal, D: thio-digalactoside (TDG)

The pyranosyl ring of galactose was docked in the same region as the galactosyl residue in the reference X-ray crystal structure. In almost all of the results the galactosyl residue is placed in a well defined region with minimal variations in position and structure. This is not true for *o*-nitro-phenyl galactoside (panel B in Figure 3.23). Here the resulting cluster is poorly defined and comprises a lot of orientations. The docking run with *p*-nitro-phenyl- β -D-galactoside (pnitro-PheGal) yielded a cluster with the least deviation. The nitro group in *para* position of the phenyl ring is extending into the region occupied by the *N*-acetyl group of the original ligand.

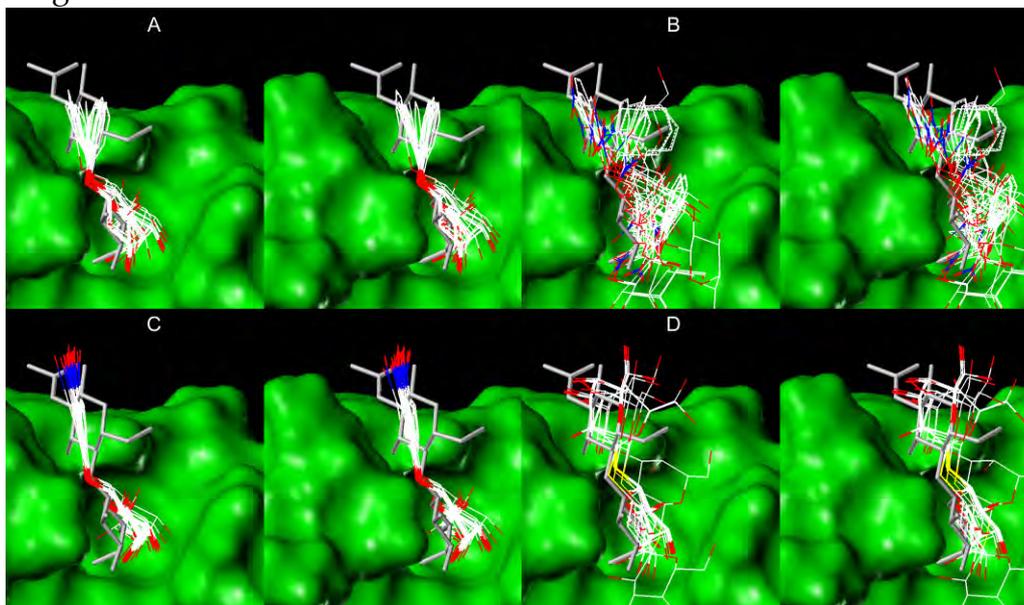


Figure 3.23 Shown above are results for some disaccharide like ligands in the database. A: β -D-O-phenyl galactoside. B: β -D-O-*o*-nitro-phenyl galactoside. C: β -D-O-*p*-nitro-phenyl galactoside. D: β -D-thio-digalactoside. In all panels the position of the heavy atom from the ligand in 1sla is indicated in white and capped stick rendering. All panels are in crossed eye stereo representation, hydrogen atoms were omitted for clarity.

As shown in the STD titration experiment with β -D-phenyl galactoside (PheGal) the affinity is reduced 6 fold compared to that of LacNAc. This may be the result of decreased spatial fitting into the binding site and the loss of polar interactions from the hydroxyl groups of the GlcNAc residue. In p-nitro-PheGal the nitro group may stabilize the complex by forming a polar interaction with the guanidine group from Arg73. In the docking result for *o*-nitro- β -D-phenyl galactoside the 20 structures saved do not form a well defined cluster. The best ranked pose for *o*-nitro phenyl galactoside may form a hydrogen bond with H δ in the imidazole ring of His52.

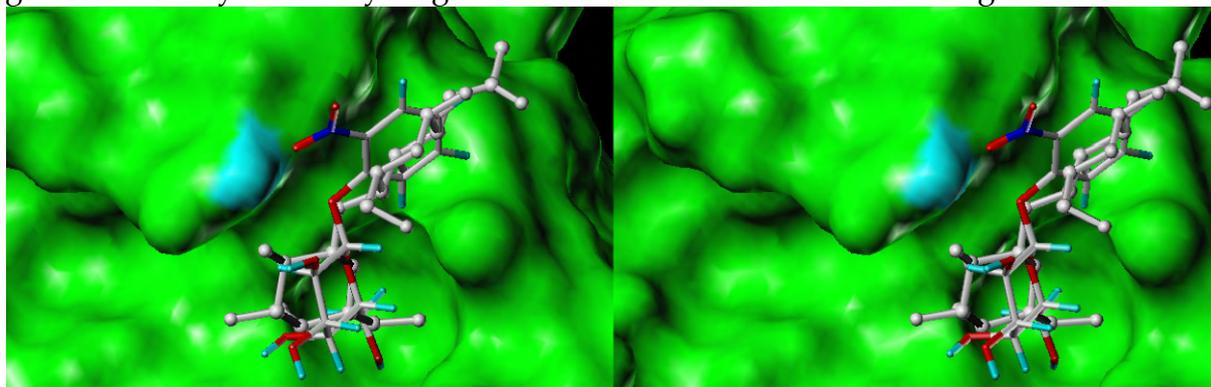


Figure 3.24 In this orientation the nitro group may form a hydrogen bond with H δ of His52 (cyan patch on surface, picture is in crossed eye stereo mode).

Thio-digalactoside (TDG) is an unusual, non natural ligand for galectins. In comparison to *p*-nitro-phenyl galactoside the resulting cluster is less consistent. The lowest ranked pose is completely displaced in comparison to the reference position of lactosamine. The quality of the docking result for TDG is comparable to the result obtained for the LacNAc disaccharide ligand taken from the X-ray structure 1sla.

H. Ahmed *et al.* report a relative activity for PheGal of 75% compared to Gal (100%), while the nitro-phenyl- β -D-galactosides were slightly more active than galactose. The relative activities were reported as 125% for *ortho*-phenyl-galactoside and 138% for *para*-phenyl-galactoside⁸⁸. Their study indicates a 528 fold stronger affinity of LacNAc compared to *ortho*-nitro-phenyl galactoside and a 472 fold stronger affinity of LacNAc compared to *para*-nitro-phenyl galactoside, determined by 50% inhibition of galectin-1 binding to asialofetuin. In the same study Ahmed *et al.* report a relative activity of 320% of TDG in comparison with lactose, equivalent to 32% relative activity in comparison with *N*-acetyl lactosamine. Schwarz *et al.* report a relative activity of 107% of TDG in comparison with LacNAc⁸³.

Polar interactions of Arg73 of galectin-1 with the GlcNAc residue of *N*-acetyl lactosamine have been described for the X-ray structures 1slt and 1sla. Weaker binding of phenyl galactoside can be explained with the loss of these interactions. The nitro group in *para* position is capable of forming a polar interaction with Arg73 when the galactose residue of *p*-nitro-phenyl galactoside is docked in the orientation of the galactose residue of LacNAc in the crystal structures. In the crystal structure 1sla the oxygen atom of the *N*-acetyl group is 4.17 Å away from one of the N η atoms and 5.41 Å away from the N ϵ atom of the guanidine group of Arg73. For the best ranked pose of *p*-nitro-phenyl galactoside the smallest distances of an oxygen atom

of the nitro group to the respective atoms of Arg73 were 4.37 Å and 5.77 Å. This binding mode is scored favorable and therefore, all poses in the resulting cluster displayed the same binding mode with minor differences occurring in their relative position to the protein surface and ligand conformation. Obviously no favorable binding mode was found for *o*-nitro-phenyl galactoside, resulting in a structurally very diverse cluster. The drastically higher diversity of structures of *o*-nitro-phenyl galactoside docked to galectin-1 in comparison to phenyl galactoside was taken as an indicator of insufficient fit into the binding site. In the best scored structure of thio-digalactoside the oxygen atom of the hydroxyl group in 3-position of the Gal residue that is not in the galactose binding site of galectin-1 is capable of forming polar interactions. The distances to both N η atoms of the guanidine group of Arg73 are 3.74 Å and 5.52 Å respectively. The distance to the N ϵ atom of Arg73 is 5.72 Å.

As stated above the planar structure of the phenyl ring in the phenyl galactosides may fit less well into the binding site of galectin-1. 2-phenyl-ethyl thio-galactoside and *p*-amino-phenyl- β -D-thio-galactoside are inhibitors of galactosyl transferase. The additional ethylene spacer of 2-phenyl-ethyl thio-galactoside should to allow a better interaction with the binding site. Additionally the influence of sulfur in the glycosidic bond on the docking results could be tested.

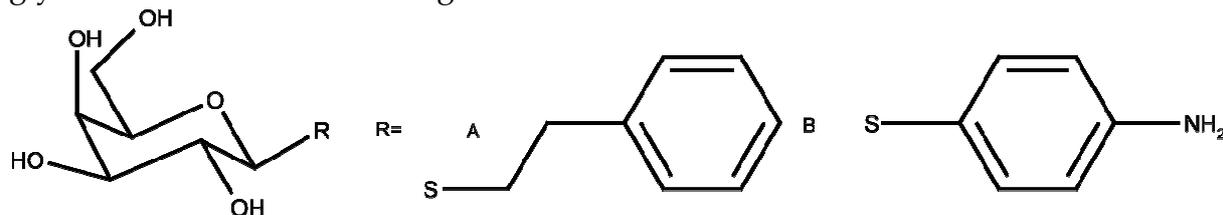


Figure 3.25: Galactosyl transferase inhibitors. A: 2-phenyl-ethyl-thio-galactoside, B: *p*-amino-phenyl- β -D-thio-galactoside

Docking experiments with these molecules did not yield consistent results. For 2-phenyl-ethyl-thio-galactoside two of twenty saved molecule poses were similar to the binding mode of LacNAc. The others were placed in a different way (Figure 3.26). It was assumed that the flexibility introduced to the molecule by the ethylene group caused the failure of the docking procedure.

The consistency of the results obtained for *p*-amino-phenyl - β -D-thio-galactoside was even worse. While the highest ranking pose still resembled that of the reference from the X-ray structure. Most of the other poses did not agree with this pose. The disappointing consistency of resulting structures for *p*-amino-phenyl-ethyl-thio-galactoside may be a result of the amino group in *para* position of the phenyl ring. When adopting an orientation similar to that of lactose in the binding site of galectin-1 this group is in vicinity to the side chain of Arg73. The side chain and amino group are charged positively under physiological conditions and in the model. Thus repulsive electrostatic effects will decrease binding energy resulting from the interaction of the ligand with the binding site.

Based on these results it was decided that NMR experiments with the galactosyl transferase inhibitors would not yield conclusive results or may not give a response due to low affinity.

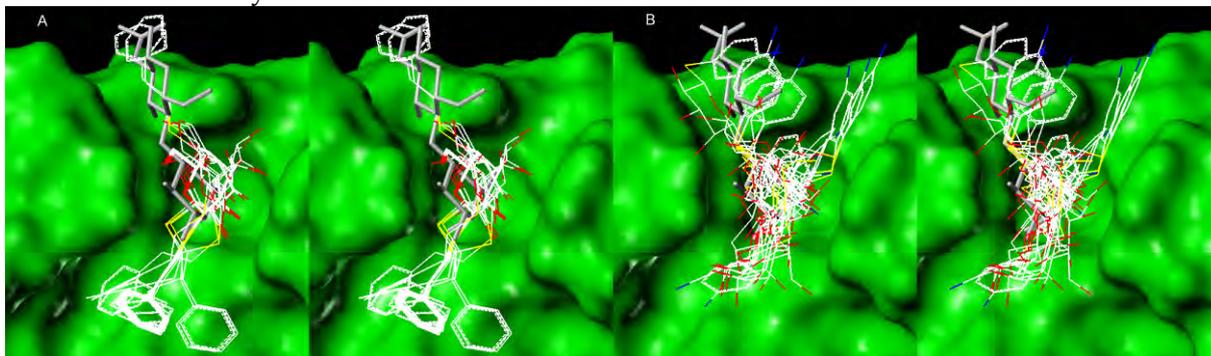


Figure 3.26 A: the result obtained for 2-phenyl-ethyl-thio-galactoside contained only two poses that were extending the phenyl ring in the same region as the GlcNAc residue of LacNAc (white, capped stick). These poses achieved rank 17 and 19. B: The results obtained for *p*-amino-phenyl thio-galactoside indicate that the docking procedure tried to avoid vicinity of the amino group to Arg73. The phenyl ring is moved away from the position where GlcNAc is positioned in the reference structure.

As a final test the result of docking the monosaccharide 1-*O*-methyl- β -D-galactoside (GalOMe) was examined. The resulting cluster for GalOMe was well defined. The two highest scoring poses were slightly shifted in comparison to the galactosyl residue of the reference.

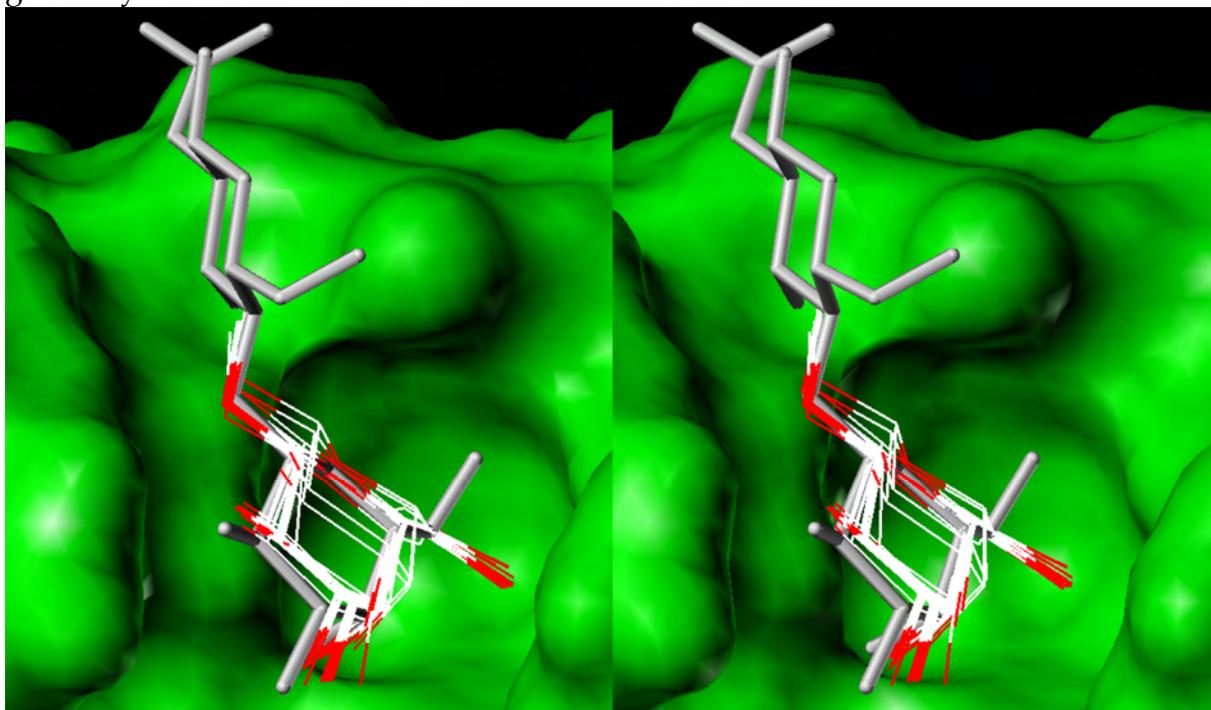


Figure 3.27 The result cluster for GalOMe is well defined. The two highest ranked poses are slightly shifted in their position.

3.5.3 Analysis of the conformation of docked *N*-acetyl lactosamine

The carbohydrates docked to galectin-1 were either extracted from the X-ray structures or were created with Sybyl. The torsion angles of the glycosidic bond and the hydroxymethyl group in 6-position of the pyranosyl ring of hexoses are key characteristics of the conformation of carbohydrates. In this work glycosidic torsion angles are defined as Φ : H1'-C1'-O-C_x, where *x* denominates the carbon that the carbohydrate is linked to and Ψ : C1'-O-C_x-H_x. Additionally the torsion of the CH₂OH-groups in 6-position is of interest and defined by the torsion angle ω : O5-C5-C6-O6. Since hydrogen atoms are usually not resolved in X-ray structures in the cited literature torsion angles of the glycosidic linkage are often given with reference to oxygen atoms. For the ligands extracted from X-ray structures torsion angles referenced to protons are given in Table 3.2. These torsion angles do not necessarily reflect the torsion angles described in the respective publications as the torsion angles may change somewhat during preparation of the structures (see experimental section, chapter 6.6.1).

Table 3.2 Comparison of the torsion angles for glycosidic bond and hydroxyl group in 6-position for the ligands prepared from X-ray structures.

ligand source	Φ [°]	ψ [°]	$\omega(\text{Glc})$ [°]	$\omega(\text{Gal})$ [°]
1a3k	51.8	16.3	297.4	59.7
1sla	41.2	10.4	321.3	72.3
1slt	63.7	336.4	299.6	60.3

In the three X-ray structures mentioned above, the ω torsion angle of the Gal residue brings the hydroxyl group in *gauche-gauche* (*gg*) conformation allowing a polar interaction with the amide of the side chain of Asn61 (Asn65 in human galectin-3). The distance from O6' to the amidic nitrogen atom of the side chain is 3Å.

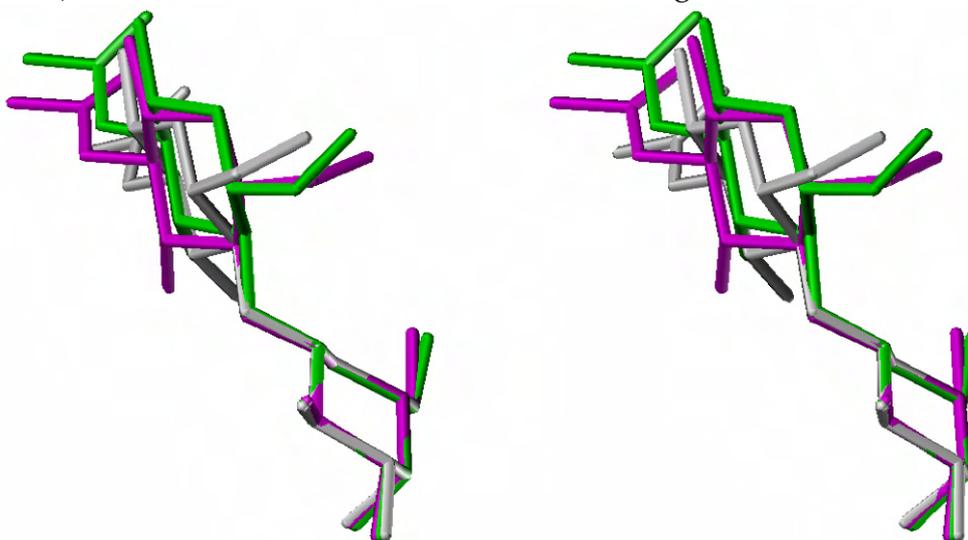


Figure 3.28 Ligand structures extracted and prepared for docking from 1slt (white), 1sla (green) and 1a3k (magenta) were matched to the pyranosyl ring of 1slt (e.g. anchor fragment). Hydrogen atoms are not displayed for clarity.

The coordinates of the carbohydrates LNB and LacNAcOEt were created using the fragments library of the program Sybyl⁸⁹. Torsion angles of the glycosidic bonds were adjusted to low energy conformations⁹⁰ prior to energy minimization. Torsion angles for the hydroxyl group in 6-position were left unchanged. The conformation of hydroxyl group of Gal was created in *gg* conformation.

Table 3.3 Torsion angles of the disaccharide ligands created from fragment library after optimization.

ligand	Φ [°]	ψ [°]	$\omega(\text{Glc})$ [°]	$\omega(\text{Gal})$ [°]
LacNAcOEt	59.2	325.9	64.0	67.2
LNB	50.0	10.0	59.7	59.8

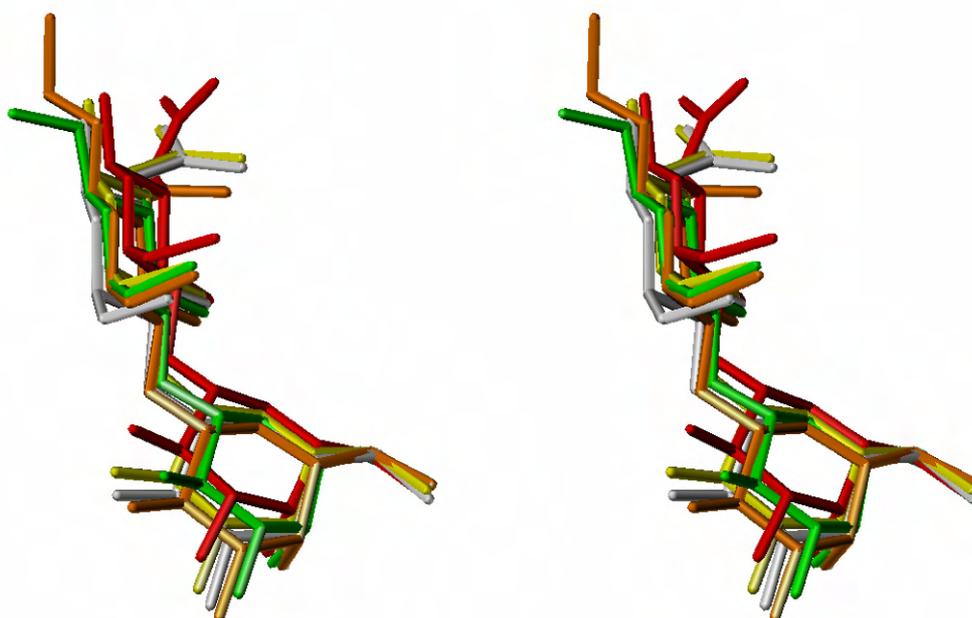


Figure 3.29 Comparison of the 1st ranked poses for LacNAc extracted from 1slt (white), 1sla (yellow), 1a3k (red), LacNAcOEt (orange) and LacOMe (green) reveals the high similarity of orientation and conformation. For all ligands OH6' is rotated into the *tg* conformation.

Table 3.4 Comparison of the glycosidic torsion angles and torsion angle of the 6-hydroxyl group of the Gal residue. All values were determined for the 1st ranked pose.

ligand	Φ [°]	ψ [°]	$\omega(\text{Gal})$ [°]
LacNAc (1slt)	44.6	6.5	189.7
LacNAc (1sla)	39.4	17.2	181.2
LacNAc (1a3k)	34.1	15.8	182.3
β -LacNAc-O-Et	61.4	-4.9	175.0
β -D-Lac-O-Me	39.2	-3.9	173.5

The structures obtained from docking experiments were exhibiting notably different torsion angles for the hydroxymethyl group of the Gal residue. The torsion angles are ranging from 173.5° to 189.7° and thus the hydroxyl group rotated into *trans-gauche* (*tg*) conformation. In this conformation the oxygen atom in 6'-position is still capable of forming a hydrogen bond with the proton attached to the amide group of the side chain of Asn61. The distance is ca. 2.4 Å. The consistent appearance of OH-6 of galactose in *gg* conformation in the crystal structures could either result from changes in the preferred conformation in the crystalline state or poor resolution and subsequent inaccurate processing of the crystallographic data. The B-factors of

C-5, C-6 and O-6 in the crystal structure 1sla are 60.6 Å², 61.3 Å² and 63.9 Å² respectively. The B-factors of the protein in 1sla are 36 Å² for the backbone and 39 Å² for the side chains. This indicates low resolution of the carbohydrate in the crystal structure. In the crystal structure 1slt the respective B-factors are 21.4 Å², 24.7 Å² and 26.4 Å², in the crystal structure 1a3k the respective B-factors are 20.5 Å², 23.5 Å² and 19.3 Å², which is the same magnitude as determined for the respective protein indicating higher reliability of the data. In 1a3k and 1slt the hydroxymethyl group of galactose is also in *gg* conformation.

The *tg* conformation consistently found in the docking experiments represents a preferred conformation found in solution⁹⁰. For the poses proposed by DOCK these conformations achieve a better energy score. It remained unclear whether this change of conformation was a result of beneficial contribution of internal energy of the ligand or improved receptor ligand interaction. No indications were found that the ligand from the X-ray structure is placing the hydroxyl group in 6'-position of *N*-acetyl lactosamine unfavorably close to the receptor. Thus a change of conformation due to unnaturally close VDW contacts could be ruled out. The fact that DOCK did change the torsion angles and scored the conformations with *tg* conformation better than the *gg* conformations that were used as input conformations underscores the importance of treating ligands as flexible molecules. This allows the docking software to optimize ligand conformations during the docking procedure. This may prove especially effective when the input conformation is taken from conformer generators such as Corina and not from crystal structures. Conformer generators of conformer libraries are sources of diverse conformers of novel ligands studied in computational screening procedures. Erickson *et al.* also determined that consideration of ligand flexibility significantly improved docking results²⁷.

The binding mode of the carbohydrate ligand in the binding site is also characterized by the angle θ , at which the normal planes of the galactose ring and the ring system of Trp68 intersect and the centroid distance R of these ring systems. The relevance of these characteristics is based on the interaction of carbohydrates with aromatic residues in the binding site of lectins⁹¹⁻⁹⁵. Recently several studies underscored the significance of hydrophobic interactions between carbohydrates, namely galactose, fucose and glucose, and aromatic residues of the receptor by quantum mechanical calculations of simplified model systems consisting of monosaccharide and an aromatic ring system⁹⁶⁻⁹⁸. M. Fernandez-Alonso *et al.* were also able to demonstrate direct interactions of galactose and phenol via changes of the chemical shifts of galactose resonances⁹⁸.

The normal planes and centroids of the galactose residue were defined using the heavy atoms of the pyranosyl ring system of the galactose residue. The normal plane definition and centroid of Trp68 were defined from the heavy atoms of the tryptophane's indole ring. A value of $\theta=180^\circ$ means parallel alignment of the normal planes.

The results for θ and R are given in Table 3.5. The results from the docking experiments displayed high consistency within the different lactose derivatives docked to galectin-1. The results were in good accordance with the findings from the X-ray structure. The distance of the *N*-acetyl lactosamine centroid, extracted from the X-ray structure 1a3k, was significantly closer to the centroid of Trp68. This indicated a shift of the ligand towards the indole ring system. A similar shift of lesser extent was also observed for the pose of methyl- β -D-lactoside.

Table 3.5 Comparison of the θ angles and centroid distances in the X-ray structure 1sla and complexes from docking experiments. All values from the docking experiments were determined for the 1st ranked poses.

ligand	θ [°]	R [Å]
LacNAc (1sla, X-ray reference)	133.2	5.46
LacNAc (1sla, docked)	132.8	5.20
LacNAc (1slt)	132.6	5.07
LacNAc (1a3k)	132.6	4.17
β -D-LacNAc-O-Et	130.5	5.18
β -D-Lac-O-Me	137.6	4.79

3.5.4 Docking of lacto-*N*-biose

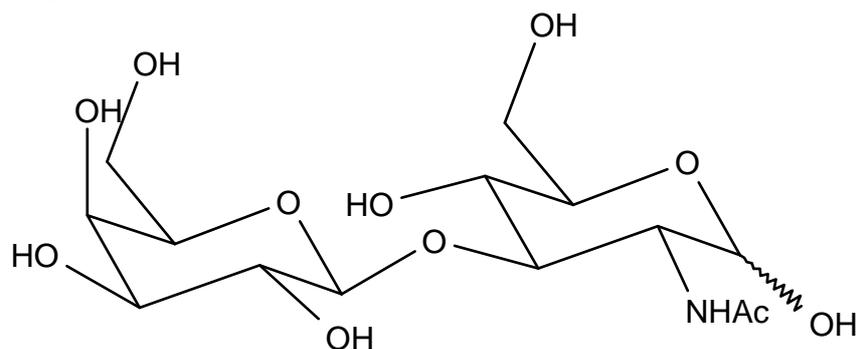


Figure 3.30: Structure of lacto-*N*-biose. The disaccharide is linked from 1' to 3 position in β -configuration.

Lacto-*N*-biose, (LacNAc type I, LNB) is a known ligand of galectins. The affinity of LNB has been determined to be 111 μ M by ITC⁸¹. Ahmed *et al.* report a 2.7 fold increase of the relative activity of Gal β 1-3GlcNAc α -OMe compared to lactose. They determined the relative activity of LacNAc to be 10.2 fold higher. To date no X-ray structure has been published of this ligand in contact with galectin-1. Y. Bourne *et al.* were proposing a binding mode for LNB based on their structure 1sla⁷⁷. LacNAc (type II) is presenting the hydroxyl group in 3-position of the GlcNAc residue, so that it is capable of polar interactions with Arg48, Glu71 and Arg73. In LNB this interaction is eliminated by the different glycosidic linkage. In their proposal Bourne *et al.* assumed torsion angles of $\Phi = -65^\circ$ and $\Psi = 140^\circ$ for the glycosidic linkage (Φ : O5'-C1'-O-C3; Ψ : C1'-O-C3-C4) corresponding to $\Phi = 58.1^\circ$ and $\Psi = 25.1^\circ$ when referenced to the proton torsion angles described in chapter 3.5.3⁷⁷. These torsion angles were assumed to be in a low energy conformation⁹⁰. In this conformation the GlcNAc residue is turned almost 180° along the O1-C3 bond. If the Gal residue remained in the same pose, the hydroxyl group in 4-position of the GlcNAc residue would be placed close to the position occupied by the 3-hydroxyl group of LacNAc. Bourne *et*

al. assumed that the interactions with Arg48, Glu71 and Arg73 could be adopted by the hydroxyl group in 4-position of the GlcNAc residue in LNB⁷⁷.

A model of LNB was created from the fragment library of Sybyl. Torsion angles for the glycosidic bonds were applied resulting in a low energy starting structure. The torsion angles were referenced to the protons of the 1-position of Gal and 3-position of GlcNAc, ϕ (H1'-C1'-O-C3) was set to 50° and ψ (C1'-O-C3-H3) was set to 10° .

Docking of LNB to galectin-1 yielded results consisting of different orientations of the ligand in the binding site and different ligand conformations. The majority of results consisted of a binding motif where the disaccharide was rotated along the normal plane of the Gal residue. This caused the 6'OH group to come out of the binding pocket and the GlcNAc residue was placed in a binding pocket defined by Trp68, His44, Cys2 and Ala1. 14 poses were in this cluster.

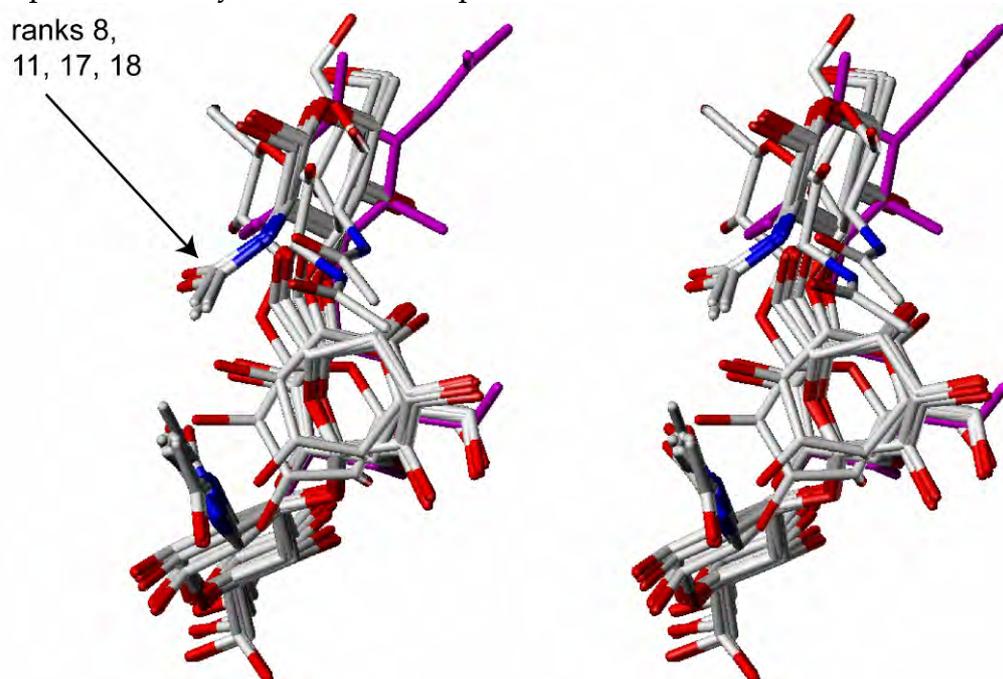


Figure 3.31 Comparison of the results from docking LNB to galectin-1 with the orientation of the ligand in the X-ray structure (magenta). Four poses, indicated by the arrow, were placed in a binding mode as proposed by Bourne *et al.* Two more poses (ranked 19th and 20th) were in a similar binding mode but had different Ψ torsion angles. These poses project their *N*-acetyl group out of the picture plane towards the observer.

Four of the poses proposed by DOCK were oriented as proposed by Bourne *et al.* These poses attained ranks 8, 11, 17 and 18. The torsion angles resulted in $\Phi=49.3$ and $\Psi=37.6^\circ$, referenced to the protons. The last two poses of the results displayed the expected orientation for the Gal residue but achieved unusual Ψ torsion angles of 73.3° and 61.5° . The torsion angles of the first pose were $\Phi=49.3^\circ$ and $\Psi=37.6^\circ$. With the exception of the lowest ranked poses the conformation of the glycosidic bond was in agreement with the low energy values of $\Phi=50^\circ$ and $\Psi=10^\circ$.

The majority of the resulting binding modes were in disagreement with the proposal of Bourne *et al.* To check whether the proposal of Bourne *et al.* was wrong or the docking protocol failed to give the expected result the outcome of the docking experiment was checked against the results from STD NMR experiments.

3.6 Docking constraints obtained from STD NMR experiments

Epitope mapping of ligands bound to a receptor obtained from STD NMR experiments is a valuable source of constraints for the selection of poses proposed by docking programs.

In the complex of ligand and receptor the magnetization applied to the receptor is transferred to the ligand. Transfer is accomplished by dipolar interactions of the protons of the receptor and protons of the ligand. After dissociation of the complex the resonance signals acquired for the saturated ligand protons are weaker. The amount of magnetization transferred and thus the attenuation of the resonance signal for a given ligand proton is depending on the amount of receptor protons in vicinity to this proton and their distances to the ligand proton. The saturation of the resonance is proportional to Σr^{-6} for all distances of receptor protons to the respective ligand proton.

This correlation of observed saturation of a resonance signal with the distance of the proton to other protons is similar to the relation found for the Nuclear Overhauser Effect (NOE) of two distinct protons which is used in 3D structure determination by NMR. Depending on rigidity of the molecule studied and the capabilities of the NMR spectrometer used, proton-proton distances may be determined in the range from ~ 1.75 to $\sim 5.0 \text{ \AA}$ ⁹⁹.

If the 3D structure of the ligand bound to the receptor is available the distances for each ligand proton to protons within the range to transfer an NOE can be determined. The sum of the reciprocal distances to the power of six should be proportional to the STD effect acquired in the STD spectrum.

To test the validity of this calculation of virtual STD intensities the distances of protons from *N*-acetyl lactosamine to protons in galectin-1 were determined. The structure 1sla of Bourne *et al.* was taken as a template⁷⁷. As the reference STD NMR data from the spectrum of β -D-1O-ethyl lactosamine (LacNAcOEt) was taken. Since the STD spectrum was acquired in D₂O as solvent all exchangeable atoms of ligand and receptor were treated as deuterium atoms. These deuterium atoms were not used in the distance calculation, since they cannot transfer saturation to the ligand.

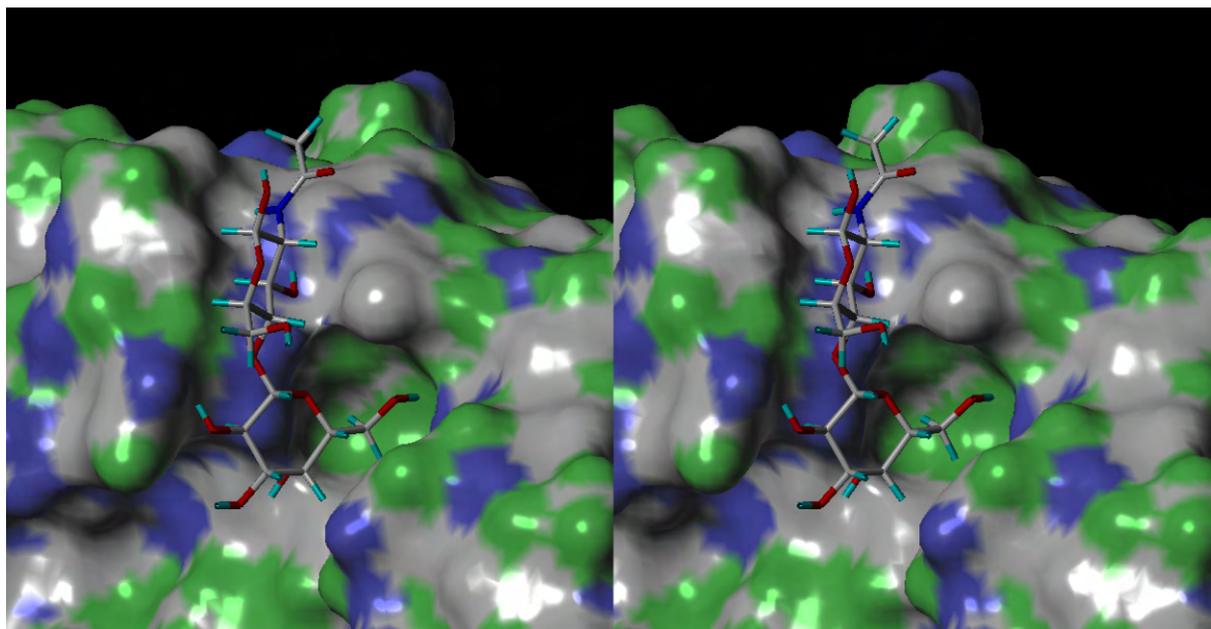


Figure 3.32: LacNAc residue and protein surface from 1sla: protons of the protein that are exchanged to deuterium are colored in blue, protons in green are capable of transferring magnetization. On the ligand only protons bound to carbon atoms are detected in the NMR spectrum.

Only protons attached to carbon in the ligand molecule are detected in the ^1H -STD NMR spectrum. Thus only distances for these protons were calculated. The cutoff distance was set to 6 Å. It was assumed that an array of protons more than 5 Å away from the ligand may still transfer detectable magnetization due to additive behavior.

A program was developed, that is capable of processing result files from DOCK and the receptor files. The reciprocal of the distances to the power of six were calculated and added up for all individual ligand protons and the result saved in a separate file.

3.6.1 Automated calculation of STD effects for docking results

The program DOCK is handling ligand files in the mol2¹⁰⁰ file format. The output is a multi-mol2 file containing all poses in a single file. In order to calculate the distances for all individual ligand protons of all poses the coordinates of the protons must be accessible.

The mol2 file format offers two ways to address individual atoms called 'static set' and 'dynamic set'. The members of the *dynamic set* are defined by a rule. The generic rule applicable in this thesis would be "hydrogen atom bound to any carbon atom". The advantage of the *dynamic set* is its flexibility because the fulfillment of the respective rule can be checked continuously. The disadvantage is that fulfillment of the rule has to be checked for each atom in each structure that is investigated. Atom coordinates can only be retrieved after checking for compliance with the rule defining the set. In addition, handling of dynamic sets requires that the respective software is capable to interpret the rule defining the set. The respective capabilities have to be implemented in the program.

Static sets are defined by a list of atoms. The atoms are identified by their identification numbers, referred to as atom IDs. The drawback of this definition is that all the atoms belonging to the respective set have to be predefined by the user. The advantage of *static sets* is based in the availability of methods to directly retrieve the atom coordinates for a given atom ID. Once the desired atoms have been defined their coordinates are accessible in each structure. Furthermore a list of atom IDs is easily handled with simple programming methods.

For the distance calculation a program was implemented in C++ that could handle mol2 files and is independent from third party software. As basis of the software the code of the OpenBabel (OB)^{101,102 103} project in the version 1.100 was chosen. The OpenBabel code supplied the functions to access atom coordinates from mol2 files and read single molecules from multi-mol2 files. The OpenBabel code did not supply functions to access set information of the mol2 files. This function was implemented in the analysis program.

The atom ID information was stored in static sets. This was done because external evaluation of the selection rules for the dynamic set could not be implemented in an effective way. The static set supplied a list of atom IDs to be used as direct input to functions of OpenBabel to access the respective coordinates of the atoms (random access).

The docking results were read into a data container consisting of OB-molecule objects (C++ vector). For the calculation of distances and calculation of virtual STD intensities the docked molecule structures were processed in a serial manner. Thus the results could be ordered according to the rank the respective molecule achieved in the docking experiment. Virtual STD intensities were calculated for all distances between ligand and receptor protons within a cutoff distance. The cutoff distance was implemented as a user definable parameter.

Additionally a minimum distance was implemented. DOCK can be allowed to dock molecules with a given number of ligand atoms closer to the receptor atoms as the sum of the respective van der Waals radii. This methodology can implement a rough model of receptor flexibility and allows the orientation search algorithm to create poses that can be optimized by the minimization step in the docking procedure. These close distances do not occur in the bound state of the ligand under the conditions in the NMR experiments. Close contacts in the model would result in disproportional high virtual STD intensities. To compensate for these unnaturally close contacts the minimum distance was implemented. Distances falling below this value would be set to the minimum distance. As stated above the maximum distance between protons contributing to the STD effect was set to 6Å and the minimum distance to 1.8Å. The results of the analysis were stored in a file format that allows the import of the data into office software for spreadsheet analysis. The program processed mol2 files at considerable speed. The 20 docked poses of the monosaccharide β -D-O-methyl-galactose were processed in less than 0.5s.

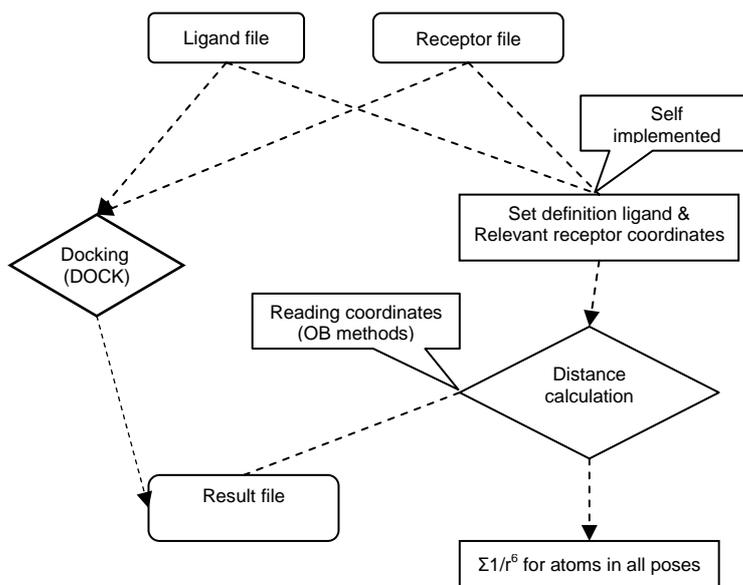


Figure 3.33: Flowchart of the analysis of docking results.

3.6.2 Comparison of calculated and experimentally determined STD effects

The X-ray structure 1sla was prepared for data analysis. After removal of crystallographic water molecules and ions of the buffer, hydrogen atoms were attached to the protein. The ligand was truncated to the lactosamine residue in the binding site by deleting atoms of the remaining carbohydrate residues (substructures 403, 404, 405, 411, 412 and 413 in the pdb file). Then hydrogen atoms were attached. Both structures were allowed to structurally relax by an energy minimization. Distances were determined for all non exchanging protons of the ligand to all non exchanging protons of the protein within 6 Å using the Sybyl⁸⁹ software. For distances less or equal 6 Å, the reciprocal of the distance to the power of six was calculated and added up for each ligand proton. The final sum for each ligand proton was referred to as the virtual or calculated STD effect.

When comparing virtual STD effects with STD intensities from NMR experiments it became clear that absolute STD values were directly comparable (Table 3.6). This may be due to the fact, that the transfer of magnetization in the bound state is also proportional to the cross-relaxation rate of the complex, and magnetization is lost once the ligand dissociates into the solvent depending on the T_1 relaxation time. Additionally influences from background noise and internal motion of the complex must be considered.

In Figure 3.34 it is obvious that relative virtual STD effect and relative STD effect from NMR determined by intensity or integration (p17) did not match exactly. But it could be seen, that especially on the Gal residue the distribution of the effects is following the same pattern. Thus it was assumed that the comparison of virtual STD

intensities with STD intensities from STD NMR experiments should give a lead to the determination of the binding mode that is obtained in solution.

Table 3.6: Comparison of the absolute and relative STD values determined by STD NMR and calculated from distances from ligand proton to receptor protons within 6 Å.

		abs. STD[%]			relative STD[%]		
		Intensities	Integrals	calculated	Intensities	Integrals	calculated
Gal	H1	0.178	0.033	0.00020	59.33	10.91	2.80
	H2	0.180	0.055	0.00171	60.00	18.10	23.81
	H3	0.173	0.082	0.00046	57.67	27.08	6.38
	H4	0.217	0.118	0.00210	72.33	38.94	29.31
	H5	0.150	0.089	0.00083	50.00	29.31	11.64
	H6a	0.293	0.303	0.00716	97.67	100.00	100.00
	H6b	0.300	0.272	0.00574	100.00	89.82	80.06
GlcNAc	H1	0.000	0.058	0.00052	0.00	19.10	7.30
	H2	0.176	0.173	0.00017	58.67	57.09	2.38
	H3	0.000	0.083	0.00664	0.00	27.48	92.70
	H4	0.150	0.089	0.00029	50.00	29.31	4.04
	H5	0.156	0.089	0.00135	52.00	29.32	18.91
	H6a	0.110	0.141	0.00013	36.67	46.53	1.84
	H6b	0.173	0.162	0.00008	57.67	53.56	1.10
	NHAc	0.094	0.082	0.00019	31.33	27.23	2.64
Ethyl	H1	0.082	0.042	--	27.40	13.74	--
	H2	0.213	0.212	--	71.00	70.18	--

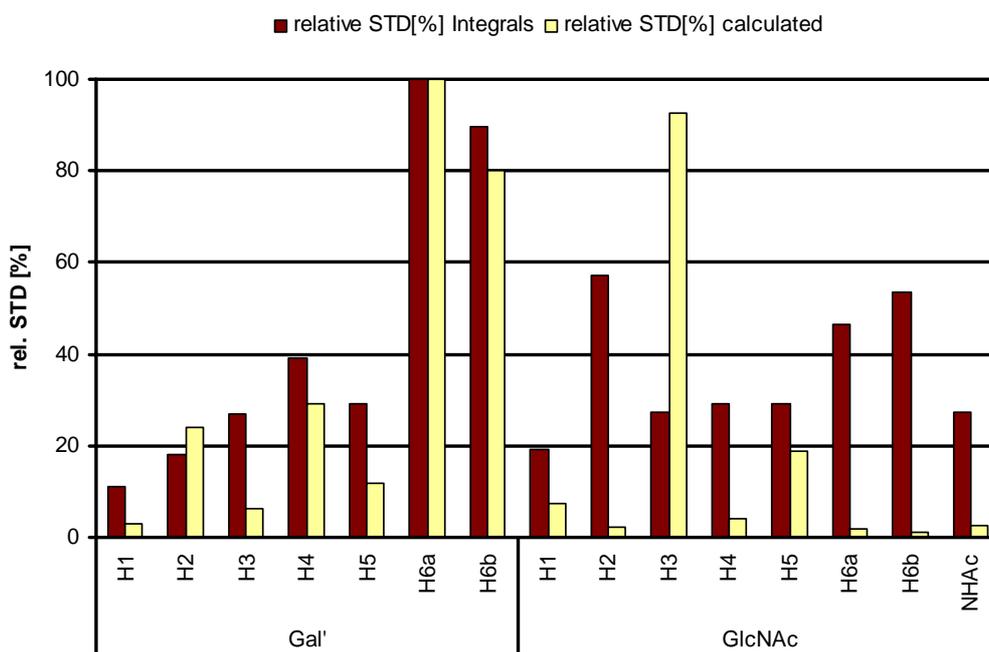


Figure 3.34: Comparison of relative STD values determined by intensity, integration and virtual STD by distance. The comparison of the STD by integration and distance reveals the broad similarities in the pattern for several protons, especially on the Gal residue.

The most outstanding discrepancy in calculated and experimentally determined STD values is detected for H3 of the GlcNAc residue. Experimentally determined relative STD effects were at 27% for LacNAcOEt and 49% in LacOMe. Based on the distance the effect should be ~90%.

One reason for this disagreement could be that the position of the octasaccharide ligand in structure 1sla is shifted closer to the protein or that this position is obtained as a result of the crystallization process. Another reason may be a conformational change in the structure of galectin-1 from solution to the crystalline state. Additionally, the resolution of the crystal structure may be too low to allow accurate determination of the protein conformation.

The high STD effect determined in the calculation is a result of close proximity between H3 of GlcNAc and the closest β -proton of His52 in galectin-1 over 2.4 Å as obtained from the crystal data. The B-factors of His52 in the crystal structure 1sla were ranging from 53.3 to 57.7 Å². The crystal structure has a resolution of 2.45 Å⁷⁷. In the crystal structure 1slt the distance of H3 to one of the protons in the β -position of His52 is 2.7 Å. The B-factors of the His52 in 1slt atoms are ranging from 25.0 to 49.1 Å² at a resolution of 1.90 Å⁷⁸. The structures of 1sla and 1slt were matched to determine the variations of the crystal structures. The heavy atom coordinates of His52 and the essential amino acid in the CRD of all galectins Trp68 were used for the matching procedure and the structure 1sla was taken as reference structure. The structure of 1slt could be matched to the reference coordinates with a rmsd of 0.44 Å. For comparison the crystal structure of human galectin-1, 1gzw⁷⁴, was matched to the reference coordinates with a rmsd of 0.60 Å. In 1gzw His52 displayed B-factors of 25.3 to 31.9 Å² at a resolution of 1.7 Å. All galectin-1 structures displayed close contacts from His52 to H3 of the GlcNAc residue. The B-factors are indicating some flexibility of the His52 residue¹⁰⁴. The same conclusion can be drawn for the GlcNAc residue in *N*-acetyl lactosamine. In the crystal structure 1sla the B-factors of the GlcNAc residue are in the range of 62 Å², in 1slt (GlcNAc) and 1gzw (Glc) the B-factors are in the range of 35 Å² and 33 Å² respectively. The structural differences and the resulting differences in the calculated distances between protons of the receptor and the ligand can have huge effects on the STD intensities calculated from these structures. The calculated STD effect for H3 of *N*-acetyl lactosamine in 1sla is $5.2 \cdot 10^{-3}$ (2.4⁻⁶). The same STD intensity calculated from 1slt is $2.6 \cdot 10^{-3}$ (2.7⁻⁶). This could explain the difference in the calculated and experimentally determined STD intensities.

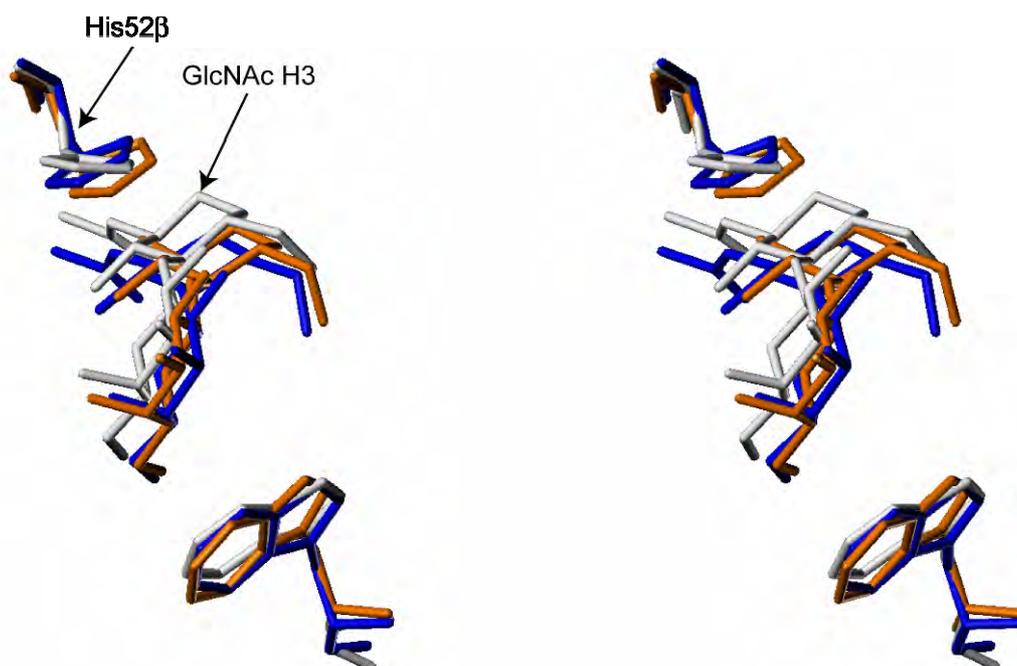


Figure 3.35: The comparison of the matched structures 1sla (white), 1slt (blue) and 1gzw (orange) reveals the structural similarities of the galectin-1 structures. Only heavy atoms of His52 and Trp68 of the protein are shown. The β -position of His52 and the position of H3 of GlcNAc in 1sla are indicated by arrows.

Another explanation could be the decomposition of magnetization transferred to H3 of LacNAc by fast relaxation processes. The T_1 relaxation times of the unbound LacNAc molecule can be determined by the inversion recovery NMR experiment. Using this experiment the T_1 relaxation time of H3 of LacNAc was determined to be 940 ms. The largest relaxation time determined was 3.59 s for H2'. The relaxation time of H2 was determined to be 910 ms. The relaxation time of H4 could not be determined unambiguously because of signal overlap with H5'. The overlapped signals yielded a T_1 time of 1.45 s. These relaxation times did not give rise to assumptions of disproportional fast decomposition of magnetization for the H3 proton. The strongest STD intensities were determined for the protons in 6'-position. The relaxation times for these protons were determined to be 563 ms and 681 ms respectively.

The discrepancy in the calculated and experimentally determined STD effects for H2 is most likely to be a result of the disadvantageous S/N ratio in the STD NMR (chapter 3.4) causing difficulties when integrating the region of the respective resonances.

3.6.3 Comparison of calculated STD for docked poses with results from STD NMR

STD NMR experiments conducted with β -D-O-ethyl-lactosamine (LacNAcOEt) yielded a well defined epitope. This was already used in the comparison of STD epitope and virtual STD effect in Figure 3.34. LacNAcOEt was also docked to galectin-1. Thus it was decided to compare the epitope determined by STD NMR with the calculated STD epitope for the results obtained using LacNAcOEt as ligand.

The relative STD values calculated from proton distances for LacNAcOEt showed strong correlation with the values determined by STD NMR for the Gal residue. But instead of H6' the proton in 3-position was evaluated to have the highest STD effect in most poses. This was mostly attributed to the distance of 2.26 Å to one of the β -protons of His52. Strong calculated intensity of H3 was already determined for the crystal structure but the magnitude was in the range of the effect determined for H6'-position. Most of the poses from docking LacNAcOEt achieved between 30 to 40% STD effect for H6'. One pose (rank 16) that exhibited the strongest interaction on H5, which was also achieved by close contact with a β -proton of His52 over 2.29 Å. Only one of the top 20 poses exhibited the highest effect on H6'. This pose was ranked 17th.

As mentioned before, the capabilities of docking software may be exceeded by the number of rotatable bonds in the ligand molecule. In comparison with the crystal structures of LacNAc in complex with galectin-1 the GlcNAc residue of LacNAcOEt is rotated along the glycosidic bond as a result of altered torsion angles. The Gal residue of LacNAcOEt is placed in a way, that the H6' protons are placed further away from the protein surface. These differences in the distances become strongly emphasized when the STD effect is calculated from the distances due to the proportionality to r^{-6} .

The LacNAc disaccharide extracted from the X-ray structure 1slt yielded a result that was in good agreement with the binding mode of the ligand in the crystal structure after docking to galectin-1 from 1sla (Figure 3.21, panel B). The structures obtained from docking LacNAc from 1slt to galectin-1 from 1sla were used to calculate the STD effects as described before. In Figure 3.37 the experimentally determined relative STD percent are compared with the calculated relative STD percent for the best three poses and the average value determined for all twenty poses saved after docking.

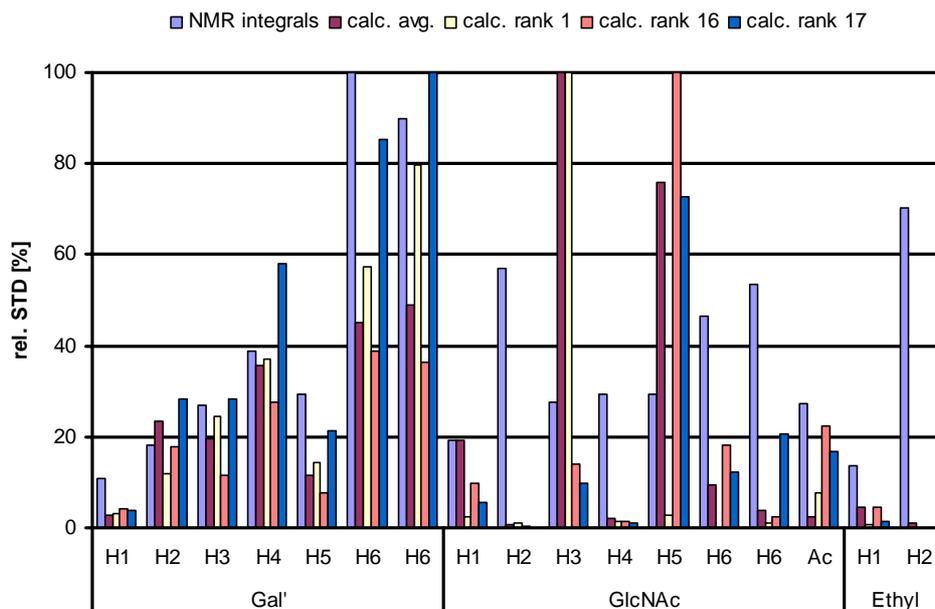


Figure 3.36: Comparison of the STD NMR intensities and the calculated STD intensities for LacNAcOEt.

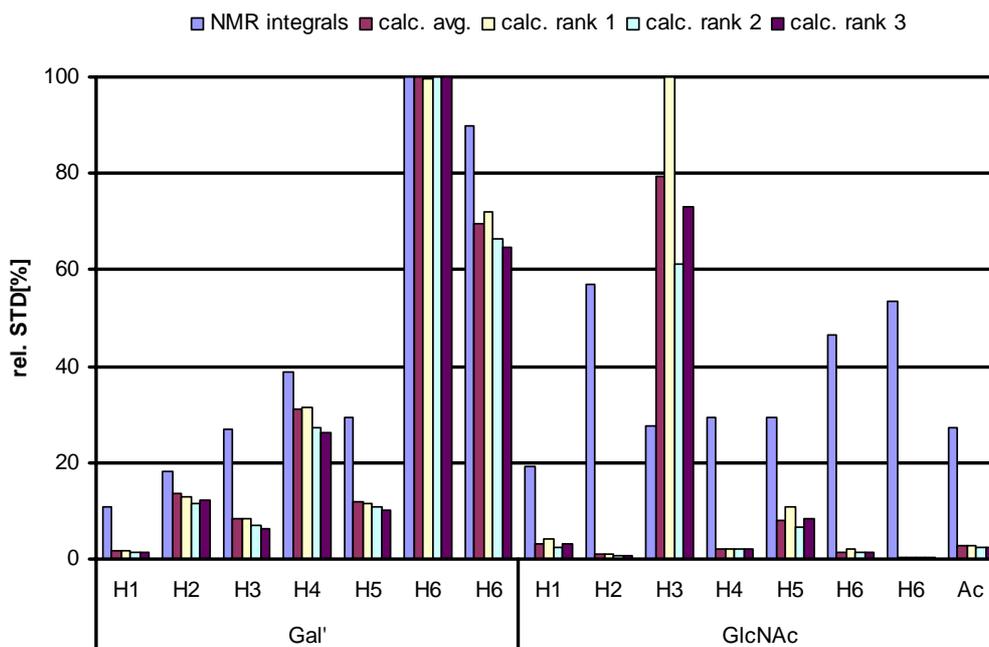


Figure 3.37: Comparison of the relative STD values from NMR experiments using LacNAcOEt and galectin-1 determined by signal integration with the calculated STD effects of the average (avg.) of 20 docked poses and the top 3 poses of LacNAc extracted from X-ray structure 1slt. High similarity of calculated and experimentally determined STD values for protons of the Gal' residue is obvious.

The calculated relative STD values were in agreement with the results from the STD NMR experiments. The best scored pose (rank 1) showed the highest calculated STD effect on H3 but H6' was calculated to have 99.8% relative STD effect. Seven out of the twenty saved docking results were placing H3 in a position resulting in the highest calculated STD effect. In all of these poses at least one of the H6' protons achieved relative STD values over 80%. For all other binding modes the highest

virtual STD effect was determined for H6'. The relative STD values from distance calculation and STD NMR determined for the Gal residue were consistent in relation to each other.

Although the experimentally determined relative STD effect exceeded the calculated STD for all atoms, except H3, it could be seen that the pattern of STD intensity distribution from NMR was matched by the calculated STD effect especially for the Gal residue. The results obtained from docked poses were comparable to the results already obtained from the bound ligand structure in the X-ray structure. Thus it was concluded that the comparison of relative STD values determined from STD NMR experiments can be compared to the calculated relative STD values determined from the distance of receptor protons to the protons of the ligand. The comparison of calculated STD effects of differently docked poses with the epitope derived from STD NMR should give indications on the validity of the respective pose proposed by the docking procedure.

It remained unclear why the calculated STD effect for H3 of the GlcNAc residue always exceeded the STD effect determined by STD NMR. The order of the experimentally determined STD effect must be a result of saturation transfer effectiveness. It is possible that protons of the receptor in vicinity to H3 are not capable of transferring their magnetization effectively. This could be a result of molecular flexibility causing inefficient magnetization transfer. This assumption is endorsed by the B-factors of the His52 residue. In the crystal structure 1sla and in docking experiments the closest contacts from H3 to the receptor were a result of contacts to one of β -protons of His52. As described in chapter 3.6.2 (p49) and Figure 3.35 this structural feature is likely to be a result of flaws in the crystal structure.

The comparison of STD values determined by STD NMR and calculated from distances was also carried out for the ligand phenyl- β -D-galactoside (PheGal). The docking result featured a cluster of molecules with the same alignment in the binding site of galectin-1. The binding modes found were similar to that of the galactose residue in the crystal structures 1slt⁷⁸ and 1sla⁷⁷. In the STD NMR experiments the strongest STD effects were determined for protons in ortho position of the phenyl ring and H4 and H6 position using the intensities of the signals in the difference spectrum. The signals of the meta protons in the phenyl ring overlapped with the proton in para position. It was assumed that based on the binding mode proposals the effect was mainly to be attributed to the protons in meta position. However the calculated STD effects determined by distance analysis of the docked complexes did not give indications to the significant STD effect determined for the protons in *ortho* position.

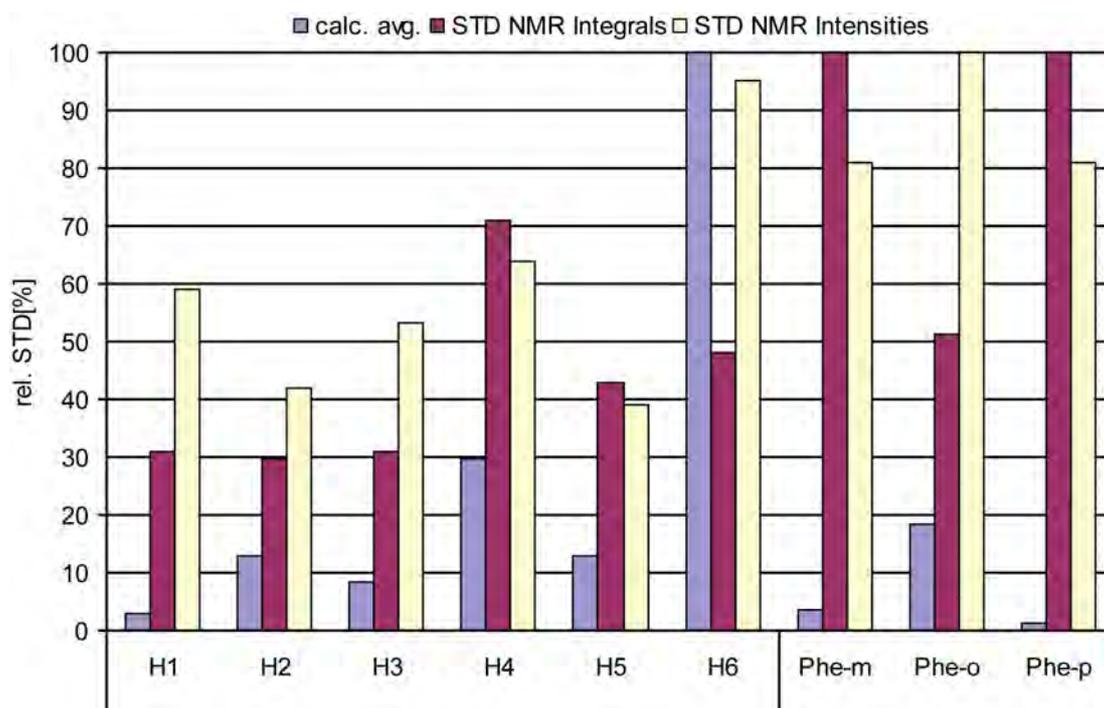


Figure 3.38: Comparison of relative STD values from STD NMR spectroscopy determined by integration and signal intensity with average STD values calculated for all 20 poses of PheGal docked to galectin-1.

Direct comparison of the STD values determined by STD NMR experiment and virtual STD values determined from distance revealed some similarities. Again neither absolute nor relative values determined for the STD effect matched exactly. But it became obvious that for the protons of the carbohydrate residue the strongest intensities were determined for H6, H4 and H5. When using integrals for the determination of the STD effect the strongest effect was determined for H4. This could be accounted for by the significant overlap of the signals of H6, H5 and H2 in the 1D ^1H spectrum. The signal of H4 is separated from all other resonances and was easy to integrate. As already mentioned in chapter 3.3.4 calculated STD effects for aromatic protons in *ortho* position of the phenyl ring are the strongest due to vicinity to the protein. The STD effects determined for the phenyl protons from STD NMR experiments are likely to be affected by the disadvantageous S/N ratio as discussed in chapter 3.4.

3.6.4 Selection of docked LNB poses based on STD NMR

The previously developed protocol was adopted to select poses of lacto-*N*-biose (LNB) describing the binding mode of LNB in complex with galectin-1. The result obtained from docking LNB docked to galectin-1 consisted of three structurally different binding modes as shown in Figure 3.31. Fourteen out of twenty structures were not in agreement with the binding mode of lactose derivatives determined by X-ray experiments and the hypothesis made on the basis of crystallographic results⁷⁷. Four of the docked poses were in agreement with the binding mode of *N*-acetyl lactosamine and obtained low energy conformations of the glycosidic bond. The best ranked pose in agreement with the hypothetical binding mode achieved rank 8. Two

poses obtained from docking LNB to galectin-1 were omitted due to energetically unfavorable torsion angles. These two poses were last in the DOCK ranking.

In the STD NMR experiments the highest intensities were determined for the protons in H6'-position (Figure 3.12). High intensities were also determined for the H4' and H5' protons. Based on the epitope determined for the galactose residue in LNB it was assumed that the binding mode of LNB was similar to that of LacNAc. The changes of the STD NMR epitope determined for the protons of the GlcNAc residue compared to LacNAc were attributed to the different linkage and resulting torsion angles.

Relative STD values were calculated for all 20 poses saved from 2000 orientations tried during the docking procedure. In Figure 3.39 the relative STD values determined by calculation and STD NMR are compared to each other.

The pose that achieved rank 1 was chosen as a representative of the 14 poses that did not agree with the proposed binding mode. For this pose the highest STD effects were calculated for H6 and H4 followed by H2. The strongest effect on the galactose residue were calculated for H2' and H3', not exceeding 15% relative STD effect. The poses ranked 8th, 11th, 17th and 18th were used to determine the calculated STD effect for the binding mode in agreement with the hypothesis for the binding mode.

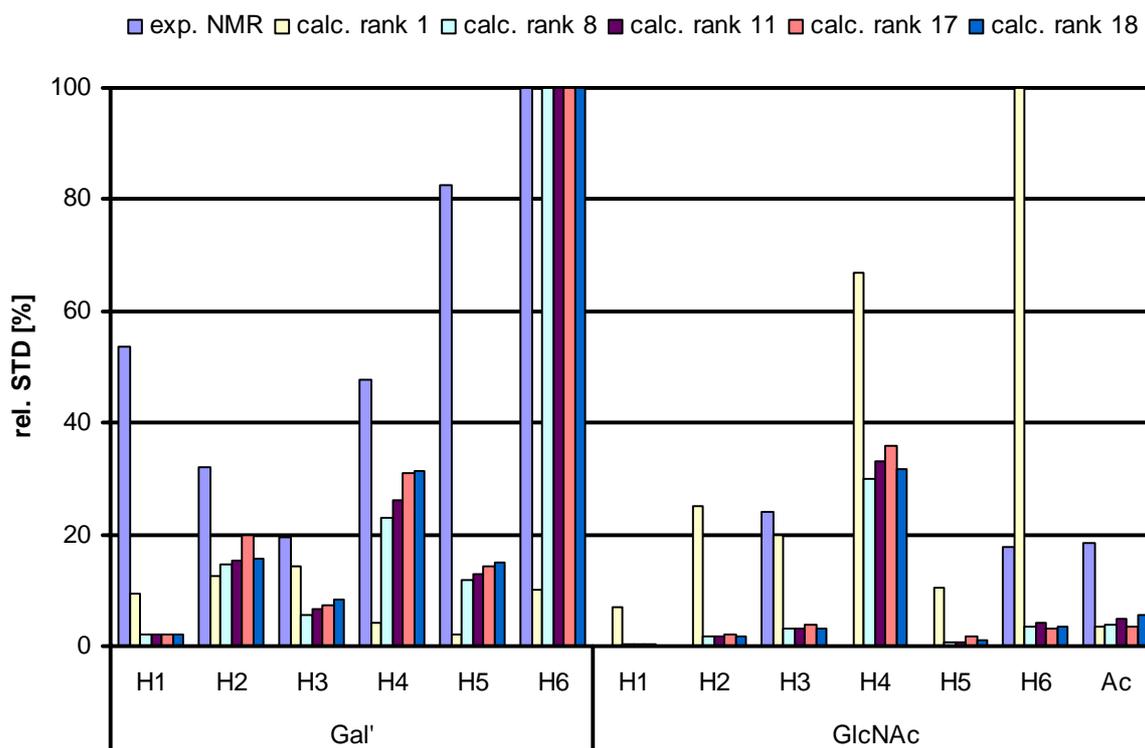


Figure 3.39: Comparison of virtual relative STD intensities obtained from proton-proton distances in the docked complexes and relative STD intensities acquired from STD NMR experiments with LNB and galectin-1. For clarity the experimental STD NMR intensities are displayed as average of both anomers.

The calculated STD intensities determined for the docked poses in agreement with the hypothesis (rank 8, 11, 17 and 18) were clearly differing from the calculated STD

intensities of the best ranked pose. These four poses were in alignment with the orientation of *N*-acetyl lactosamine in the reference crystal structure (Figure 3.41). The pattern of calculated STD intensities determined for the Gal residue in LNB matched the pattern of the Gal residue in LacNAc (Figure 3.37). For the Gal residue the strongest calculated STD intensities were determined for the 6'-position and 4'-position of LNB. The calculated STD intensities determined for the GlcNAc residue differed significantly. The strongest intensity for the GlcNAc residue was determined for the H4-position. The order of magnitude is similar to that of H4'-position. This is in agreement with the proposal of Bourne *et al.* that in a low energy conformation of the LNB disaccharide H4 is occupying the space that is occupied by H3 in the LacNAc disaccharide⁷⁷. As described in chapter 3.6.2 the calculated STD effect arises from a close contact of H4 with His52. No STD effect could be determined by STD NMR experiments.

In contrast the first ranked pose displayed the strongest calculated STD intensity on the H6-position of the GlcNAc residue followed by the H4-position (67%) and H2-position (25%) of the GlcNAc residue. For the first ranked pose calculated relative STD intensities did not exceed 20% for protons of the Gal residue. The strong interactions calculated for the H6-position are a result of close vicinity of both H6 protons with a methyl group of Leu31 of galectin-1. The close distance of these protons allows four contacts over less than 3.0 Å and two more contacts at 3.4 Å and 4.0 Å respectively. The strong intensity of H4 is a result of a close contact with H ϵ of His44 of galectin-1 with distance of 2.4 Å.

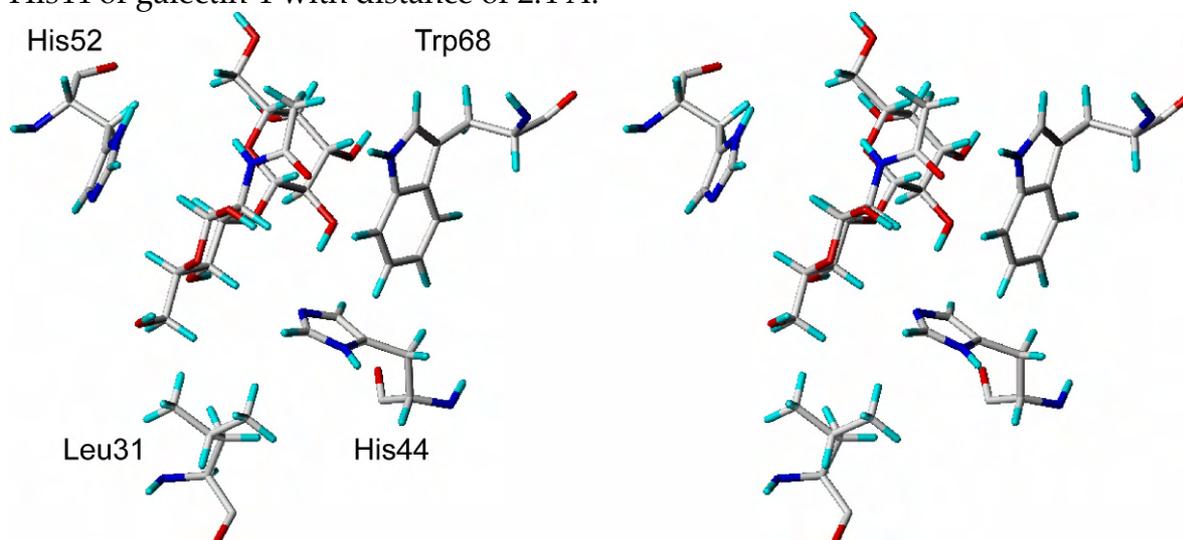


Figure 3.40: The interactions of the first ranked pose of LNB docked to galectin-1. The strongest interaction was calculated for H6 of GlcNAc as a result of close contact with a methyl group of Leu31. The second strongest interaction is calculated for H4 of GlcNAc as a result of close contact with His44. His52 and Trp68 are shown for reference.

From the comparison of STD effects calculated from distances of ligand and receptor protons with the results from STD NMR experiments (chapter 3.3.3) it was deduced that LNB is binding in a similar manner to that of LacNAc (type II) to galectin-1. The STD NMR experiments showed that the strongest intensities were found on H6' of the LNB disaccharide and only minimal intensity was found on H3

and the acetyl group of GlcNAc in LNB. This is in agreement with the poses ranked 8th, 11th, 17th and 18 in the docking. The pose achieving the best rank and poses in similar orientation must be wrong, because the strong STD effect calculated for H6 and H4 were not found in the STD NMR spectrum.

The selection of the poses was based on the assumption of a similar binding mode and the epitope from STD NMR confirming this assumption. The poses that were in accordance with the hypothesis of Bourne *et al.* and the STD NMR epitope were capable of forming the polar interaction with Arg48, Glu71 and Arg73 via the hydroxyl group in 4-position of the GlcNAc residue as proposed by their publication⁷⁷. Similar to the results for LacNAc the torsion angle of the OH6'-group changed, so that a *gg* conformation was obtained.

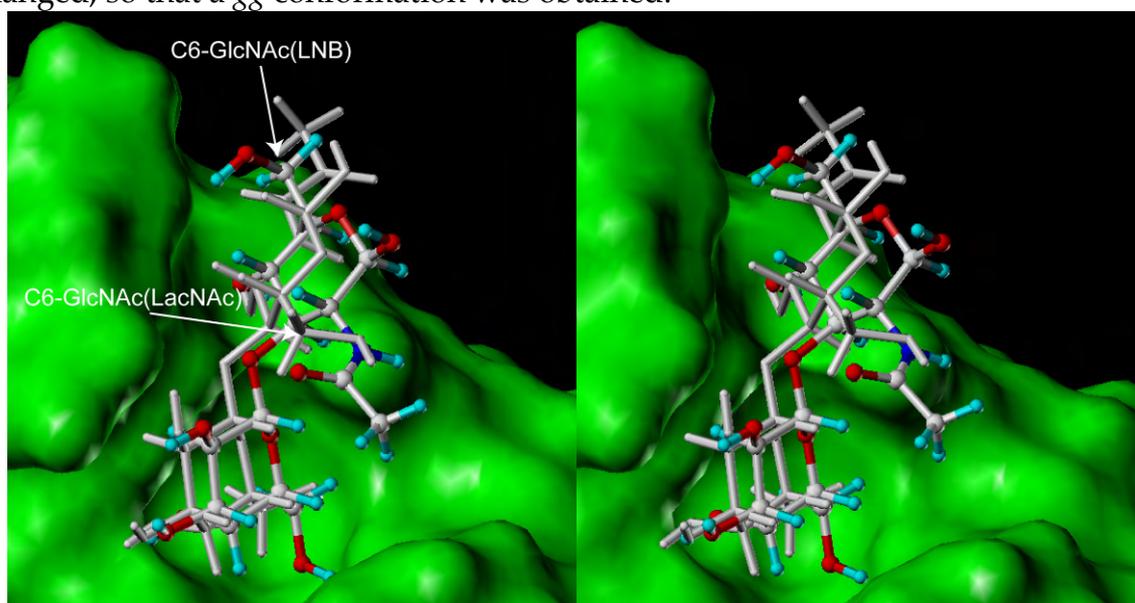


Figure 3.41: Superposition is shown for LacNAc (white, capped stick rendered) extracted from 1sla and LNB docked to the receptor from 1sla (ball and stick rendered). The pose of LNB shown in this picture was ranked 8th of 20 poses.

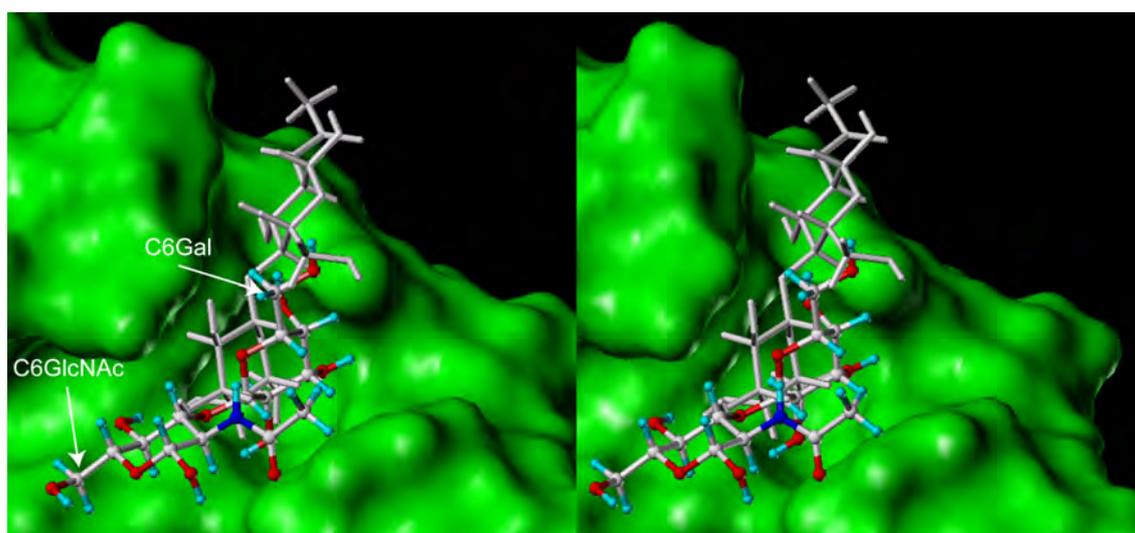


Figure 3.42: The first ranked pose from the docking experiment is oriented in a different pose as the LacNAc residue reference.

3.6.5 Binding mode of 3'-Sialyllactose

Several observations lead to assumption that the binding site of galectin-1 can accommodate additional carbohydrate residues attached to galactose residues in the binding site. Namely the 2'- and 3'-positions of lactose derivatives are assumed to be key positions for the attachment of further carbohydrates while maintaining specificity for galectins.

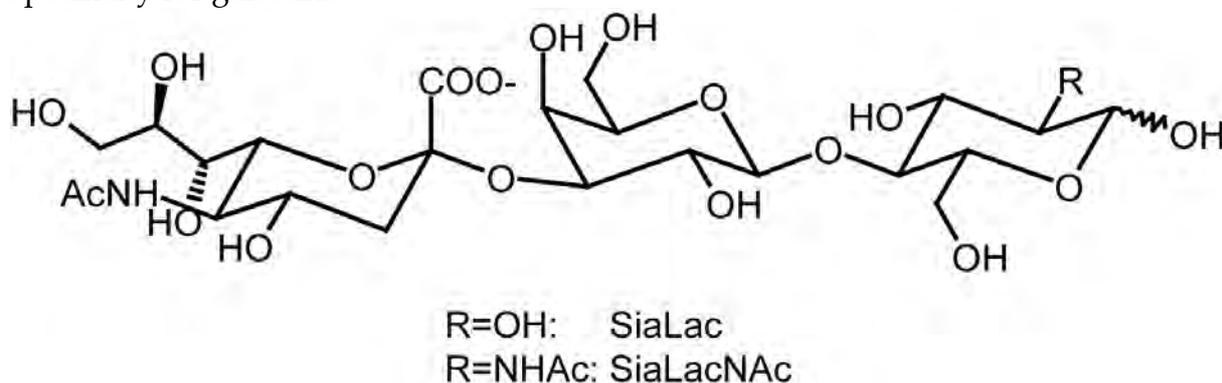


Figure 3.43: Structures of sialyllactose and sialyllactosamine

The specificity of galectin-1 for 2,3-sialyl lactose (SiaLac) and 2,3-sialyl lactosamine (SiaLacNAc) has been reported differently by several groups. C.P. Sparrow *et al.* determined 40% drop of relative activity compared to lactose, determined by an inhibition assay versus galectin-1 from human lung¹⁰⁵. The same assay applied to SiaLac binding to galectin-1 from rat lung indicated a 70% drop in relative activity¹⁰⁶. H. Ahmed *et al.* report 80% residual relative activity of SiaLac to galectin-1 from human spleen, determined by ELISA⁸⁸. Ahmad *et al.* report a K_D value of 435 μM for SiaLacNAc binding to galectin-1 from bovine heart muscle, determined by ITC⁸¹. A. Leppänen *et al.* determined an almost two fold increase in affinity of recombinant dimeric human galectin-1 to immobilized SiaLacNAc, compared to immobilized LacNAc. They concluded that galectin-1 exhibits higher affinity towards immobilized glycans than to glycans in solution as a result of the cross-linking capability of dimeric galectin-1. In their study monomeric galectin-1 showed distinctively lower affinity towards immobilized oligosaccharides. This would support the theory that galectins bind to glycans presented by the cell surface¹⁰⁷. M.G. Ford *et al.* report a three fold increased relative affinity of SiaLacNAc to galectin-1 compared to lactose determined by surface plasmon resonance¹⁰⁸.

To date no X-ray structure of SiaLac or SiaLacNAc in complex with a galectin has been published. Ford *et al.* examined SiaLacNAc in their MD simulation of galectin-1 in complex with various ligands¹⁰⁸. Data from these simulations were taken as reference for the docking results. STD NMR experiments were performed using SiaLac and galectin-1 for reference experimental data.

The first attempt to dock SiaLac to galectin-1 failed to deliver a consistent result cluster. SiaLac was extracted from the X-ray structure 1qfo¹⁰⁹. Only the first three of the poses proposed by DOCK were consistent with the binding mode of LacNAc. These three poses again incorporated the *tg* conformation at the OH-6'-position

already found for the other lactose derived ligands. In these poses the carboxyl group pointed in the direction of the side chain amino group of Lys63 over a distance of ~ 5 Å. This is in disagreement with the findings of Ford *et al.* where the carboxyl group is in close contact with the $N\epsilon$ of His52 (3.1 and 4.4 Å for both oxygen atoms). The massive difference of these proposed poses is based on the torsion angles of the NeuAca2-3Gal linkage. The ligand extracted from 1qfo contained torsion angles of $\Phi=290.1^\circ$ ($C1''-C2''-O-C3'$) and $\Psi=342.4^\circ$ ($C2''-O-C3'-H3'$) and was docked with $\Phi=288.7^\circ$ and $\Psi=8.3^\circ$. Ford *et al.* report average torsion angles of $\Phi=158.8^\circ$ and $\Psi=8.1^\circ$ for the bound state of SiaLacNAc.

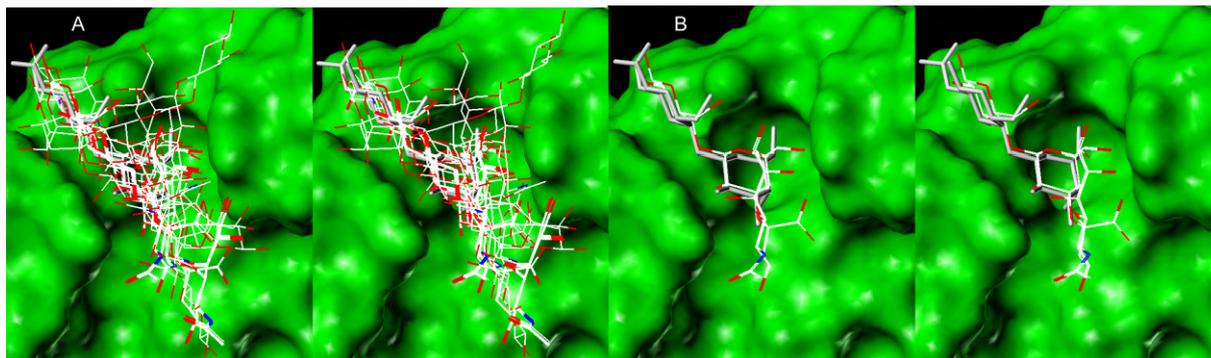


Figure 3.44: Panel A: complete resulting cluster generated by docking sialyllactose to galectin-1 (1sla). Panel B: the top 3 poses are in agreement with the binding mode of LacNAc.

The template structure of SiaLac was modified accordingly. In addition a model of SiaLacNAc was created with the same torsion angles. Both structures were energy minimized after creation over 100 steps. Docking experiments with the modified model of SiaLac still yielded poses that pointed the carboxyl group towards Lys63 with instead towards His52. To assess if this effect was based on the protein structure, new docking experiments were carried out using the protein from 1slt^{78;108}. Only one chain of the dimeric protein was used (A-chain). In order to provide improved conformational sampling a multiple random conformation search strategy was adopted (see also script 'rndscreen' on p138)²¹. For each of these experiments the top ten ranking structures were saved.

All these experiments yielded well defined clusters. Not all of the resulting clusters were oriented and placed into the binding site in accordance with the position of LacNAc from the X-ray structure, but most of the proposals did agree with the general orientation of the reference structure. However the orientations of the *N*-acetyl neuraminic acid residue were differing in most of the results and differed from the proposal of Ford *et al.* The structure proposed by the results of Ford *et al.* results projects the carboxyl group of Neu5Ac towards His52 and the glycerol side chain is pointing in the direction of Lys63 allowing a hydrogen bond with $N\epsilon 1$ of Trp68 via the 7''-hydroxyl group. An additional hydrogen bond is assumed to form between the 9''-hydroxyl group and the side chain amide of Asn33¹⁰⁸.

Most of the results from docking consisted of poses that pointed their carboxyl group of Neu5Ac'' past Trp68 towards Lys63. Only two results clusters contained structures where one of the oxygen atoms of the carboxyl group came close to the

He1 of His52. In these poses the glycerol side chain stretches into a ‘canyon’ between Cys2 on one side, Ala51 and His52 on the other side and Gly124 at the bottom. This twisted orientation is a result of the changes in the torsion angles of the α 2-3 glycosidic linkage. The torsion angles were found at $\Phi=59.5^\circ$ and $\Psi=-10.5^\circ$ for the best scored pose. The remaining poses in the cluster were very similar in orientation and conformation. Compared to the binding mode proposed for lactose derivatives the galactose residue of SiaLacNAc was shifted out of the cavity harboring the OH6’ group. The residue was generally placed closer to His52. Since OH6’ is not buried deep in the cavity of the galectin-1 binding site, the ω torsion angle remains almost unaltered in *gt* conformation. In this pose the ring stacking of the galactosyl ring with the aromatic side chain of Trp68 is defined by an angle θ between the normal planes of the ring systems of 139.8° and a distance R between the geometric centroids of the ring systems of 5.2 Å. Ford *et al.* report $\theta=144^\circ$ and $R=5.8^{106}$ Å. Liao *et al.* report $\theta=142^\circ$ and $R=5.1$ Å for LacNAc in the crystal structure 1slt⁷⁸⁷⁸. Thus it was concluded that the binding mode found in the docking is in agreement with the binding mode determined for LacNAc by X-ray crystallography. Only the proposal for the Neu5Ac residue is differing in the proposals made by this study and Ford *et al.*

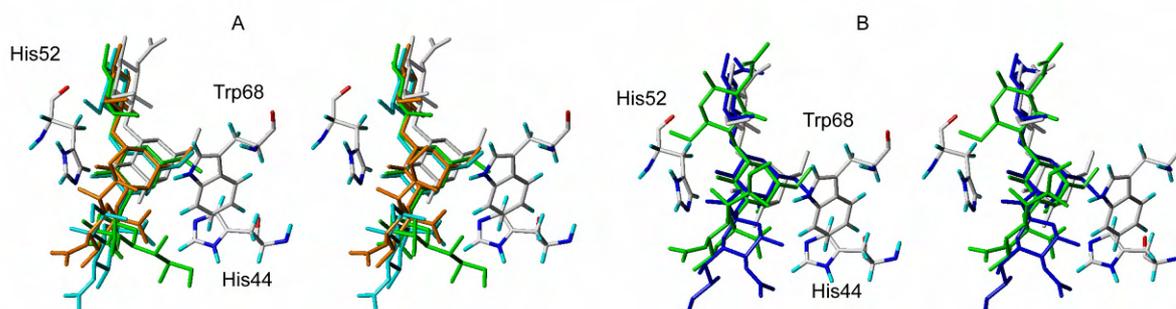


Figure 3.45: Display of some results (first ranks) from the random conformational search. Dockings were carried out with the protein structure 1slt, the position of the ligand is shown in white, His44, His52 and Trp68 are shown for orientation. A: results of SiaLac docked to galectin-1. Most of the resulting poses projected their carboxyl group past Trp68 towards Lys63 (not in picture, exemplary poses in cyan and orange). The green poses represents a result where the carboxyl group is pointing in the direction of His44. B: exemplary results of SiaLacNAc docked to galectin-1. The pose displayed in green points the carboxyl group past Trp68 towards Lys63 (not shown). The blue pose projects the carboxyl group away from the receptor surface.

A possible explanation of the differences between binding modes of SiaLacNAc proposed by MD simulation and docking experiment could be the absence of water in the docking experiment. For various X-ray structures of galectins in complex with carbohydrates the influence of hydrogen bonds mediated by water molecules located in vicinity to the binding site has been discussed⁷⁵⁻⁷⁸. It is reasonable to assume a stabilizing effect to arise from water molecules in the environment of the Neu5Ac residue in respect to the orientation of the residue. In addition the effect of water in attenuating the effect of polar interactions may have an influence as most poses direct their carboxyl group towards Lys63 over distances exceeding 5 Å. This space could easily be filled with up to three water molecules when a radius of 1.4 Å for one

molecule is assumed. Water molecules may also form intermolecular hydrogen bonds stabilizing the ligand conformation.

To assess whether this influence could be attained by increasing the dielectric constant used in the Coulomb term of the scoring functions the dielectric factor was increased to from 4.5 to 35. Ford *et al.* included an implicit counter-ion for the carboxyl group during their MD simulations, but there is no equivalent protocol that could be implemented in DOCK. Neutralization of charged functional groups by attachment of protons is feasible but would seriously alter the partial charges of the atoms involved. Non-covalently bound counter-ions can not be kept in vicinity of the charged group by means of features of DOCK. The increased dielectric constant did not result in docking results that were in the conformation described by Ford *et al.*¹⁰⁸.

The validity of the orientation of SiaLac from docking experiments was checked by comparison of the STD effect from NMR experiments and calculated from the structures from docking.

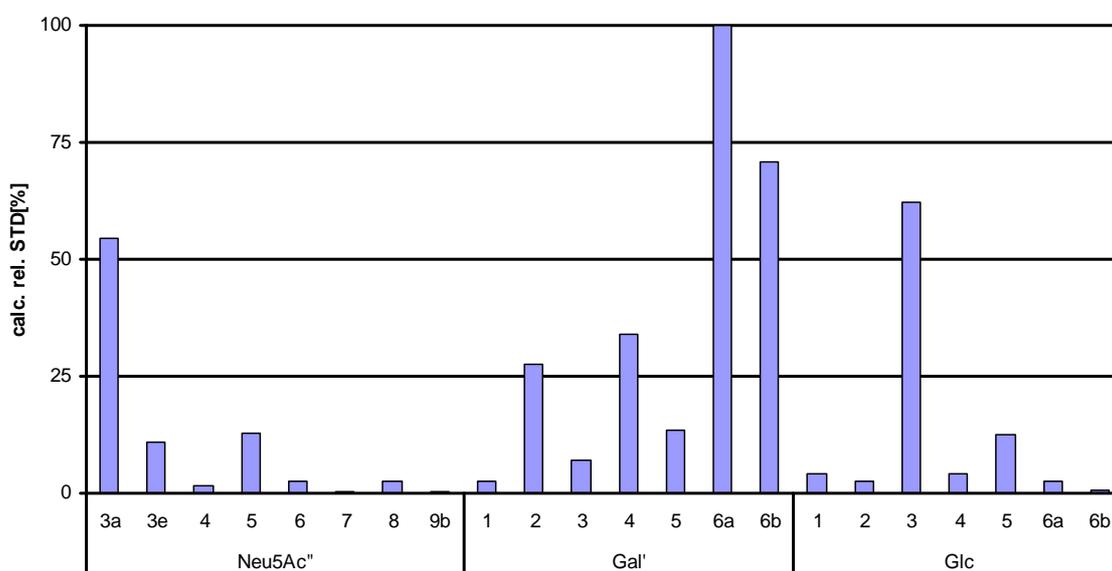


Figure 3.46: Calculated STD intensities for SiaLac, based on the average values of the top three docked results

The calculated relative STD effect determined for the average values of the three poses shown in Figure 3.44b exhibited the pattern expected on the basis of the lactose like binding mode with 6'- and 4'-position rendering the highest intensities. This pattern has already been determined for the docking results of other lactose derivatives. H2' is stronger accentuated in these orientations as for the other lactose binding modes. As before the 3-position of the Glc residue displays strong intensity as was found for the binding mode of lactose and *N*-acetyl lactosamine. Only the equatorial proton in 3-position (3e'') of the Neu5Ac'' residue is showing significant intensity of almost 55%. The axial proton in 3-position (3a'') and 5'' are exceeding 10% relative intensity and the remaining protons of the Neu5Ac'' residue are displaying less than 3% relative intensity.

The 1D STD NMR of SiaLac displayed some distinct STD effects. The spectrum was difficult to interpret as the 1D spectrum of the trisaccharide exhibited significant signal overlap. To reduce sample volume and receptor required the STD spectra were acquired in 3 mm NMR tubes containing 9 μM galectin-1 and 750 μM SiaLac (~80 fold excess) in 160 μL sample volume. 3 mm tubes proved to be more useful in water suppression experiments in the 700 MHz NMR spectrometer and displayed similar spectral quality for proton spectra when compared to the classical 5 mm sample tubes. Unfortunately no hetero nuclear STD spectrum could be acquired with this sample. The STD HSQC experiment would have allowed the distinct determination of the STD effect for all protons. In the 1D ^1H STD spectrum some signals could not be assigned unambiguously. Artifacts were determined for the signals of 3a'', 3e'' and Ac'' when comparing STD intensities. For the determination of the binding epitope artifacts were subtracted from the STD NMR results. The determination of the binding epitope of sialyllactose by STD NMR was difficult due to signal overlap.

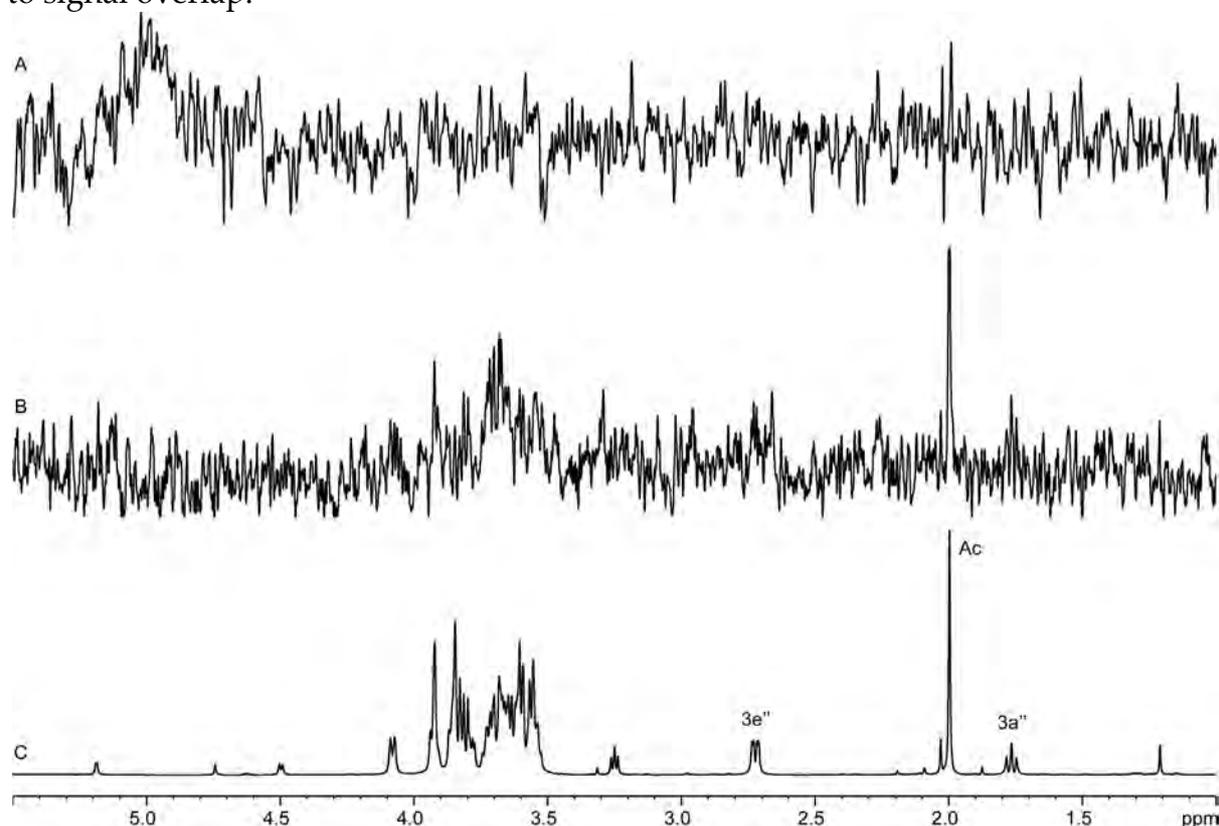


Figure 3.47: Panel A: blank test STD spectrum of 750 μM SiaLac on 500 MHz spectrometer with 1024 scans at 300 K, *on resonance* irradiation at -500 Hz and 50 dB pulse power. Panel B: STD spectrum of 80fold excess of SiaLac with galectin-1. *On resonance* radiation was applied at -525 Hz at 49 dB pulse power at 300 K. Panel C: *Off resonance* spectrum. *Off resonance* irradiation was applied at 20 kHz. (B and C: 2048 scans, 700 MHz spectrometer)

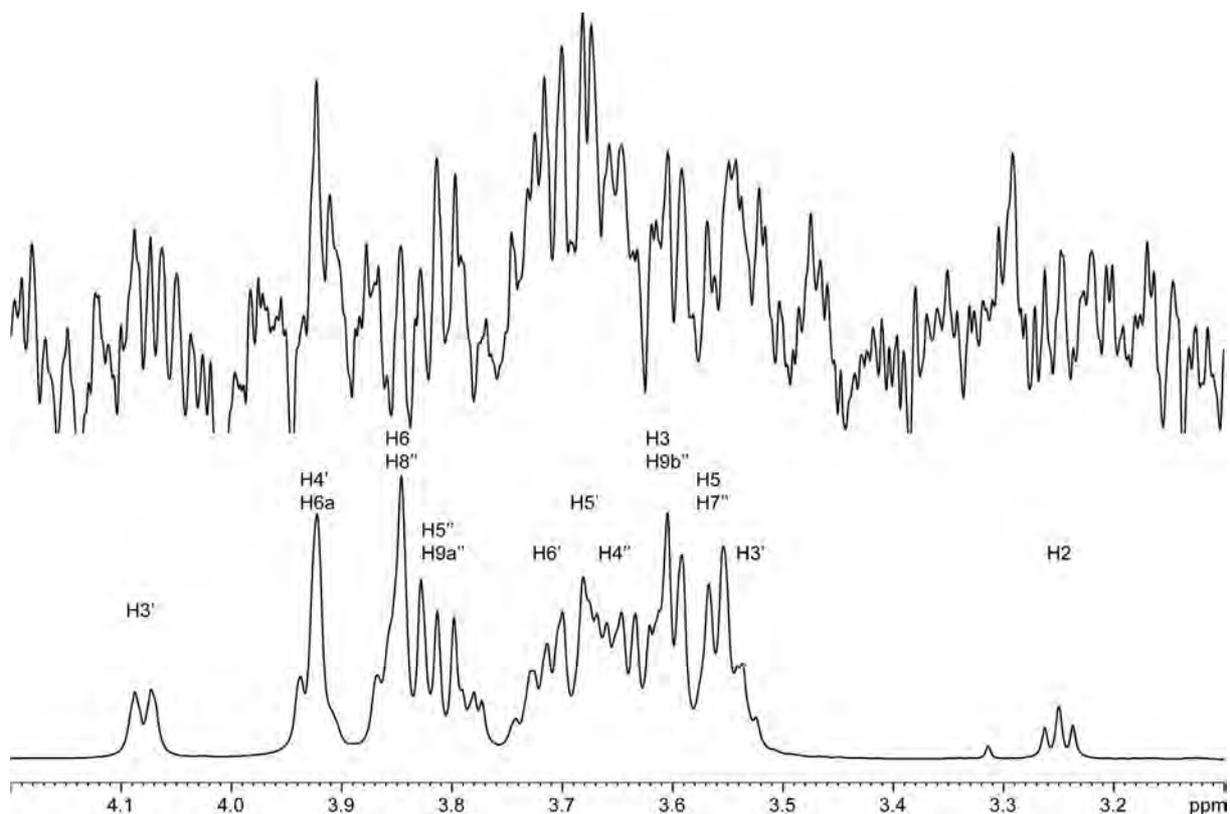


Figure 3.48: Expansion of the STD spectrum (top) and *off resonance* spectrum (bottom with assignments) of sialyllactose.

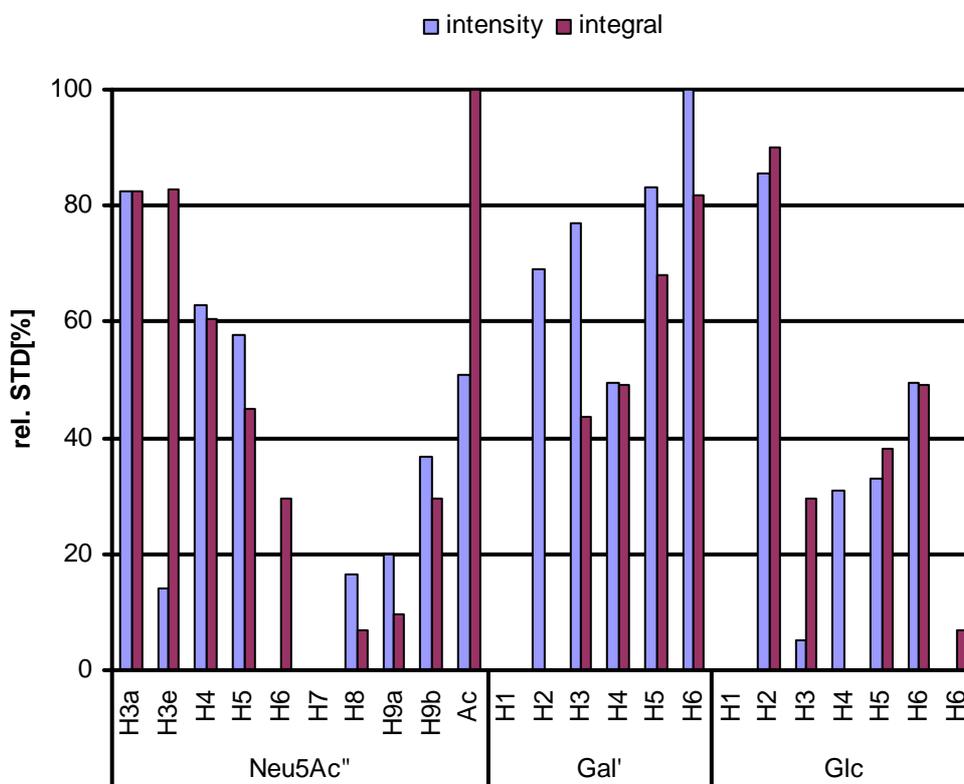


Figure 3.49: Binding epitope of SiaLac determined by STD NMR using signal intensities and integration. Some signals overlap and thus were assigned the same relative STD effect. Artifacts were subtracted.

The STD epitope resembles that of other lactose derivatives in contact with galectin-1. H6' of the Gal' residue gives strong intensity in agreement with all other STD NMR experiments. In contrast to LacNAc the 5'-position yielded a stronger STD signal than the 4'-position. This was alleged to be a result of signal overlap. The intensity for H5' may be overrated due to overlap with H6', bearing strong intensity itself, and H4''. This effect was observed for both methods to determine the STD effect, by signal intensity and integration of signal.

H4' is overlapping with H6a. Based on the docking results and the X-ray structures the protons in 6-position of the Glc residue should not be subjected to large magnetization transfer. From these findings it could be deduced that the STD signal entirely resulted from saturation transferred to H4'. This would increase the STD intensity of H4' and deplete STD intensity for H6a. However the STD spectrum of LacNAcOEt revealed relative STD intensities from 45 to 55% for the protons in 6-position of the GlcNAc residue.

H2 of the GlcNAc residue yielded a weak signal in the *off resonance* and difference spectra. As before, in the case of methyl lactose the signal of H2 did not exceed the noise level in the difference spectrum (chapter 3.3.1, p21 and Figure 3.5). Thus the signal was discarded from further discussion.

The best three poses did give strong indication for the difference in the intensities of H3a'' and H3e'' observed when using signal intensity as measure. H3a'' is pointing directly towards His52. The distance to Hε1 of the histidine side chain is 2.28 Å. H3e'' is making several contacts with the side chains of Leu31 and His44 over distances exceeding 4 Å, resulting in weak calculated STD effects.

Although artifacts were incorporated in the determination of the epitope determined by STD NMR the intensities of H5'', H9a'', H9b'' and the acetyl group in the Neu5Ac'' residue were not supported by the binding modes from the docking experiments. This is either due to inadequate treatment of the artifacts in the STD NMR experiments or inaccurate docking results.

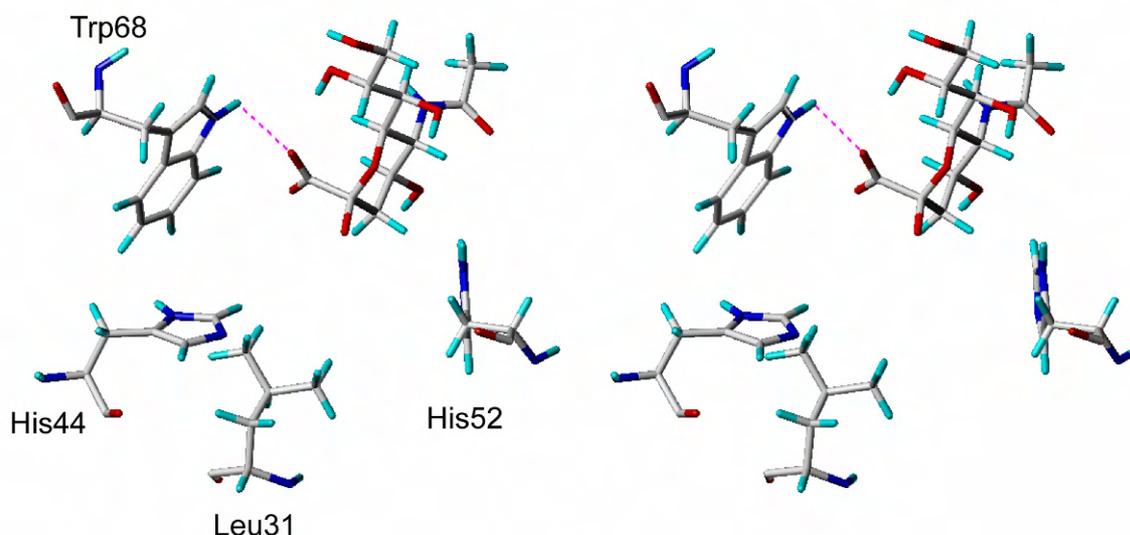


Figure 3.50: Orientation of the Neu5Ac residue in SiaLac from docking (rank 1). The interactions with the neighboring amino acids of galectin-1 are shown from the perspective of the Lac residue of SiaLac (omitted for clarity). Strong interactions of H3a'' determined from STD NMR and calculations arise from vicinity to H ϵ of His52. H3e'' may interact with His44 and Leu31 over distances exceeding 4 Å. In this orientation the *N*-acetyl residue points away from the protein. The dotted line (magenta) indicates a hydrogen bond between the carboxyl group of Neu5Ac to H ϵ of Trp68.

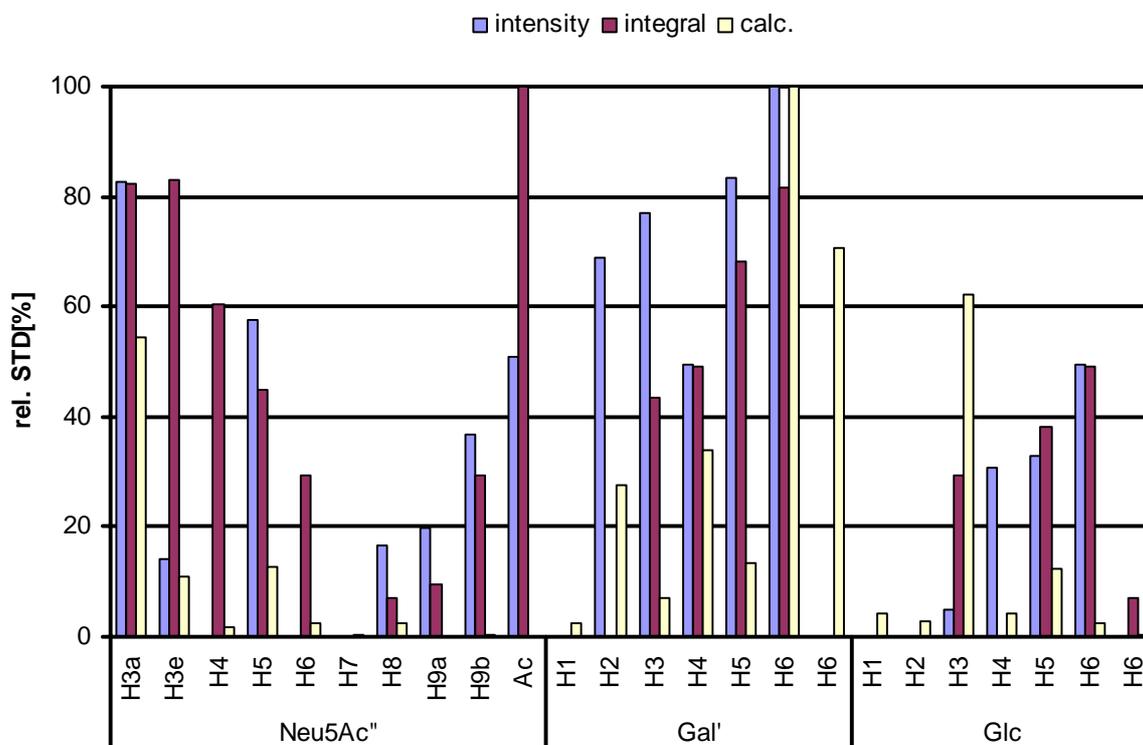


Figure 3.51: Comparison of experimental and calculated relative STD effects for SiaLac. The intensity for H2 has been omitted (see text).

3.7 Influence of the histidine protonation state on the docking results

The binding site and its surrounding area of galectin-1 from *bos Taurus* contains two histidine residues, His44 and His52. His44 is considered to be part of the CRD in all galectins, His52 is conserved in human and bovine galectin-1. His52 neither constitutes a conserved amino acid in different galectins, nor is it considered as part of the binding motive in galectin-1^{60,62}.

Histidine poses a challenge to the generation of receptor models for computational binding studies. Histidine may occur in three different states of protonation (Figure 3.52). Each of these states may have influence on the binding mode of ligands. The differences may occur as results of spatial requirements of the histidine side chain, electrostatic behavior and the possibilities to form hydrogen bonds. These influences may have an extraordinary impact when the docked ligand is carrying a charged group which may form additional interactions with the receptor.

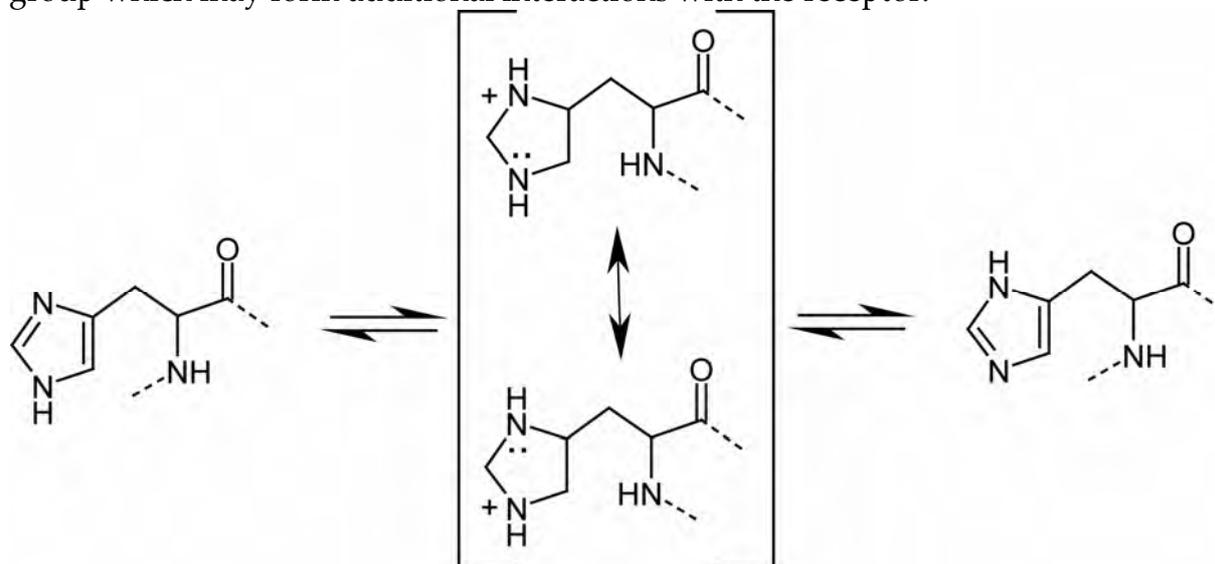


Figure 3.52: The different states of histidine protonation: on the left side the ϵ position is protonated, on the right side the δ position is protonated.

In their study Ford *et al.* assumed all histidine residues to be protonated at the $N\epsilon$ position. Preparation of the galectin-1 model as described yielded histidine with the $N\delta$ position protonated. This is a result of the residue recognition implemented in Sybyl⁸⁹, where the amino acid residue 'His' is interpreted as $N\delta$ protonated residue (residue: 'Hid'). Alternative protonation states can be incorporated into the model by changing the residue to 'Hie' for the $N\epsilon$ protonated or 'Hip' for the di-protonated and positively charged histidine.

3.7.1 Docking of *N*-acetyl lactosamine with different histidine types in the receptor model

To assess the impact of the protonation state of the histidine residues the receptor model of galectin-1 was changed to yield a receptor with either both histidine residues with the proton attached to $N\delta$ according to Ford *et al.*¹⁰⁸, or the

di-protonated state. For both models grid points were created. No further modifications were applied to the docking setup.

To test the influence of the protonation state of the histidine residues on the binding mode lactosamine type I (LNB) and type II (LacNAc) were docked to the receptor models using the same parameters as for SiaLac. These ligands do not exhibit charged functionalities. Interactions rely solely on nonionic polar and unpolar interactions. Polar interactions like hydrogen bonds may be influenced by the change of protonation of the histidine residue. Additionally the change of spatial requirements and new possibilities to form hydrogen bonds may have influence on the binding mode of the ligands.

The docking results achieved from the disaccharides did not support assumptions that their proposed binding mode was depending on either position of the hydrogen atom at the histidine side chain or the position of the spheres used for the initial ligand placement. Results from these docking experiments were in agreement with experiments carried out earlier with the initial receptor structure and sphere cluster. This again depicted the consistency of the docking procedure in respect to this class of molecules. The deviation of the first poses in the respective dockings varied from 1.27 Å to 1.98 Å. In their study J. Erikson *et al.* treated RMSD values less than 2.0 Å as successful docking and RMSD values exceeding 2.0 Å as failure²⁷. Thus the results were treated as successful docking indicating that the protonation state of the histidine residues His44 and His52 is not relevant for docking of uncharged carbohydrates. One reason for the change of the ligands orientation may be the occurrence of a new hydrogen bond from NH ϵ of His44 to O4'' in the disaccharide. This hydrogen bond can only form when N ϵ of His44 is protonated.

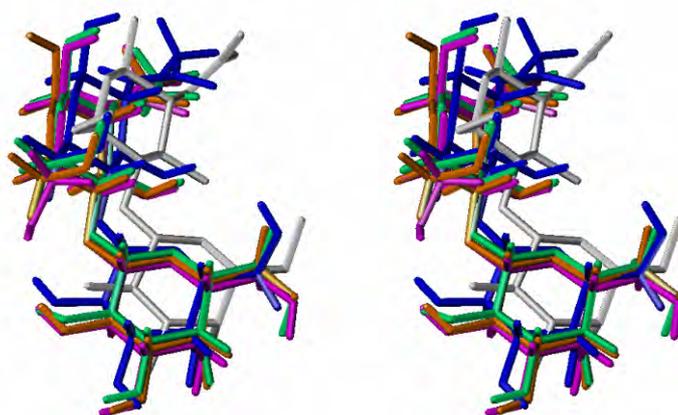


Figure 3.53: Comparison of LacNAc poses from different docking setups: white: X-ray structure (1slt); magenta: histidine with N ϵ H, RMSD: 1.27 Å; orange: same receptor, alternative sphere cluster (Figure 3.65), RMSD: 1.62 Å; blue-green: histidine with N δ H and N ϵ H, alternative sphere cluster, RMSD: 1.86 Å; blue: histidine with N δ H original sphere cluster, RMSD: 1.98 Å. The offset from the reference structure may be a resulting from the random search approach, effects from the force field (Lennard-Jones potential)²⁷ or additional hydrogen bonds.

3.7.2 Docking of 2,3-sialyl lactose with different histidine types in the receptor model

The results obtained from docking sialyllactose to galectin-1 were distinctly more dependent on the docking setup regarding receptor structure and sphere composition. This is easily explicable with the influence of the charged carboxyl group of the neuraminic acid residue. The resulting cluster obtained for sialyllactose docked to galectin-1 (1slt) with the proton at the N ϵ position of His44 and His52 represented a consistent binding mode. Members of this group represented the top ten of the ranking but did not match the binding mode of *N*-acetyl lactosamine. The top ranking structures obtained an orientation where the OH6-group of the Gal' residue is located outside the usual pocket. Instead the OH6-group is placed between His44 and His52 allowing the formation of a hydrogen bond from the hydroxyl group to the proton at N ϵ of His44. This orientation also represented an unusual ψ torsion angle from Neu5Ac to Gal (-50.9°).

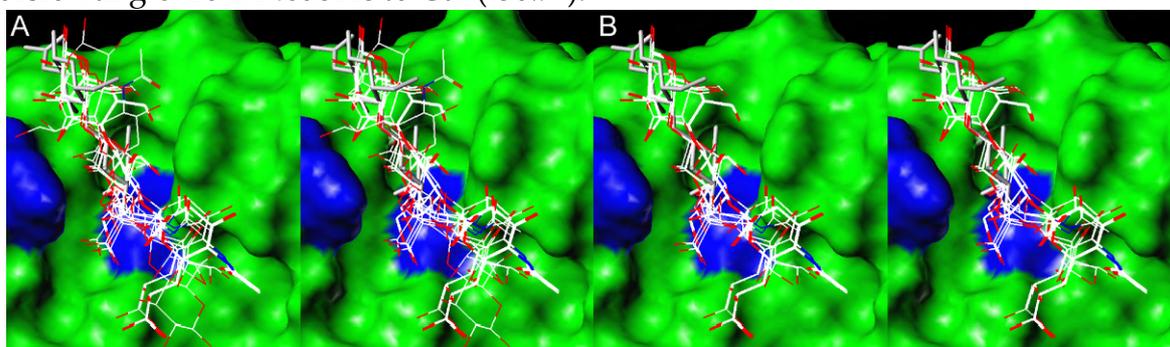


Figure 3.54: Results achieved from docking SiaLac to galectin-1 (1slt) with Hie residue type. Blue: surface of histidine residues, His44 at the bottom, His52 left of the ligand. A: complete result cluster; B: top 18 poses. The position of LacNAc in 1slt is shown in white and capped stick rendering for comparison.

Table 3.7: Comparison of conformations of SiaLac: the template structure is the original structure read in by DOCK. Ranks 14 and 17 to 19 are in accordance with the binding mode of LacNAc in the crystal structure. Rank 1 and rank 14 have similar conformation but differ in placing and orientation. For comparison the structural values given by Ford *et al.*¹⁰⁸ are given, the energy score corresponds to ΔG_{bind} with dielectric constant $\epsilon=4$.

	Neu5Ac α 2-3Gal		Gal β 1-4Glc		GalOH6	Energy score
	Φ [°]	Ψ [°]	Φ [°]	Ψ [°]	ω [°]	
template	172.4	4.5	48.7	14.0	180.1	N.A.
rank 14	182.1	9.7	63.3	17.0	187.5	-27.99
rank 17	182.1	9.7	63.3	17.0	187.5	-27.51
rank 18	176.9	9.6	63.9	18.4	187.5	-25.22
rank 19	191.2	346.1	14.6	322.8	302.0	-24.92
rank 1	187.0	329.1	57.2	14.3	194.3	-29.61
Ford <i>et al.</i> (avg.)	158.8	8.1	45.1	14.2	--	-32.80

The result obtained with galectin-1 presenting the double protonated histidine side chains displayed higher consistency and were in accordance with the binding mode of *N*-acetyl lactosamine in the crystal structure. Only the last three poses were placed in a different binding mode. The remaining poses were displaying similar

conformations that differed only slightly from the template conformation. The Φ torsion angle from Gal' to Glc changed from 49° to 65° .

The conformations obtained were in accordance with the results of Ford *et al.* In their study they report torsion angle deviations of 12.9° and 11.5° for the glycosidic linkage from Neu5Ac'' to Gal' and 10.3° and 9.5° for the glycosidic linkage from Gal' to Glc¹⁰⁸. Still none of the poses displayed the distances between N δ of His52 and the oxygen atoms of C1 in the neuraminic acid residue reported by Ford *et al.* The results obtained for histidine protonated at both nitrogen atoms of the imidazole residue are of limited relevance regarding healthy tissue. Histidine occurs in the positively charged state only in environments exceeding an acidity of pH 6.

The results obtained for different protonation states of histidine were similar to each other. It was concluded that the protonation state of histidine was irrelevant to the binding capabilities of galectin-1 towards sialyllactose. The results obtained with histidine protonated in N ϵ position yielded poses allowing saturation to be transferred to the *N*-acetyl group of the neuraminic acid residue as found in STD NMR. None of the results were in agreement with the proposals of Ford *et al.* The differences can be ascribed to the different setups of the modeling approaches and the influence of water and explicit counter ions used in the MD simulations¹⁰⁸.

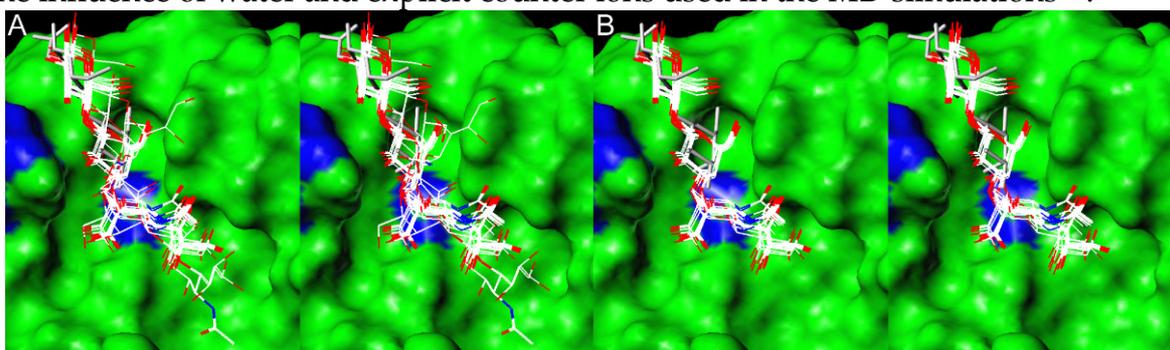


Figure 3.55: Docking results achieved from docking SiaLac to galectin-1 (1slt) with residue type Hip, with both nitrogen atoms in the histidine side chain protonated. Blue: surface of histidine residues, His44 at the bottom, His52 left of the ligand. A: complete resulting cluster; B: consistent top 17 scoring poses. (LacNAc reference in white capped sticks)

Table 3.8: The comparison of the structures docked to galectin-1 with double protonated histidine residues reveals high similarity in their conformations. Rank 1 to 17 were docked in similar conformation and orientation. The remaining poses were differing notably. For comparison the structural values given by Ford *et al.* are given, the energy score corresponds to ΔG_{bind} with dielectric constant $\epsilon=4$ determined by the AMBER software used in their study¹⁰⁸.

	Neu5Ac α 2-3Gal		Gal β 1-4Glc		GalOH6	Energy score
	Φ [$^\circ$]	Ψ [$^\circ$]	Φ [$^\circ$]	Ψ [$^\circ$]	ω [$^\circ$]	
template	172.4	4.5	48.7	14.0	180.1	N.A.
rank 1	186.4	8.8	65.0	15.9	68.8	-28.35
rank 5	186.4	8.8	65.1	15.9	68.8	-27.44
rank 17	186.4	8.8	65.0	15.9	68.8	-26.14
rank 18	71.0	343.0	14.2	323.1	186.3	-25.84
rank 19	270.9	16.8	62.9	18.8	97.3	-25.77
Ford <i>et al.</i> (avg.)	158.8	8.1	45.1	14.2	--	-32.80

3.7.3 Binding mode of lacto-*N*-tetraose

Lacto-*N*-tetraose (LNT) has been described as a ligand for galectin-1 and galectin-3. F.C. Brewer reports binding affinities of several oligosaccharides towards galectin-1 from bovine heart muscle, recombinant murine galectin-3 and recombinant human galectin-7 by isothermal calorimetric titration (ITC). The affinity of LNT towards galectin-1 and galectin-3 has been determined with K_D values of 100 μM and 38.5 μM respectively (K_a 1.0 $\text{M}^{-1}\cdot 10^{-4}$ and 2.6 $\text{M}^{-1}\cdot 10^{-4}$). In an earlier experiment N. Ahmad *et al.* determined the same binding affinities for these proteins. They also detected weak binding of the trisaccharide $\text{GlcNAc}\beta(1-3)\text{Gal}\beta(1-4)\text{Glc}$ with a $K_D=1.1$ mM ($K_a=0.88$ $\text{M}^{-1}\cdot 10^{-4}$) to galectin-3⁸¹. These results indicate that galactose containing saccharides may be recognized even if the galactose is not the terminal residue at the non reducing end of the oligosaccharide. However, A. Leppanen *et al.* did not observe detectible binding of dimeric galectin-1 towards oligosaccharides lacking terminal galactose residues at the reducing end¹⁰⁷. It is widely assumed, that interactions of the hydroxyl group in 4-position of the galactose residue are relevant for selective galactose recognition by galectins. The hydroxyl group in 3-position does not seem to be important for binding and could be used for further modifications i.e. glycosylation or synthetic modifications, which were tested by some groups to create potent galectin-3 inhibitors¹¹⁰.

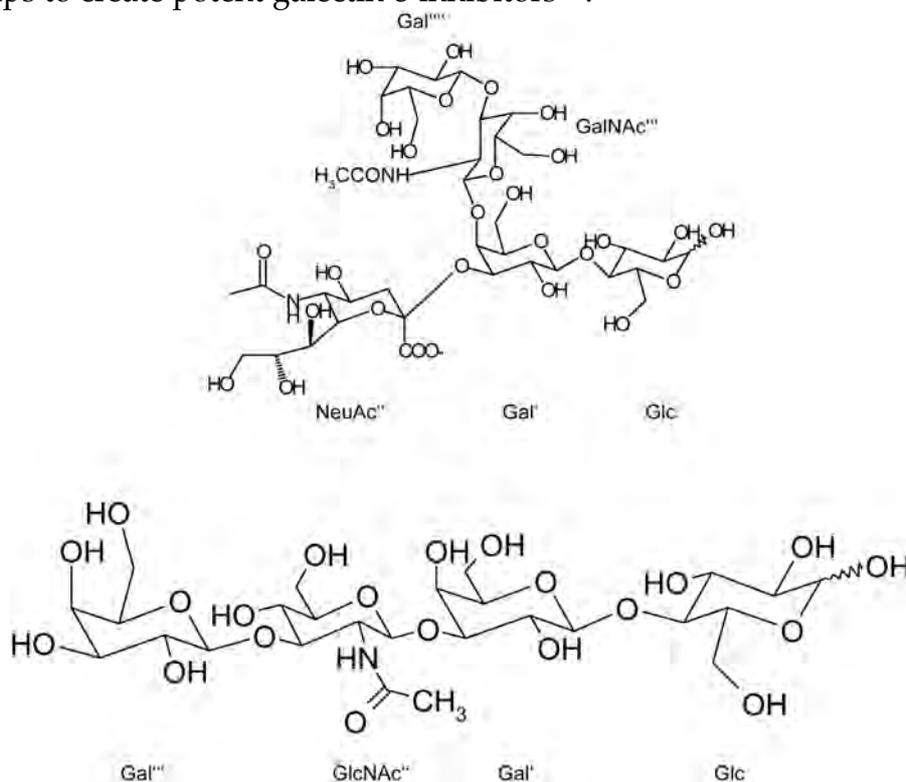


Figure 3.56: Top: structure of GM1 as studied by Siebert *et al.* (naming of residues as in the report). Bottom: structure of lacto-*N*-tetraose (LNT).

LNT contains two β -galactoside residues, one is the terminus at the non reducing end, and one is second in the sequence ($\text{Gal}\beta(1-3)\text{GlcNAc}\beta(1-3)\text{Gal}\beta(1-4)\text{Glc}$). This could allow binding modes of the tetrasaccharide where either one of these galactose residues occupies the binding pocket in the CRD of the galectins or the

tetrasaccharide is changing its binding mode dynamically, where the ligand may reorient in the binding pocket to alternate the galactosyl residue in contact with the CRD. H.C. Siebert *et al.* determined the bound structure and binding epitope of the ganglioside GM1 bound to galectin-1¹¹¹. Their results indicate that only the terminal galactose residue is in close contact with the protein. The binding mode seems to be similar to the crystallized complexes of various galectins.

In this work the binding mode of LNT in contact with galectin-1 was studied by docking experiments and STD NMR experiments as described before.

The STD NMR blank test with LNT displayed the presence of artifacts. The intensity correlated with the respective resonance frequency and the irradiation frequency.

The 1D ^1H STD spectrum of LNT exhibits distinct intensities over the whole spectral window of the ligand. The intensities are not linked to the resonance frequencies of the signals. Thus, it was assumed that a selective binding event was detected by STD NMR. Significant overlap in the region from 3.35 to 4.05 ppm caused problems in assignment of the resonances and calculation of their respective STD effects in 1D ^1H STD NMR experiments.

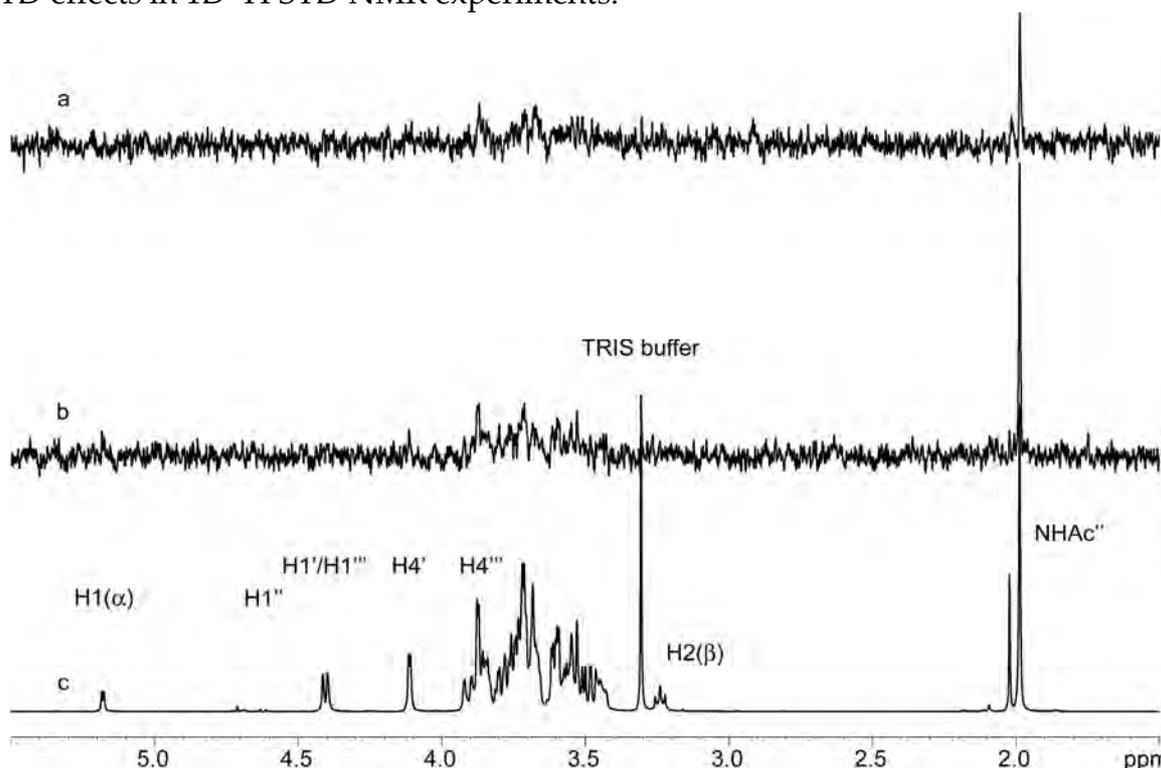


Figure 3.57: Determination of artifacts in the spectra of LNT ($c= 1$ mM). a: irradiation at -500 Hz (-1 ppm); b: irradiation at 0 Hz (0 ppm); c: reference spectrum with *off resonance* irradiation at 40 kHz. All spectra were recorded at 300 K with $T_{\text{sat}}=2$ s and a pulse power of 45 dB for the saturation pulses. Assignment was made for the separated peaks that could be assigned from 1D spectrum. The spectra were recorded on a 500 MHz spectrometer.

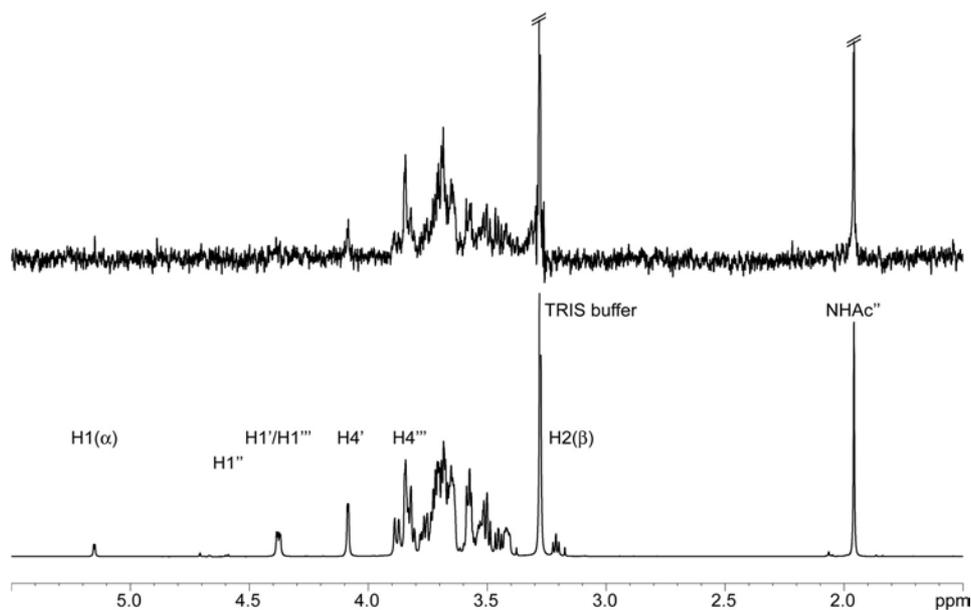


Figure 3.58: Top: STD spectrum of LNT. Bottom: *off resonance* spectrum of LNT. Irradiation was applied at -525 Hz (*on*) and 40 kHz (*off resonance*) on a 700 MHz spectrometer at 300 K, 45 dB saturation pulse power and $T_{\text{sat}}=2$ s ($c(\text{LNT})=1.68$ mM, 100 fold excess).

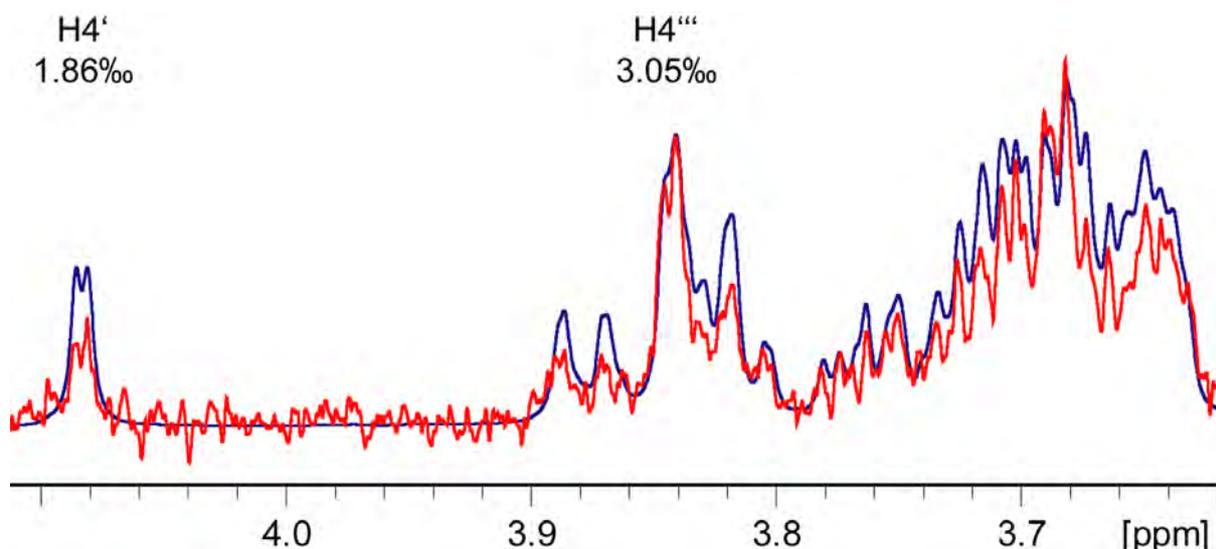


Figure 3.59: Comparison of the STD effect determined from signal intensity for $\text{H4}'$ and $\text{H4}'''$: obviously $\text{H4}'''$ is more involved in the binding process (red: STD spectrum, blue: *off resonance* spectrum). In the region from 3.6 to 3.9 the resonances of the H6 groups of all carbohydrates are subject to overlap with each other and further protons of the carbohydrate residues.

Since overlap prevented proper assignment of the resonances in the 1D spectra 2D spectra were employed. HMBC, phase sensitive HSQC, TOCSY and COSY spectra were recorded to allow determination and assignment of chemical shifts. Although most of the resonances could be assigned unambiguously, the protons in 6-position of the carbohydrates proved to be difficult to assign to a specific carbohydrate residue. Unambiguous assignments were only possible for $\text{H6}''$ and one proton in 6-position of the terminal Glc in β -configuration.

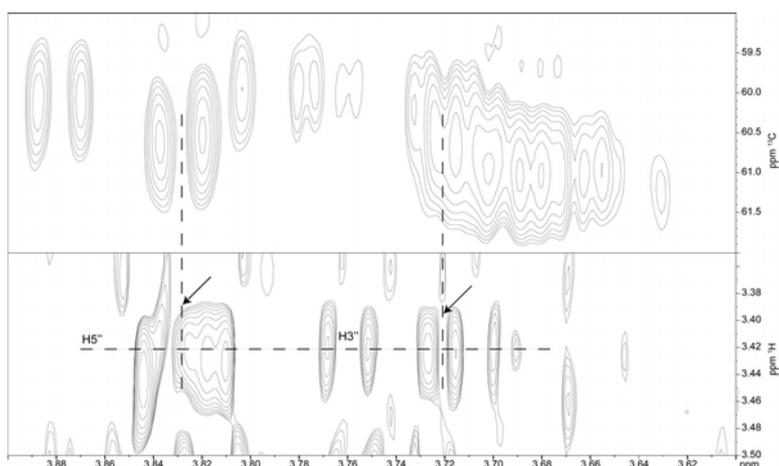


Figure 3.60: Overlay of the region of a phase sensitive HSQC spectrum (top) including the resonances of the H6-groups and a section of a TOCSY spectrum (bottom). The section of the TOCSY displays the trace of the H5'' proton. In this trace coupling to H3'' and the two protons in 6-position is detectable (arrows).

The significant signal overlap also prevented unambiguous determination of STD effects for discrete proton resonances. 2D NMR experiments like STD TOCSY and STD HSQC spectra allow the discrete determination of STD effects for dispersed signals. The sample used for the STD HSQC experiment contained 1.68 mM LNT and 18 μ M galectin-1. The STD HSQC experiment with this sample did not yield a difference spectrum. Two dimensional STD experiments require more signal intensity than could be observed in this sample. The resulting lack of information on the binding epitope of LNT binding to galectin-1 posed difficulties in the determination of the binding mode of LNT binding to galectin-1.

As stated before the number of rotatable bonds in the ligand structure is a key factor for the accuracy of the docking procedure. Although the initial setup chosen in this study proved successful in finding the binding mode of disaccharides in comparison with experimental data from X-ray crystallography and STD NMR, demonstrated in this study, performance fell off significantly in the prediction of the binding mode for the trisaccharide SiaLac (chapter 3.6.5).

The docking experiments with the model of LNT were carried out two times. In one experiment the terminal Gal''' residue was chosen to be the anchor fragment. In a second model the Gal' residue was chosen to be the anchor fragment. As before only the ring atoms of the residue were defined as anchor.

Both approaches to dock LNT into the binding site of galectin-1 from the crystal structure 1sla failed to yield a consistent result. None of the resulting poses represented a binding mode similar to that of *N*-acetyl lactosamine. Additionally the carbohydrate structures created by dock did not represent preferred glycosidic torsion angles.

Table 3.9: The torsion angles of the glycosidic bonds in LNT docked to galectin-1 did not represent low energy conformations⁹⁰. Unfavorable torsion angles were found for the top scoring ranks and the last ranks saved (rank 20).

		Φ Gal'''- GlcNAc'' [°]	Ψ Gal'''- GlcNAc'' [°]	Φ GlcNAc''- Gal' [°]	Ψ GlcNAc'' - Gal' [°]	Φ Gal'- Glc [°]	Ψ Gal'- Glc [°]
	expected	50	10	50	10	40	15
Gal''' anchor	top 12	-24	-32	67	66	61	61
	rank 20	-31	-42	41	-6	70	62
Gal' anchor	top 14	34	-61	58	54	67	67
	rank 20	34	-61	20	-23	-170	65

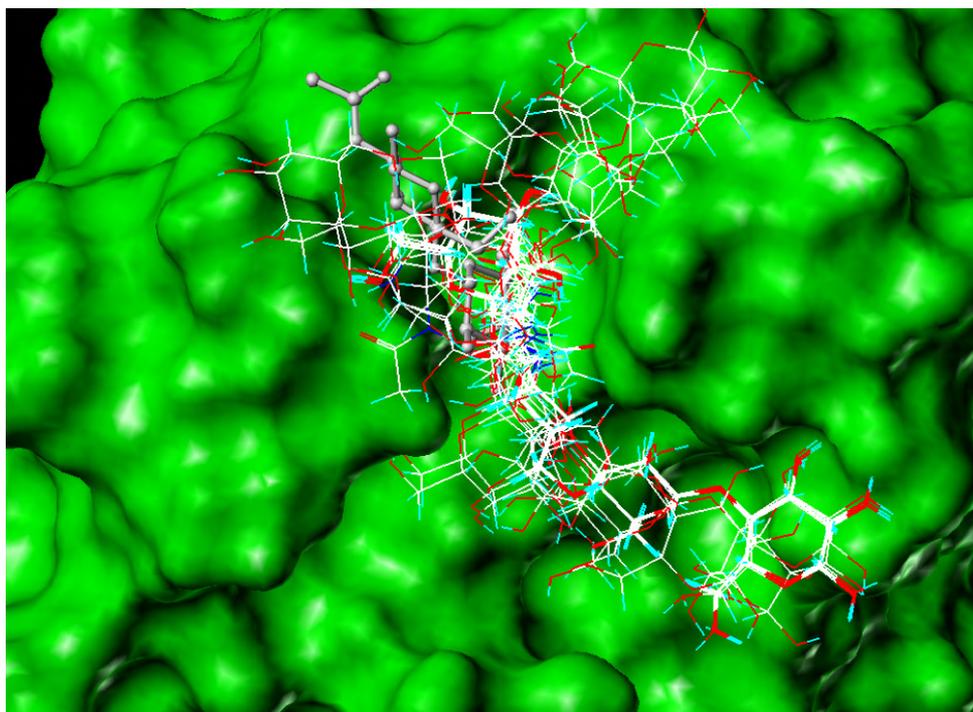


Figure 3.61: The results obtained for LNT docked to galectin-1 (1sla) with the terminal Gal''' residue as anchor fragment did not include poses in agreement with the binding mode of lactosamine (white, ball and stick rendering).

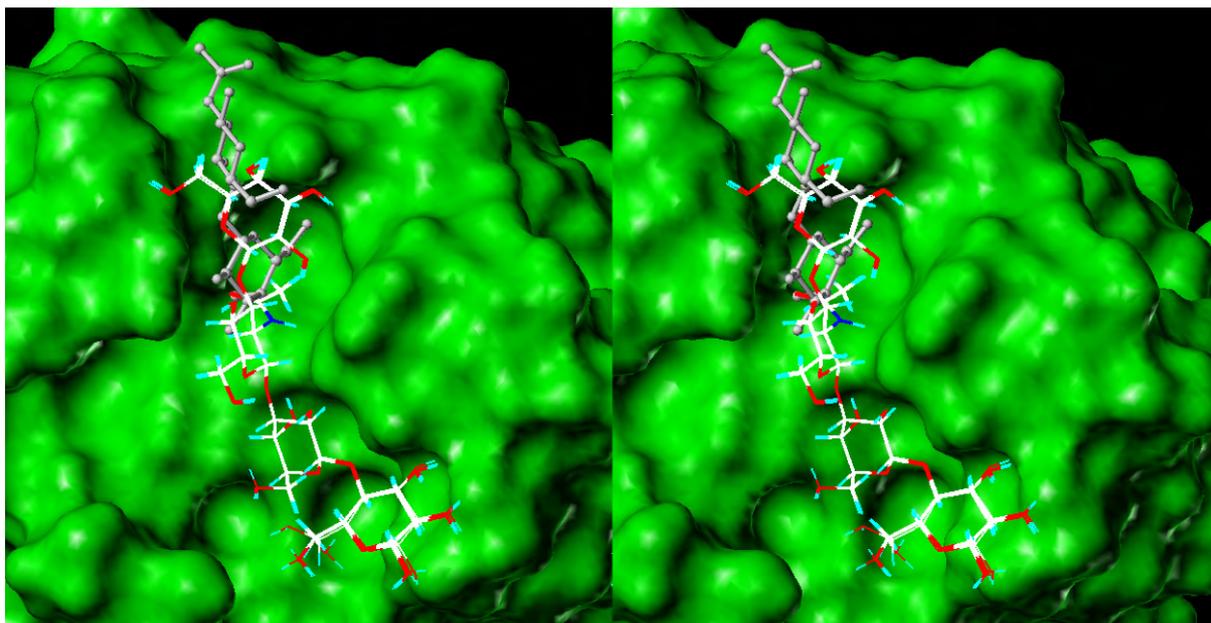


Figure 3.62: The top 12 poses of the docking experiment, using Gal''' as anchor, obtained a consistent binding mode, 'wrapped' around Trp68. The conformation of the glycosidic bonds of tetrasaccharide did not represent low energy torsions (Table 3.9). In the X-ray structure the reducing end of the carbohydrate is extending towards the top of the picture. The docked structures extend towards the bottom of the picture.

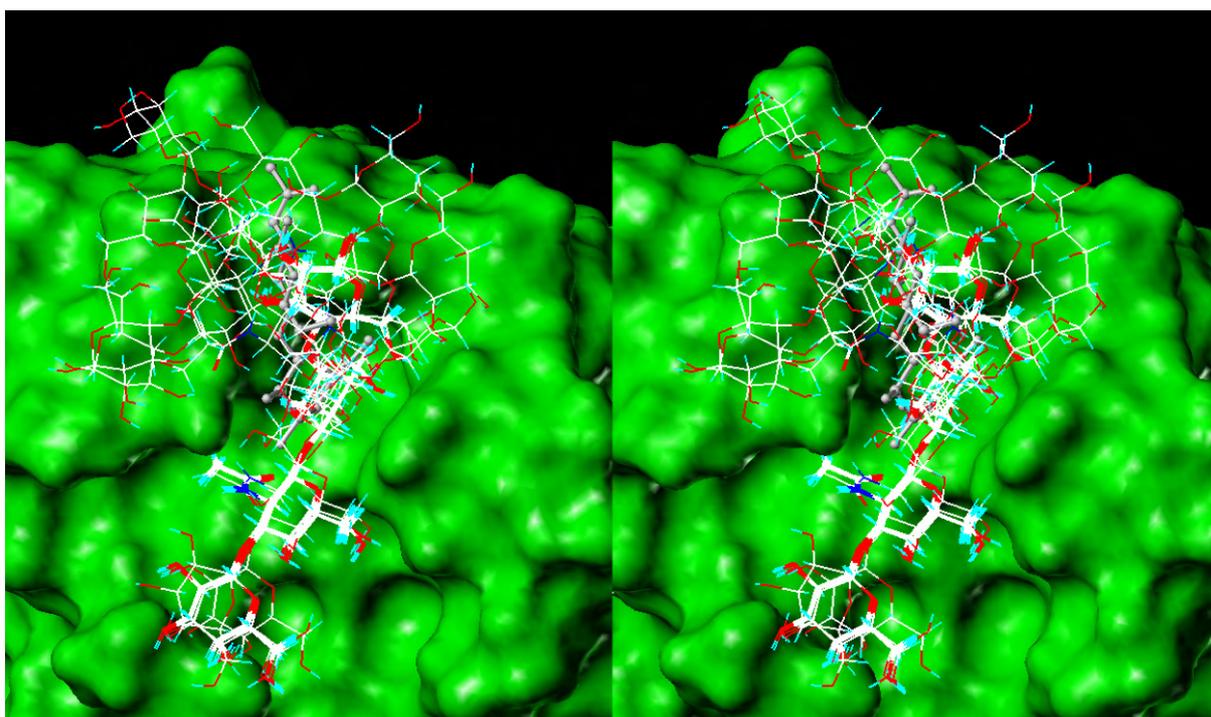


Figure 3.63: Complete resulting cluster obtained for LNT docked to galectin-1 (1sla) with Gal' as anchor fragment.

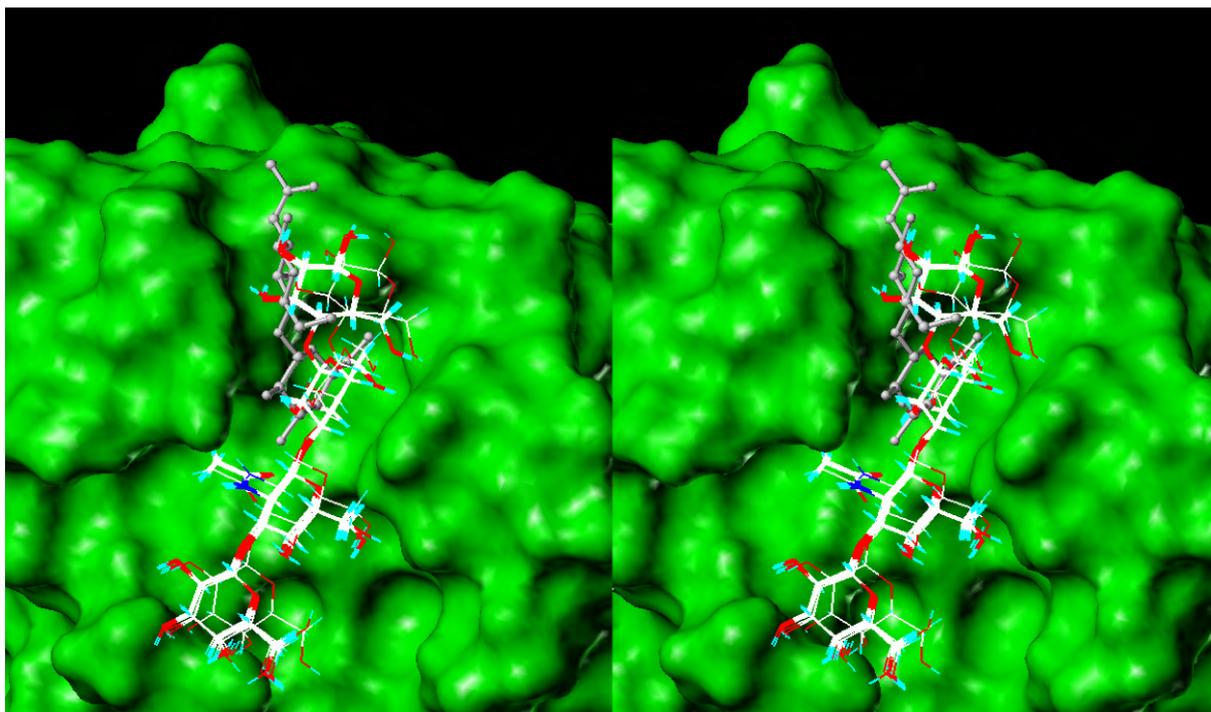


Figure 3.64: The top 14 poses of the docking, using Gal' as anchor, obtained a consistent binding mode. The Gal' residue is placed in a similar manner as in LacNAc (white, ball and stick rendering), but the torsion angles of the glycosidic bonds were unfavorable (Table 3.9).

In order to improve the proposals made by DOCK further constraints were applied to the docking setup. The results from STD NMR indicated a binding mode similar to that of lactosamines type I and type II, with the terminal galactosyl residue in the binding pocket of galectin-1. This assumption was made on the basis of the strong STD effect determined for H4'''. The non reducing end of LNT constitutes of a lacto-*N*-biosyl residue with the Gal β 1-3GlcNAc linkage. The binding mode of lacto-*N*-biose (LNB) was successfully determined from docking experiments with constraints from STD NMR. The new setup for docking was based on a pose of LNB, docked to galectin-1 from the crystal structure 1slt, in accordance with the binding mode of lactosamine and the STD NMR experiments (Figure 1.1). The positions of the spheres to which the ligand is docked were taken from the heavy atoms of the docked ligand. Especially the atoms of the glycosidic bond and the 6'-position of LNB were included into the cluster of spheres. The sphere enclosing grid box in which the ligand was to be placed was generated with a large extra margin of 12 Å. The initial sphere cluster included 14 spheres. This proved insufficient, DOCK exited with an error code. This was eliminated by adding spheres until a cluster of 20 spheres was achieved. The additional spheres were also created on the coordinates of heavy atoms from the docked LNB structure. This cluster of spheres was used to dock LNT to the binding site of galectin-1 from the crystal structure of 1slt. To test the general validity of this procedure additional docking experiments were carried out using *N*-acetyl lactosamine, extracted from 1slt and the model of LNB that had been used in the previous docking studies.

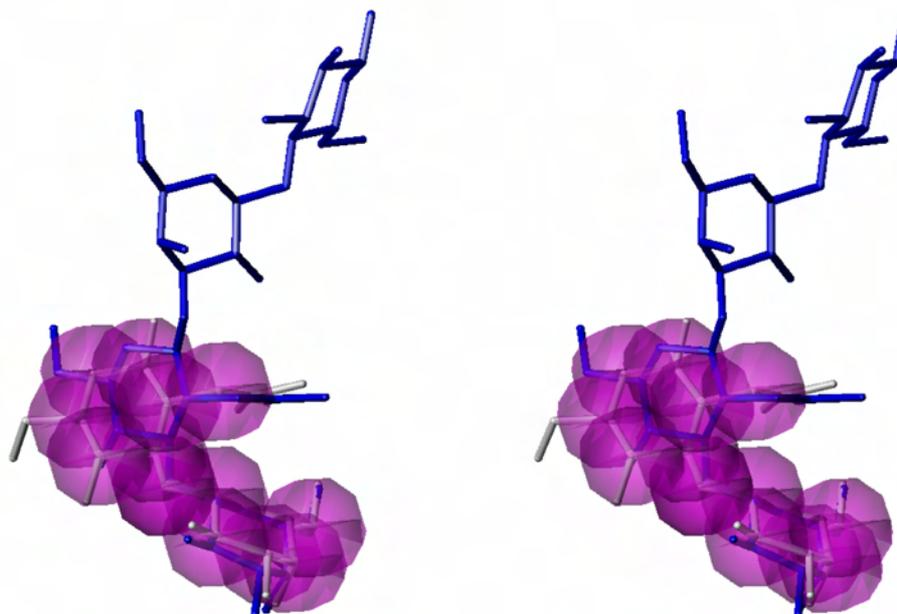


Figure 3.65: White: template structure of LNB; magenta: initial spheres generated for ligand matching; blue: starting structure of LNT, matched to the ring of the galactose residue of LNB.

The results achieved for LacNAc and LNB this way were supporting the validity of the setups as they were in agreement with the experimental data from crystallography and STD NMR. The resulting cluster for LNT with Gal''' set as the anchor fragment revealed high consistency with the binding modes of the disaccharides mentioned before. Significant differences of the proposed poses were obvious for the Glc and Gal' residues at the reducing end of the tetrasaccharide.

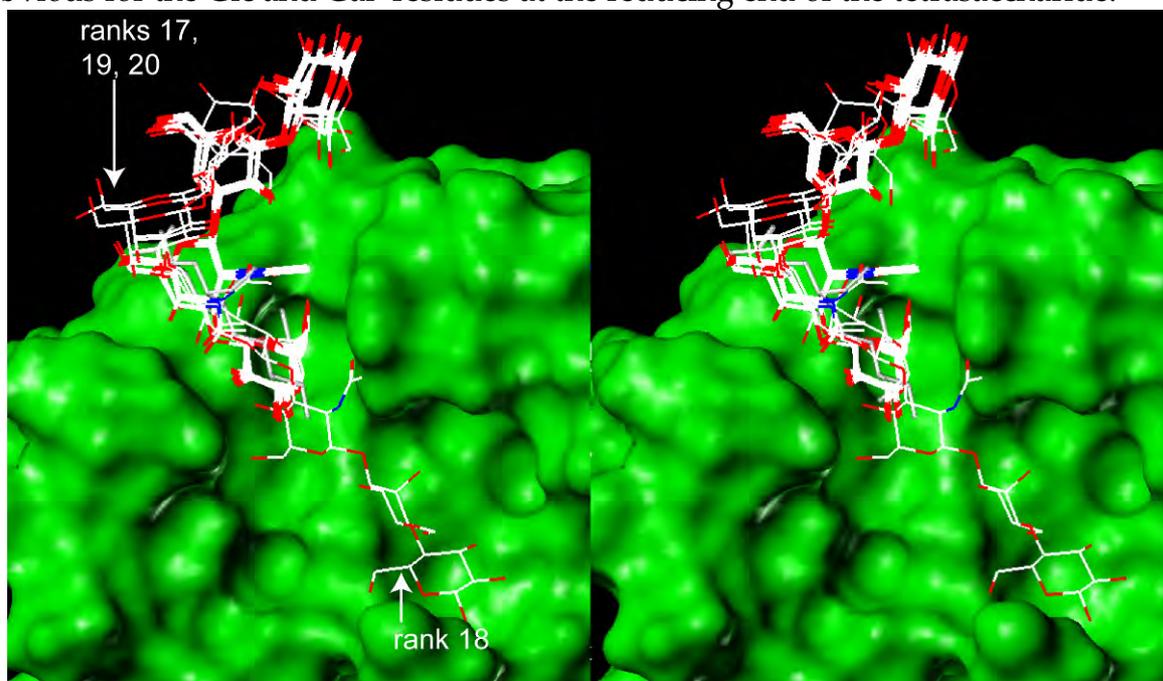


Figure 3.66: The result achieved from the LNT model with the terminal Gal''' residue as the anchor fragment revealed high consistency. Most of the results were in agreement with the binding mode of lactosamine (white, capped stick rendering). All hydrogen atoms omitted for clarity

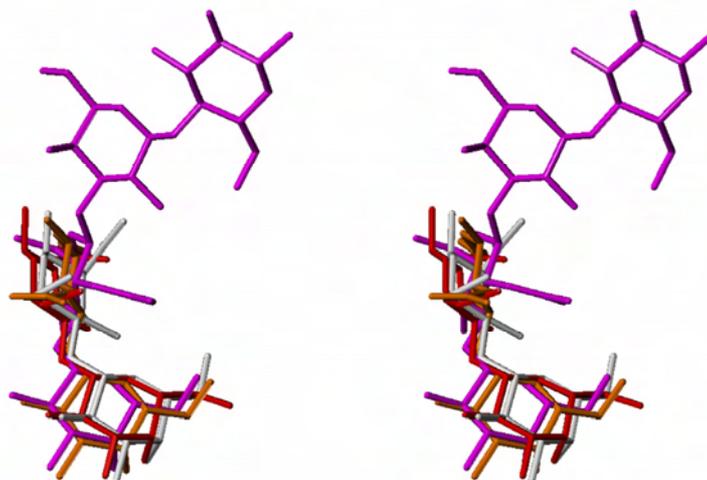


Figure 3.67: Results of the docking experiments with LacNac (red), LNB (orange) and LNT (magenta, Gal''' as anchor) performed with the novel sphere cluster. The position of the original ligand is shown in white. Shown are the first ranked poses for each molecule. All hydrogen atoms omitted for clarity.

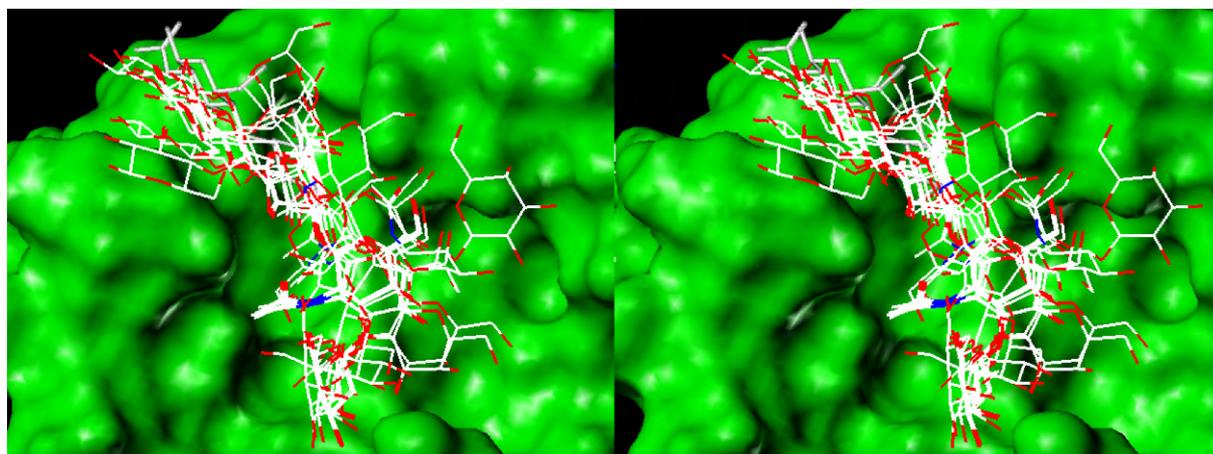


Figure 3.68: The result achieved for LNT with Gal' set as anchor fragment lacked the high consistency observed for LNT with Gal''' as anchor fragment. The position of LacNac in 1slt is shown in white and capped stick rendering. All hydrogen atoms omitted for clarity.

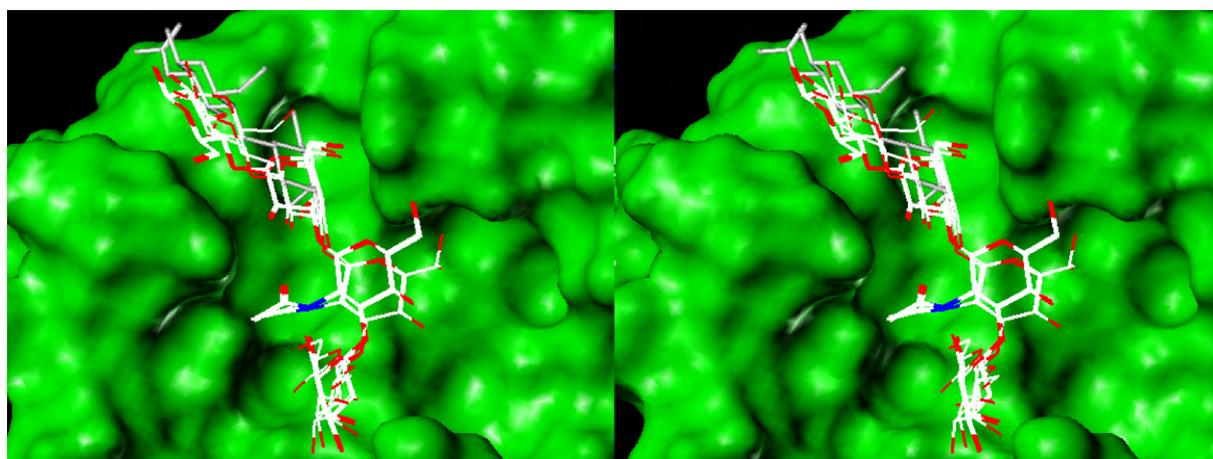


Figure 3.69: The top 13 poses of LNT docked to galectin-1 (1slt) with Gal' as anchor fragment were obtaining consistent orientations similar to that of LacNac (white, capped stick rendering). The poses ranked 12th and 13th were shifted in respect to the other proposals. All hydrogen atoms omitted for clarity.

In this docking study the poses in agreement with the binding mode of LacNAc adopted torsion angles of the Gal' β 1-4Glc bond of approximately $\phi=40^\circ$, $\psi=-5^\circ$ for the top 13 ranks. In the Gal''' anchored model the Gal''' β 1-3GlcNAc'' bond adopted values in the range of $\phi=65^\circ$ and $\psi=48^\circ$ for the top 16 poses in agreement with the expected binding mode.

In general, both models yielded less consistently defined torsion angles for the carbohydrate residues attached to the anchor fragment. The torsion angles between the carbohydrate residue docked into the binding site first (anchor fragment, p32) and the carbohydrate attached at the reducing end of the anchor fragment were well defined and in low energy conformation. In the model using the terminal Gal''' residue as anchor fragment the torsion angles between Gal''' and GlcNAc'' were equal for all poses, while torsion angles of the remaining glycosidic bonds were displaying somewhat higher variability. In the model using Gal' as anchor the torsion angles between Gal' and Glc were displaying equal values. The best ranked poses of this model also showed high consistency in the torsion angles between GlcNAc'' and Gal' of $\Phi=56^\circ$ and $\Psi=19^\circ$ approximately.

The high consistency of torsion angles of the glycosidic bonds in direction of the reducing end of the anchor is explicable by the docking procedure performed by DOCK. In the first step of the docking procedure, the ligand is matched to the spheres in the binding site. This applies to all atoms including those that are part of the glycosidic bond. The coordinates of the spheres were taken from the coordinates of the heavy atoms of the disaccharide lacto-*N*-biose docked to galectin-1 in agreement with the experimental data (chapter 3.6.4). During the placement of the anchor fragment and the subsequent reconstruction of the full molecule the ligand's atoms can be matched to these coordinates and remain these as long as the orientation of complete molecule does not form unfavorable interactions with the protein. Most of the remaining influence on the orientation and conformation of the ligand is based on the fitting of the remaining carbohydrates into the binding site. In the resulting cluster obtained for the Gal''' anchored model all poses are projecting the residues Gal' and Glc away from the protein surface.

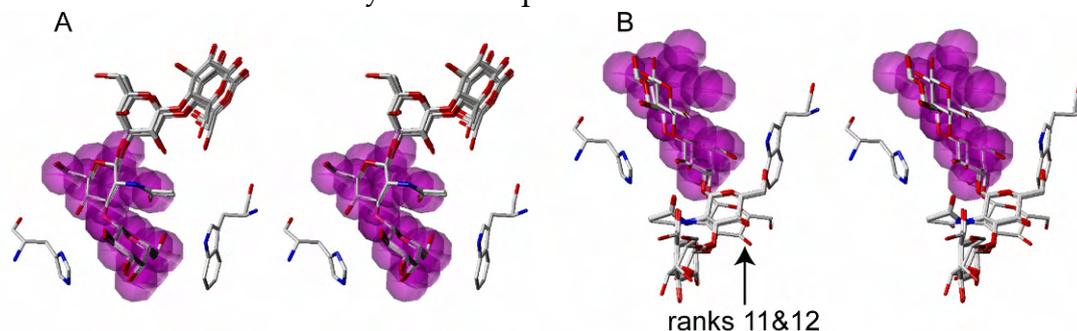


Figure 3.70: Comparison of LNT docked to galectin-1 (1slt, position of His52, left, and Trp68, right, shown for reference). The spheres used for docking are represented as transparent magenta. **A:** the 10 best poses of LNT using Gal''' as anchor fragment. **B:** the 12 best of LNT using Gal' as anchor fragment. The ten best poses are almost indistinguishable. Poses 11 and 12 are slightly shifted. Conformational diversity increases with the distance to the spheres in both approaches.

The top 10 poses obtained for the Gal' anchored model share an orientation where GlcNAc'' is placed in the same region as the Neu5Ac residue in the studies with sialyllactose derivatives (Figure 3.44). The terminal Gal residue at the non reducing end is then placed between Ala1 and His52 on the protein surface. The differences of the first 10 poses proposed by the docking of the Gal' anchored model are marginal. The two poses ranked 11th and 12th are also almost indistinguishable from each other. Most of the differences occur in the orientations and conformations of Gal'''. The poses ranked 11th and 12th displayed a counter clockwise rotation of the Gal' residue in comparison to the first poses.

Along with the similarity of the binding mode of LNT compared to LNB and LacNAc the binding modes were also satisfying the alignment of the Gal''' or Gal' residue respectively with the aromatic ring of Trp68. In the reference structure 1slt the angle of the normal planes of the aromatic residue of Trp68 and the ring atoms of Gal, which were defined as anchor, is 132.5°. The top ranked pose of the Gal''' anchored LNT model placed the Gal''' residue in a pose that yielded an angle between the normal planes, following the above definition, of 134.4°. The first ranked pose for the Gal' anchored LNT yielded an angle between the normal planes of Gal' and Trp68 of 134.2°. For comparison the plane angles of LacNAc and LNB docked with the same sphere cluster are given. For LacNAc the plane angle was 132.3° and for LNB the plane angle was 135.0°.

The high consistency of resembling docking poses is also revealed in the calculated STD effects for the respective docking poses.

Since the STD NMR spectra did not reveal unambiguous information regarding the binding mode, the epitopes calculated for the two models were employed to explain the experimental data from STD NMR. To further improve the conclusions from the comparison of docking result with experimental data the average values of the first ten docked poses for each model were averaged.

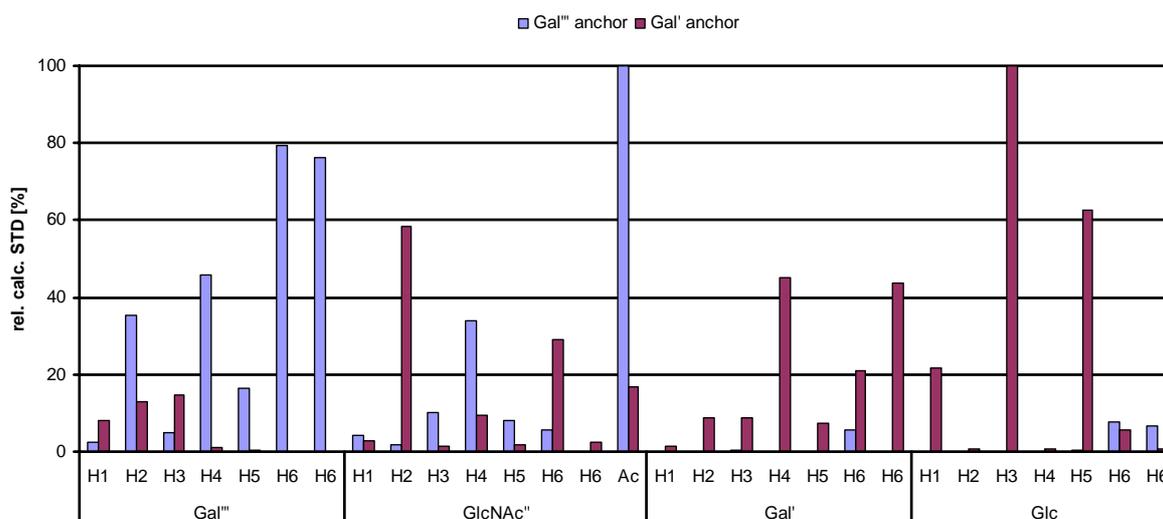


Figure 3.71: The comparison of the average calculated relative STD effects for both docked models revealed the different epitopes.

The epitopes derived from the averaged values of the top ten poses differ significantly. The poses generated for the model with the anchor on the ring atoms of Gal' displayed the strongest contacts on the terminal residues on the reducing end of the carbohydrate. GlcNAc'' and Gal''' are placed elevated over the protein surface and project into the direction of the residues Cys2 and Gly124. Gal''' is only weakly involved in the binding event, for the H6''' protons no contact was calculated. The profile of the calculated intensities on Gal' slightly resembles that of galactose residues determined before. H4' exposed stronger STD intensity than H5'. One of the H6' protons achieved significantly less STD effect than the second H6' and H4'. The strongest effects were determined for H3, H5 and H2''.

The poses generated for the model with the anchor on the ring atoms of Gal''' displayed the strongest STD effects on the *N*-acetyl group of GlcNAc'' and the 6'''-position. Gal' and Glc are projected away from the binding site, bending over the residues Gly69 and Ala70. Only the protons in 6-position of Glc are involved in the epitope. Except one of the H6-protons, influence of Gal' was insignificant. The profile of intensities found for the protons of Gal''' strongly resembled that already established for lactosamine disaccharides type I and type II. Both protons in 6'''-position were subject to almost equally distributed STD effects. H4''' displayed lower STD effect than H6''' protons and was significantly stronger than H5'''. The *N*-acetyl group of GlcNAc'' displays the strongest calculated STD effect. Additionally H4'' is significantly involved in the binding epitope.

Experimental data from STD NMR experiments with lacto-*N*-biose (Figure 3.12) indicated that in comparison to lactosamine H6 protons and *N*-acetyl group of the GlcNAc residue may display stronger STD effects as a result of the different orientation of the GlcNAc residue in the binding site. For the top ten poses H6'' of LNT did not display significant calculated STD effect on both protons. This may be a result of a shift in the placing of the ligand in the binding site as a result of additionally occurring interactions from the remaining carbohydrate residues. None of these averaged results is capable of explaining STD effects observed in the STD NMR experiments. STD NMR experiments revealed interactions with H4''' and H4' (Figure 3.59 and Figure 3.72). In the binding mode proposals derived from the top ten ranks significant STD effects on one of the respective protons of the galactose residues mutually eliminates interaction observed on the other H4 proton.

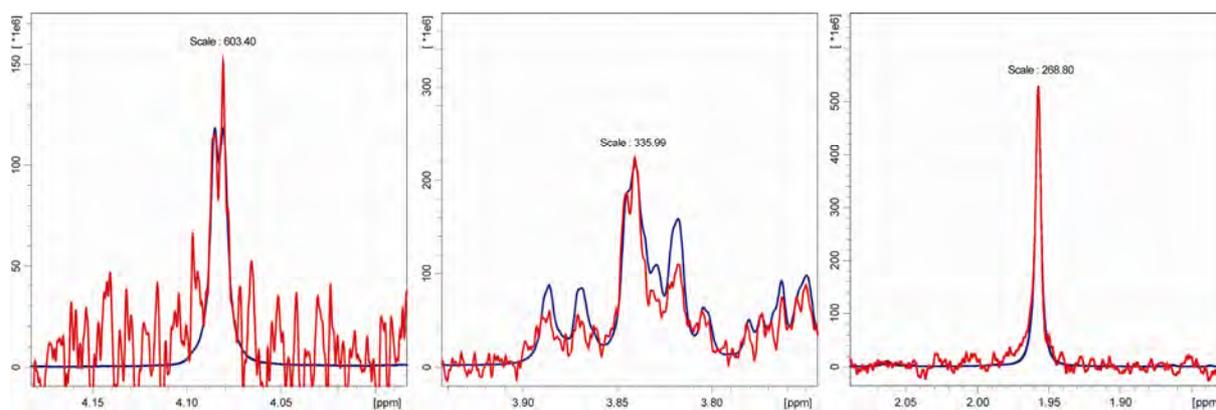


Figure 3.72: Only three signals of LNT could be assigned unambiguously in the 1D ^1H STD spectrum. From left to right (low field to high field): $\text{H4}'$, $\text{H4}'''$ and the *N*-acetyl resonance of GlcNAc'' . The scale shown in the graphs is needed to bring the intensity (signal height) of the difference spectrum (red) to the magnitude of the *off resonance* spectrum (blue).

Binding modes from the Gal' anchored model resulted in poses with strong contacts on $\text{H4}'$ and did not display contacts on $\text{H4}'''$. The results from the Gal''' anchored model generally obtained binding modes that resulted in strong contacts with $\text{H4}'''$. The pose ranked 18th was an exception. This binding mode differed significantly from binding modes proposed for LNT or other oligosaccharides. Gal''' is not buried in the binding site and the reducing end of the tetrasaccharide extend in the direction of Val5. This pose resulted in a strong contact on $\text{H4}'$ but minimal contact on $\text{H4}'''$.

Table 3.10: Relations of experimentally determined and calculated STD intensities of $\text{H4}'$, $\text{H4}'''$ and the *N*-acetyl group. Values were referenced to $\text{H4}'$ which gave the weakest signal in the STD NMR spectrum that could be assigned unambiguously.

	STD NMR	calc. Gal''' anchor	calc. Gal' anchor
$\text{H4}'''$	1.80	11.38	0.03
$\text{H4}'$	1.00	1.00	1.00
Ac	2.24	361.39	0.49

None of the proposals from the docking experiments could be aligned to the experimental results from STD NMR. Several reasons could be the cause for this. One reason could be that both binding modes proposed from the docking result occurred in the NMR sample. This hypothesis is contradicted by experimental data indicating that lactosamines bind 1.2 to 1.4 fold (type I) or more than 5 fold (type II) better to galectin-1 than lactose. The terminal disaccharide at the reducing end of LNT constitutes a lactose residue. The terminal disaccharide at the non reducing end of LNT constitutes type I lactosamine (LNB)^{106;112}. The difference in the STD intensities observed for the $\text{H4}'$ and $\text{H4}'''$ proton did not support this distribution of binding modes. The results compiled in Table 3.10 indicate that LNT occupies the binding site of galectin-1 using Gal''' only 15% of the time while Gal' is used 85%. This is in contradiction with the data published. The reason could be the analysis of the modeling data with the proportionality to r^{-6} in combination with inaccurate docking results. Errors in the docking results of ~40% can lead to an inversion of the proportions.

Intramolecular magnetization transfer may cause attenuation of strong STD intensities while non-existent or weak STD intensities may be amplified. This explanation was also not supported by the STD intensities observed. Taking the first ranked pose of the Gal''' anchored model as reference H4' is in vicinity to protons of the Gal' residue and H1, H2 and H6 of GlcNAc''. Based on the model none of these protons except H6'' should receive substantial saturation from the receptor. The H6'' protons are 5.2 and 5.6 Å away from H4' making effective saturation transfer unlikely. The docked structure displayed a low energy conformation with torsion angles of $\Phi=36.8^\circ$ and $\Psi=21.2^\circ$ for the GlcNAc1-3Gal linkage. Taking into account that the strongest STD intensity was found for the signal of the *N*-acetyl group of GlcNAc the proposals made from the Gal''' anchored model were in agreement with the STD results to a greater extent. For these models the docking results had the strongest computed effect on the acetyl group as found in the STD spectrum. This assumption is also backed by experimental data showing that the GlcNAc β 1-3Gal β 1-4Glc trisaccharide is only weakly binding to galectin-1 and galectin-3^{81;107} while LNT is exhibiting enhanced binding capacity to these proteins¹¹². Also oligomers of LacNAc (type II) residues do not seem to exhibit significantly altered binding affinity to galectin-1^{60;112}. Thus it was decided that lacto-*N*-tetraose is binding to galectin-1 via the terminal galactose residue at the non-reducing end of the tetrasaccharide. Furthermore, errors in the docking results may cause drastic deviations in the calculated STD effects (see also p49).

3.8 Molecular dynamics simulations of galectin-1 complexes

Some of the findings from STD NMR and docking experiments are not in agreement. To further investigate the binding mode of oligosaccharides bound to galectin-1 molecular dynamics (MD) simulations were applied. Several groups have studied the conformations of disaccharides in contact with various galactose specific lectins by combination of experimental data from NMR and MD simulations. M. Gilleron *et al.* studied the conformation of type I lactosamine and Gal β 1-2Gal β when binding to a galactose specific lectin from *viscum album*¹¹³. They found significant change of Φ torsion angle upon binding of Gal β 1-2Gal bound to the receptor in comparison to the binding mode when binding to galectin from chicken liver¹¹⁴. M. Ford *et al.* investigated the binding modes of lactosamine type I and type II, 3'-SO₃-LacNAc, silalylactosamine and di-LacNAc (Gal β 1-4GlcNAc β 1-3Gal β 1-4GlcNAc) in contact with bovine galectin-1. They also tested the binding behavior of the nonbinding carbohydrates GlcNAc and Glc β 1-4GlcNAc as negative test. Ford *et al.* used the crystal structure 1slt as starting point and included explicit water in either a droplet or a box with periodic boundary conditions (PBC) applied¹⁰⁷.

MD simulations are costly in respect to the time required to get a result especially in comparison with common docking procedures. However MD protocols have been applied in rational drug development process. The key applications of MD simulations are the refinement of ligand structures bound to their receptors and the assessment of receptor flexibility. MD simulations are often combined with docking

protocols where the starting structure is derived from docking and the results from MD simulations can be scored again by the docking software¹¹⁵.

In this study several strategies were applied to probe the conformation of a carbohydrate bound to galectin-1 and structural variations of the complex over time. It was assumed that certain findings from STD NMR experiments may become explicable with dynamic processes that docking experiments failed to resolve.

3.8.1 MD simulation of *N*-acetyl lactosamine and galectin-1

In this experiment the crystal structure 1slt was taken as a starting point for a MD simulation. For the MD simulation the X-ray structure was prepared by removing all atoms not belonging to galectin-1 and the carbohydrate. Hydrogen atoms were added to the protein and the ligand after validity of atom types was assured. Water molecules were added and Kollmann charges were applied to all atoms of the protein, charges for the ligand and water molecules were applied by Gasteiger method. The complex with solvent molecules was placed a box with periodic boundary conditions applied. The complete ensemble was minimized over 1000 steps using the Tripos force field. The minimized ensemble was taken as the starting point of the MD simulation over 100 ps, snapshots were taken every 50 fs.

Although the MD simulation setup allowed full flexibility for ligand and the receptor the overall structure of the ensemble did not change drastically, which results from the relative short simulation time. Simulation of 100 ps required approximately 54.1 h of processor time on two processors on a SGI Octane computer.

One of the unexplained findings when comparing STD effects and proton distances from ligand and receptor is the magnitude of STD intensity for H3 of GlcNAc. The structural analysis hinted that this contact should be among the strongest detected in the spectrum. The calculated STD effect from the X-ray structure was a result of a close contact of H3 with the beta protons of His52 over a distances of 2.6 Å and 3.9 Å. Additionally, H α and H δ of His52 have distances to H3 of 4.6 Å and 6 Å respectively. In the MD simulation the shortest distance averaged to 2.7±0.2 Å. Although the His52 residue was not displaying huge flexibility, the change of distances can explain the differences in the STD effects calculated from structure and acquired from STD NMR experiments. At a distance of 2.5 Å the calculated STD effect is 4.1*10⁻³ and decreases to 1.7*10⁻³ at a distance of 2.9 Å.

The trajectory displayed a stable complex of *N*-acetyl lactosamine in the binding site of galectin-1. Additionally the conformation of the disaccharide remained stable and the torsion angles in the expected range⁹⁰. The glycosidic torsion angles averaged to $\Phi=45.6\pm 8.8^\circ$, $\Psi=5.1\pm 11.7^\circ$. The torsion of the OH6'-group averaged to $\omega=71.1\pm 9.6^\circ$ which is in the order of the crystal structure but differing from the docking results. Ford *et al.* reported $\Phi=45.9\pm 9.3^\circ$, $\Psi=15.1\pm 8.5^\circ$.

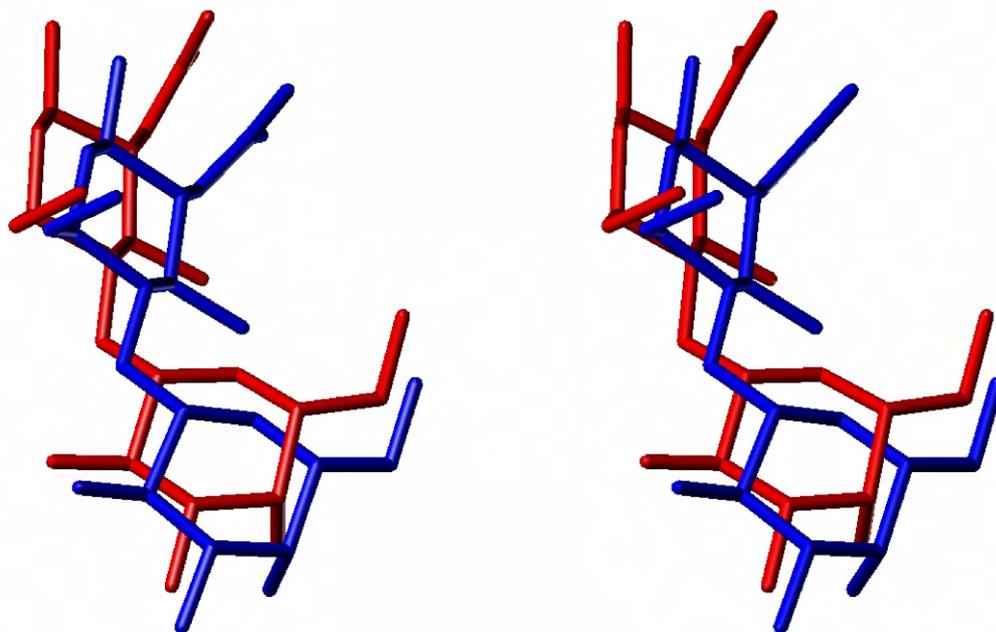


Figure 3.73: The comparison of the average structure of 2000 structures from the MD simulation (red) and the original pose of the ligand (blue) from X-ray crystallography reveals high similarity. The rmsd of the average structure is 1.1 Å. The Gal residue of the reducing end is at the bottom, the GlcNAc residue at the upper part of the picture.

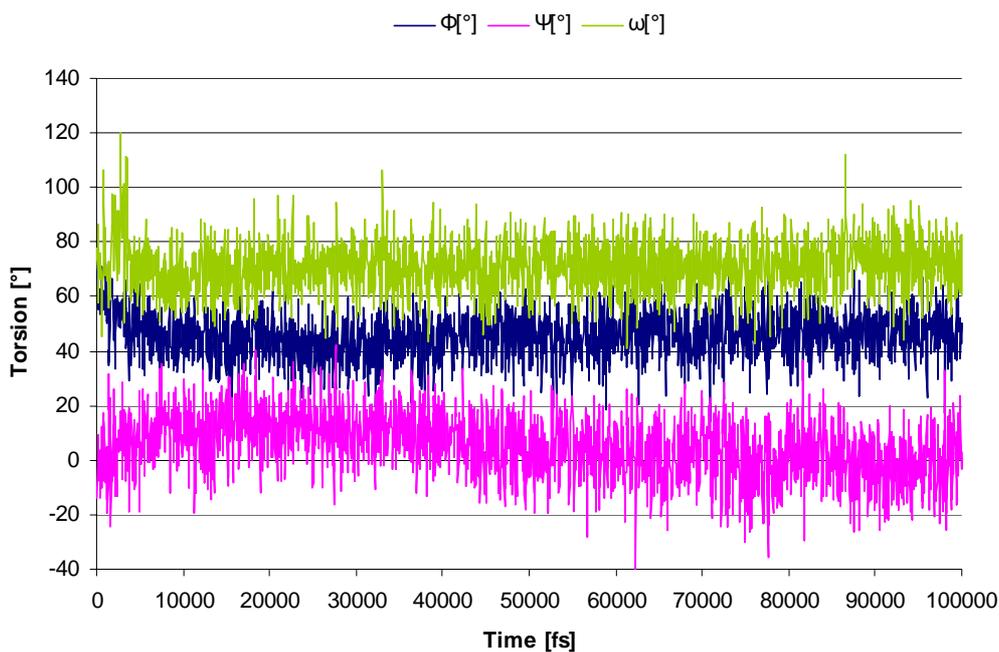


Figure 3.74: The torsion angles of the carbohydrate remained in the expected range. The MD was carried out in a box with PBC applied, containing 2900 water molecules at 300 K target temperature (avg.: 291 ± 4 K).

3.8.2 Complexes of galectin-1 with lacto-*N*-tetraose

The MD simulations were applied to more complex ligands bound to galectin-1.

The docking experiments with lacto-*N*-tetraose (LNT) did not explain the occurrence of STD effects for both protons H4''' and H4'. To test, if a binding mode exists that would allow contacts of both protons with protons of the receptor two MD simulations were conducted to check on the conformational flexibility of the ligand in the binding site. The ligand was taken from the docking experiment with the terminal galactose residue in the binding site^{81,107}. The starting structure equaled the first rank of the docking experiment with the terminal galactose docked in the binding site. The ensemble of ligand and receptor was minimized over 200 steps. In order to prevent the ligand from leaving the binding site the terminal galactose residue was also kept in its initial place.

The first MD simulation consisted of a 2 ns simulation with snapshots taken every 500 fs. To reduce computation time requirements the MD simulation was conducted without explicit water molecules and with rigid receptor. To reduce the effect of electrostatic interactions after omitting explicit water the dielectric constant was set to $\epsilon=65$. Calculation of the simulation required 71 h computation time on two processors.

The MD simulation displayed stable torsion angles over most of the simulation time. Over the simulation the average temperature was 293.2±17.7 K. The average torsion angles are given in the following table. The Φ torsion angles remained in acceptable ranges of 58° to 71° the Ψ torsion angles assumed angles ranging from -36° to 53°.

Table 3.11: Torsion angles of the input structure and the average MD simulation and expected⁹⁰ values for comparison.

	Φ Gal'''- GlcNAc'' [°]	Ψ Gal'''- GlcNAc'' [°]	Φ GlcNAc'- Gal' [°]	Ψ GlcNAc'- Gal' [°]	Φ Gal'- Glc [°]	Ψ Gal'- Glc [°]
expected	50	10	50	10	40	15
input	65.6	47.8	36.8	21.2	44.0	-17.1
MD avg.	67.96	53.52	58.71	43.43	79.29	-36.80
MD dev.	7.51	7.57	20.84	44.36	49.33	29.02

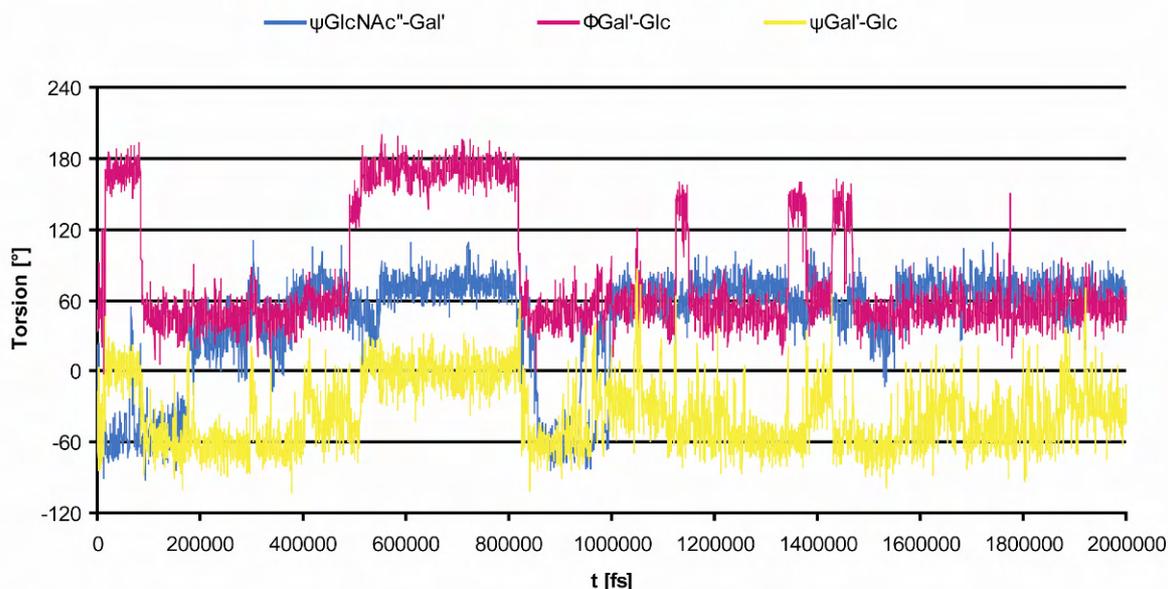


Figure 3.75: Trajectory of the torsion angles displaying the highest flexibility over 2 ns simulation time. The remaining torsion angles remained more stable in the course of the simulation (Table 3.11).

One of the reasons for the failure to reproduce reasonable torsion angles may be the input conformation from the docking experiment. MD simulations usually are not long enough to allow a system to explore the conformational space to find global minima. Additionally, keeping the position of the Gal''' residue fixed may have had an unfavorable influence. The change of orientation in the remaining carbohydrate could not be compensated by movement of this residue. During the simulation H4' was frequently pointing into the direction of Ala70 or Arg73, allowing a possible contact. The angle of the normal planes evaluated to $\theta=43.2^\circ$ and the centroids of Gal''' and Trp68 were 5.7 Å separated. Since Gal''' was not allowed to move freely in the course of the simulation these results are of limited use.

To test whether different starting conformation simulation setup would influence the result of the MD simulation a second simulation experiment was started. The initial structure of the ligand also was a pose from the docking result and achieved rank17. The altered torsion angles of the molecule project the residues Gal' and the terminal Glc at the reducing end in a direction bringing them closer to His52 and Arg73. In this MD simulation the Gal''' residue was defined flexible, only atoms of the receptor were kept rigid. The setup was also altered in setting of the dielectric constant to $\epsilon=20$. The enhanced influence of electrostatic interaction was thought to keep the ligand in the binding site. The remaining setup was equivalent to the previous simulation setup and the simulation length was set to 2 ns. The calculation required 83 h computation time on a two processor SGI Octane.

In the resulting MD simulation, the ligand did not leave the binding site of the receptor. Again the torsion angles of the glycosidic linkages displayed unusual values. Immediately after start of the simulation the Gal''' β 1-3GlcNAc'' and Gal' β 1-

4Glc Ψ torsion angles tended towards an angle of -60° . After ca. 75 ps simulation time the GlcNAc'' β 1-3Gal' Φ torsion angle rapidly changed from ca 60° to values around 180° . Obviously the change of parameters did have an effect on the MD simulation result but not in the anticipated way. In the trajectory some poses were found that did exhibit contacts from H4' to receptor protons less than 6 \AA away. In the averaged structure three protons of the receptor were less than 6 \AA away from the receptor. The angle of the normal planes of Gal''' and Trp68 evaluated to $\theta=143.2^\circ$ and the distance of the centroids was $R=4.92 \text{ \AA}$.

The simulation results did not represent low energy conformations of the glycosidic bonds. The MD simulation was carried out with the Tripos force field. The tendency to change glycosidic torsion angles resulting in unfavorable conformations indicate that the Tripos force field is not capable of handling a bimolecular system consisting of a tetrasaccharide in complex with a protein. Since the force field does not seem capable to handle flexibility and conformational preferences of carbohydrates the approach to simulate LNT in complex with galectin-1 needed further adjustments.

Table 3.12: The torsion angles of the glycosidic bonds in LNT over 2 ns MD simulation and expected⁹⁰ values for comparison.

	Φ Gal'''- GlcNAc'' [°]	Ψ Gal'''- GlcNAc'' [°]	Φ GlcNAc''- Gal' [°]	Ψ GlcNAc''- Gal' [°]	Φ Gal'- Glc [°]	Ψ Gal'- Glc [°]
expected	50	10	50	10	40	15
input	34.1	36.0	31.7	31.9	59.3	4.1
MD avg.	47.90	-59.27	174.29	56.89	56.43	-48.02
MD dev.	8.47	14.82	23.27	15.68	34.00	27.54

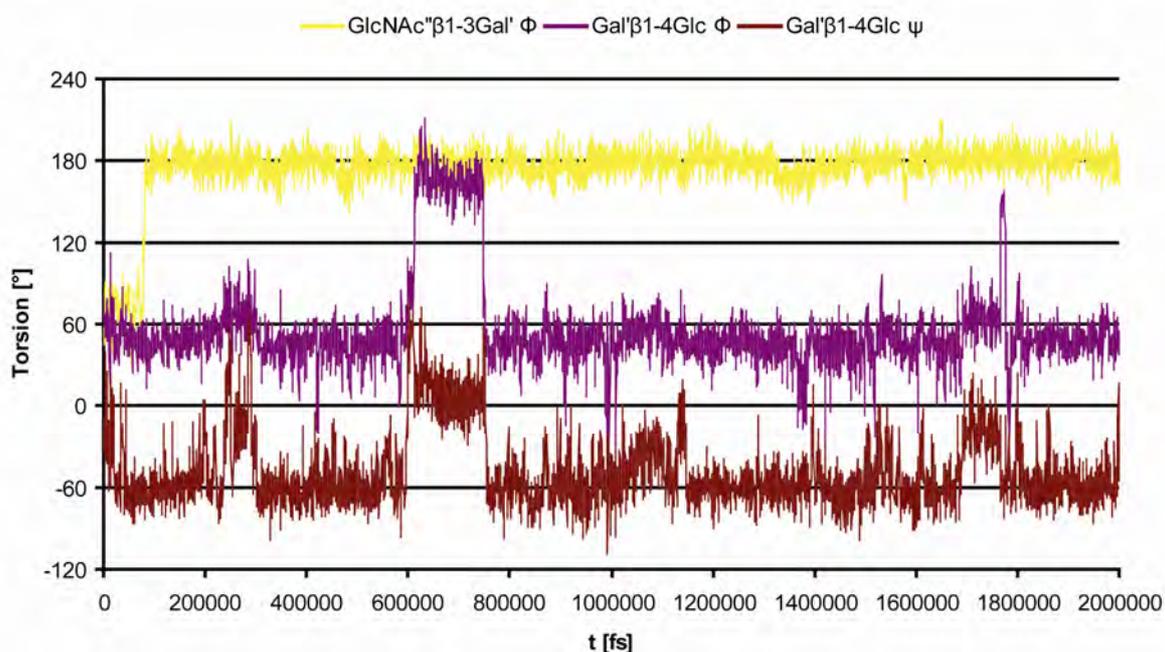


Figure 3.76: Trajectory of the torsion angles displaying the highest flexibility over 2 ns simulation time. The remaining torsion angles remained more stable in the course of the simulation (Table 3.12).

3.8.3 Constrained MD simulations of LNT in complex with galectin-1

It became obvious that the Tripos force field⁸⁹ used for the MD simulations was not able to properly handle glycosidic torsion angles in the tetrasaccharide LNT. Several attempts were made to include constraints on the torsion angles of the carbohydrate in order to maintain low energy conformations over the simulation.

Torsion angle constraints of $0.075 \text{ kcal/mol}[\text{°}]^2$ and $0.05 \text{ kcal/mol}[\text{°}]^2$ in two different simulation setups proved to be too strong. The torsion angles of the carbohydrates were only deviating over 2.7° and 3.5° respectively. This did not represent the expected flexibility and did not allow any additional contacts between carbohydrate and protein to arise from ligand flexibility. The simulation mainly consisted of rigid body movements. This even resulted in the dissociation of the complex in the second constrained simulation ($0.05 \text{ kcal/mol}[\text{°}]^2$ angle constraint). In the first constraint simulation an additional distance constrained of $5 \text{ kcal/mol}\text{Å}^2$ was employed between O6''' of LNT and a proton of the side chain amide of Asn61 to keep the ligand in the binding site. This distance constraint was omitted in subsequent simulations of the complex because it was assumed that it might interfere with possibly interesting rearrangements of the complex. The starting conformation and orientation of the ligand was taken from the respective MD simulation carried out before. The respective ligand structure was accepted as starting structure if low energy conformation and orientation in alignment with the ligand orientation from the X-ray structure were fulfilled.

Table 3.13: The results from the constrained MD simulations exhibit only minimal variations of the torsion angles.

	$\Phi\text{Gal}'''$ - GlcNAc''	$\Psi\text{Gal}'''$ - GlcNAc''	$\Phi\text{GlcNAc}''$ - Gal'	$\Psi\text{GlcNAc}''$ - Gal'	$\Phi\text{Gal}'$ - Glc	$\Psi\text{Gal}'$ - Glc
expected	50	10	50	10	40	15
input (2ns, $0.075 \text{ kcal/mol}[\text{°}]^2$)	50	10	50	10	40	15
MD avg. (2ns, $0.075 \text{ kcal/mol}[\text{°}]^2$)	50.8	10.4	50.6	10.5	41.4	15.0
MD dev. (2ns, $0.075 \text{ kcal/mol}[\text{°}]^2$)	2.6	2.7	2.7	2.9	2.7	2.8
input (1 ns , 0.05 $\text{ kcal/mol}[\text{°}]^2$)	48.9	25.9	52.5	33.9	63.7	-27.8
MD avg. (1 ns , $0.05 \text{ kcal/mol}[\text{°}]^2$)	49.8	12.1	50.8	10.6	42.0	14.9
MD dev. (1 ns , $0.05 \text{ kcal/mol}[\text{°}]^2$)	3.2	3.3	3.5	3.6	3.4	3.7

Since the penalties imposed on torsion angle were set too high to allow significant flexibility of the glycosidic bonds the penalty was drastically reduced to $0.005 \text{ kcal/mol}[\text{°}]^2$. With this setup torsion angle deviation of 30° resulted in a penalty of 4.5 kcal/mol . It was assumed that this would allow enough flexibility for the glycosidic bond to sample interesting conformations of the ligand. In order to optimize the binding site during the MD simulation the residues Leu31, ⁴⁴HFN⁴⁶, ⁵¹AHGD⁵⁴, ⁵⁹VCNSK⁶³, ⁶⁸WGAEQR⁷³ and Arg111 were defined as flexible while the

remaining part of the protein was kept rigid. 1 ns simulation time required 66 h computation time on two processors.

In the course of the simulation the glycosidic torsion angles were more flexible as in the previously described constrained MD simulations. Although an equal penalty was imposed for all torsion angles in the tetrasaccharide in this simulation Φ torsion angles showed significantly less deviation than the Ψ torsion angles. Φ torsion angles deviated about 7.5° and Ψ torsion angles deviated about 12.9° . In the previous simulations the deviations were almost equal for all torsion angles. This proves an individual evaluation of changes in the conformational features of the ligand.

Analysis of the trajectory revealed a dislocation event of the ligand from the binding site similar to the prior MD simulation with $0.05\text{kcal/mol}[\text{^\circ}]^2$ penalty applied. From 91.5 ps to 380 ps simulation time the tetrasaccharide left the binding site and resided on the protein surface. After 381 ps simulation time the ligand was again positioned in the binding site and was in alignment with the reference ligand of the crystal structure 1slt. Although the ligand was displaced over a period of 579 frames, corresponding a time of 289.5 ps simulation time, the averaged structure over the complete trajectory was in general agreement with the X-ray structure

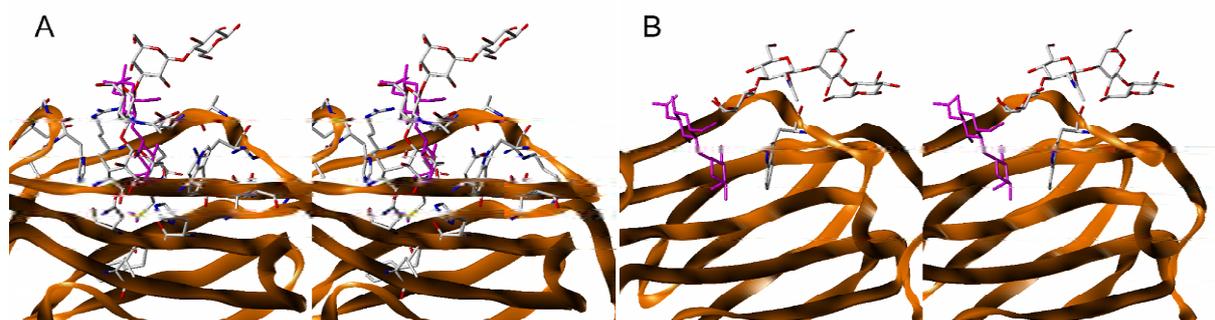


Figure 3.77: A: The starting structure of the constrained MD simulation of LNT in complex with galectin-1 is in agreement with the binding mode of LacNAc (magenta). The protein is shown in ribbon representation (orange) except for the amino acids that were kept flexible in the simulation. B: The position of LNT differed massively from the reference position of LacNAc (magenta) during the displacement phase. This snapshot was taken after 236 ps simulation time. The protein is shown in ribbon (orange) except for Trp68 for orientation. The placement of the ligand in the X-ray structure is shown in magenta.

The trajectories of the glycosidic torsion angles were not notably different during the displacement event compared to the frames where the ligand was bound to the protein in the expected mode. In the following analysis of the binding mode of LNT in complex with galectin-1 the structures of this period with LNT outside the binding site were also analyzed.

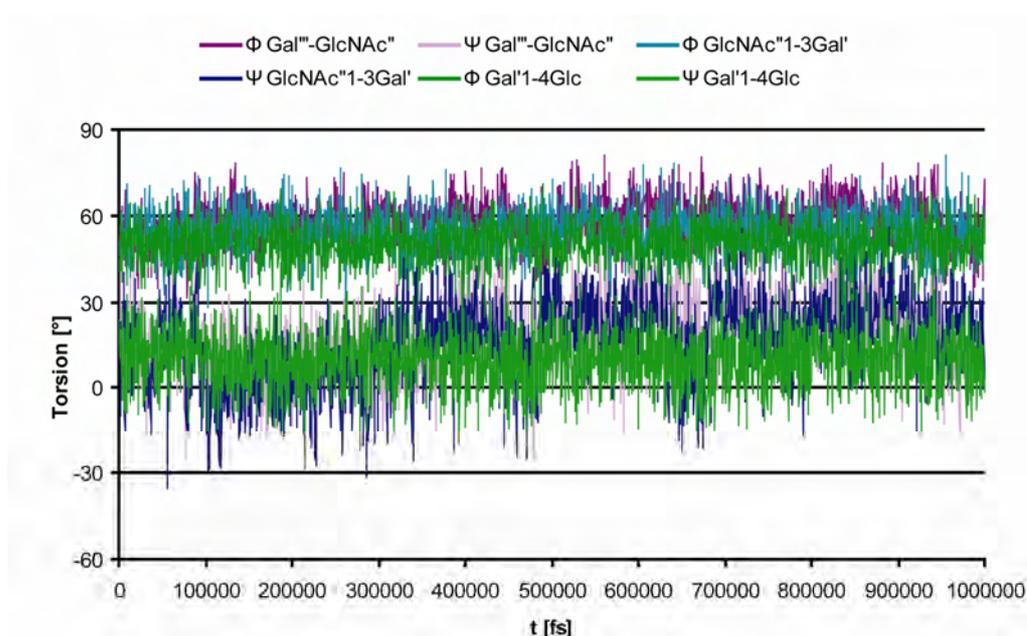


Figure 3.78: The trajectory of the glycosidic torsion angles reveals the preference for the expected values while broad flexibility is still notable.

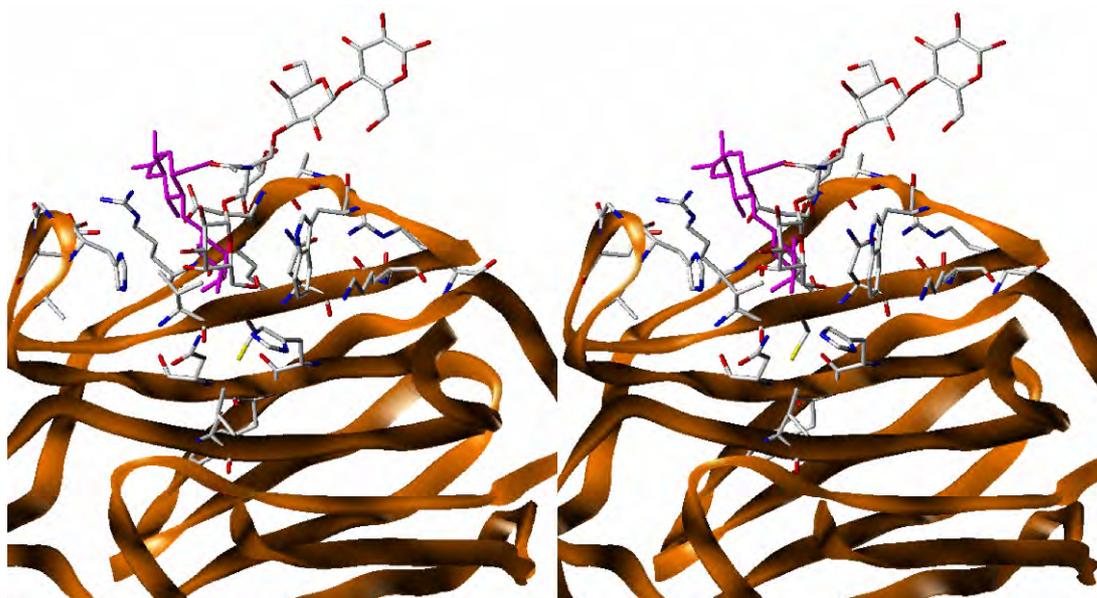


Figure 3.79: In the final structure of the MD simulation the tetrasaccharide is located in the binding site of LacNAc (magenta) but is oriented differently. The protein is shown in ribbon representation (orange) except for the amino acids that were kept flexible in the simulation.

Table 3.14: Average torsion angles determined for the MD simulation with $0.005 \text{ kcal/mol}[\text{°}]^2$ penalty for torsion angle variations. The values were averaged over the complete simulation time of 1 ns.

	$\Phi \text{Gal}'''\text{-GlcNAc}''$ [°]	$\Psi \text{Gal}'''\text{-GlcNAc}''$ [°]	$\Phi \text{GlcNAc}''\text{-Gal}'$ [°]	$\Psi \text{GlcNAc}''\text{-Gal}'$ [°]	$\Phi \text{Gal}'\text{-Glc}$ [°]	$\Psi \text{Gal}'\text{-Glc}$ [°]
expected	50	10	50	10	40	15
input	51.52	13.74	51.24	13.06	40.85	15.17
MD avg.	57.50	18.63	54.96	16.53	50.55	9.88
MD dev.	7.84	13.50	7.63	15.02	7.02	10.11

To assess whether conformations of the MD trajectory were displaying distances of H4' to protons of the receptor justifying the effects detected in the STD NMR experiments two clusters of ligands were created. The first cluster of ligand poses was created on the basis of the potential energy calculated for the respective time frame of the MD simulation. For analysis of both clusters the proton coordinates of the protein were taken from the protein structure averaged over the simulation time.

The design of the first cluster of results was intended to simulate the outcome of a force field based energy scoring as implemented in DOCK. In the course of the simulation the potential energy of the complex ranged from 270.74 kcal/mol to 502.52 kcal/mol. The lowest potential energy was calculated for the starting structure of the MD simulation. The cluster of structures comprised the 15 best potential energies calculated for the complex during the MD simulation, including the starting structure and the five worst potential energies of the MD simulation.

The analysis of the results did not give conclusive indications capable to explain the results from STD NMR. The averaged results obtained for the structures selected before gives a weak calculated signal for H4'. The magnitude of this signal is unproportional to the magnitude of H4''' which is 17.2 times stronger. STD NMR results displayed a 1.6 times stronger signal for H4''' in comparison with H4'. The averaged values of the cluster show a 1.2 times stronger signal for the N-acetyl group in comparison with H4''. In the STD NMR analysis the N-acetyl group gives a 1.06 times stronger signal than H4''. Since the results from the structures averaged over the complete MD simulation also included information of the binding epitopes during the displacement phase the analysis was also performed for the starting structure of the MD simulation and the final structure of the simulation. The calculated STD intensities for the structures from the start and the end of the simulation were differing from the epitopes determined for derivatives of lactose and were not displaying the pattern of intensities for H4''' and H4' determined by STD NMR.

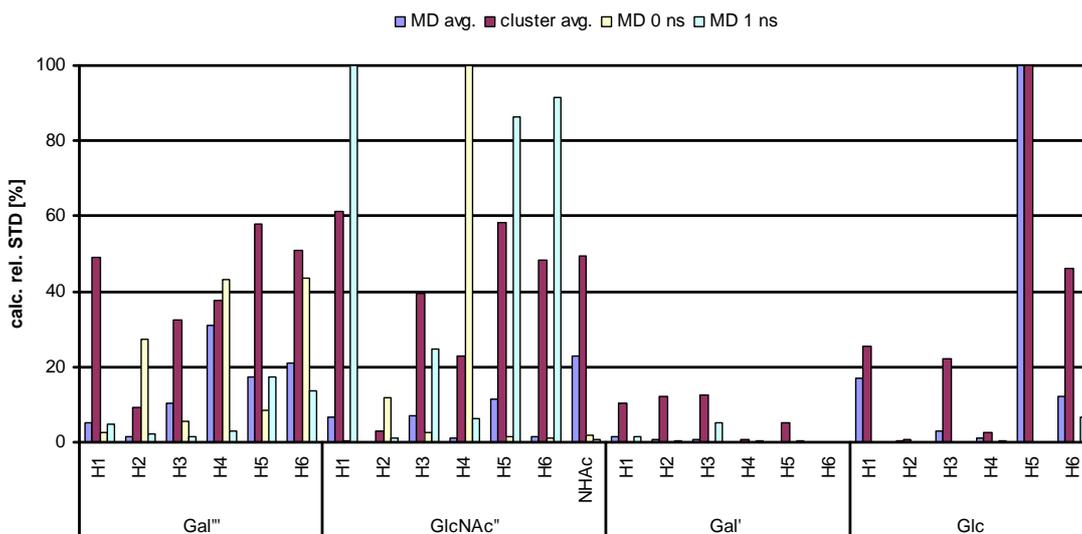


Figure 3.80: The analysis of the average structure of the complete MD simulation (MD avg., blue) and the averaged values obtained from the structures selected by potential energy of the MD simulation (cluster avg., red) reveals different binding modes. Both binding modes are differing from the common motif of LacNAc with strong intensities on H4 and H6 of the terminal galactose residue at the non reducing end. Both epitopes also fail to reproduce the result from STD NMR with the strongest intensity found on the *N*-acetyl group. For comparison the binding epitopes calculated from the structure at the start of the MD (MD 0 ns, yellow, Figure 3.77A) and the final structure (MD 1 ns, cyan, Figure 3.79) are given.

The second cluster of ligand structures was created by sampling the ligand structure every 50 ps in the MD simulation starting with the first structure at 0 ps simulation time. This cluster also included the ligand structure averaged over the MD simulation. The results obtained from these ligand structures were not able to explain the results from STD NMR. The resulting calculated epitopes neither resembled the epitopes that are common for lactose derivatives in complex with galectin-1 nor were they capable to explain the results acquired using STD NMR. The average results exhibited the strongest calculated intensity at the H5 position of the glucose residue at the reducing end of the tetrasaccharide. The pattern of intensities calculated for the terminal galactose at the non reducing end of the tetrasaccharide is also differing from results previously determined in this study. The strongest intensity for Gal'' is calculated for H5 followed by H6 and H1. No significant STD effect could be determined for H1'' in the STD NMR experiments. The averaged result obtained for the results in this cluster of structures resembles that of the result obtained for the structure averaged over the MD simulation. The explanation for this occurrence is based in the compilation of the structures which resembles a low resolution sampling of the MD simulation.

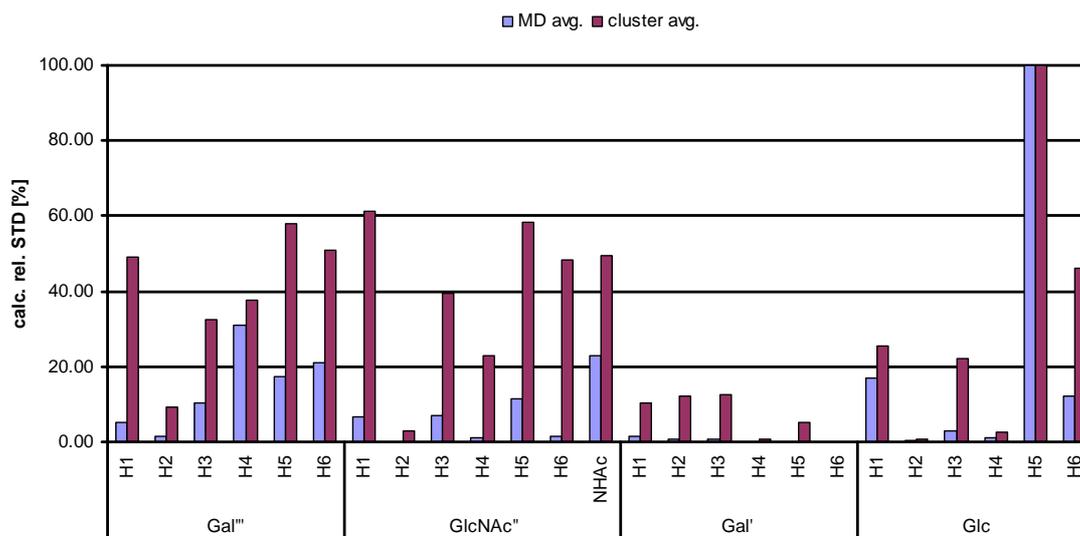


Figure 3.81: The averaged result obtained by taking structures in periodical intervals (red, cluster avg.) resembles the averaged structure over the complete MD simulation (blue, MD avg.).

The possible reasons for the inability to reproduce the experimental findings can be manifold. The MD simulations omitted explicit water molecules. The experiments carried out with *N*-acetyl lactosamine in a water box displayed high stability of the complex and torsion angles within the expected ranges. The combination of a ligand with more degrees of freedom and substitution of explicit water by choice of a rational dielectric constant may have caused the occurrence of unfavorable conformations. This effect may have been reinforced by the application of uniform constraints for all glycosidic torsion angles. These circumstances led to a binding mode that was similar to that expected from docking and other results but displayed notably different calculated STD effects than those determined experimentally in this work. In order to properly handle the complex simulation of a potentially highly flexible ligand in complex with the receptor more experimental data on ligand conformation bound to the receptor might be helpful. The transferred Nuclear Overhauser Effect (trNOE) allows the determination of intra ligand proton-proton distances. Based on trNOE constraints the conformation of the bound ligand can be modeled and be used as input to further docking or MD simulation experiments. Additionally, the inclusion of explicit water molecules should be considered for modeling approaches and in the analysis of possible saturation transfer pathways. Furthermore, it should be considered that the results from STD NMR could indicate an additional, independent binding mode of the ligand in contact with the protein. Whether this binding mode occurs in the same binding site or at a different position on the receptor surface must be confirmed by other experimental procedures.

3.8.4 Simulation of the complex of sialyllactose and galectin-1

The docking results obtained for sialyllactose bound to galectin-1 were not able to explain the strong intensity found for the *N*-acetyl group in the STD NMR spectrum. Additionally, the orientation of the Neu5Ac residue was not in accordance with the findings of Ford *et al.*¹⁰⁸.

The setup strategy for the MD simulation of sialyllactose bound to galectin-1 followed the setup applied to the MD simulations of lacto-*N*-tetraose bound to galectin-1. The simulation time was set to 2 ns, the dielectric constant to $\epsilon=20$, atoms of the receptor were kept rigid while all atoms of the ligand were kept flexible. The target temperature was set to 300 K and snapshots of the structure were taken every 500 fs. The input structure of the ligand was the first ranked pose of sialyllactose docked to galectin-1 with histidine residues His44 and His52 protonated at the N δ position as in the initial docking procedure.

Over the simulation time the ligand remained in the binding site. The average temperature was 293.7 ± 18.3 K. The ligand occupied two main orientations. The first orientation corresponded to the input orientation. In this orientation the Glc residue projected in the direction of Ala70 and the overall distance of the ligand to His52 is increased. The second orientation is characterized by the projection of Glc towards the position of Arg73 and increased vicinity to His52. The torsion angles again were subject to change in the course of the simulation, especially the Gal β 1-3Glc ψ torsion angle. After ca. 50 ps the torsion angle alters from ca. 0° to ca. -60° , after 570 ps the torsion angle altered again yielding ca. 180° . The Neu5Ac α 2-3Gal Ψ torsion angle rotated from the unexpected value of 27° to -40° . This value is expected in combination with a torsion angle of $\Phi=-156^\circ$, but the average Φ angle was -71° . Some conformations computed by the MD simulation were displaying a Neu5Ac α 2-3Gal torsion of $\Phi < -120^\circ$.

In the average structure of the MD simulation the carboxyl group of the neuraminic acid residue points towards the bottom of the receptor binding site placing the oxygen atoms in distances of 3.8 Å and 3.6 Å relative to N ϵ of His44 respectively. This would allow the formation of a hydrogen bond if N ϵ was protonated. The distances to the proton bound to N δ were 4.7 Å and 5.2 Å respectively. In this pose the acetyl group of the Neu5Ac residue points into the direction of Cys2 and Gly124 with distances of ca. 4.5 Å. In the averaged structure a total of nine protons of the receptor were within a distance of 6 Å away from protons of the acetyl group, two of these contacts spanned less than 5 Å. In this pose the axial H3'' of the Neu5Ac residue is pointing away from the receptor surface while the equatorial H3'' proton is facing in the direction of His52, namely H δ , and the side chain of Ala51. A total of 13 receptor protons were found within 6 Å distance. One of these contacts was less than 4 Å and five less than 5 Å away from the receptor allowing transfer of saturation. The angle of the normal planes of Gal' and Trp68 resulted to $\theta=154.5^\circ$ and the distance of the respective centroids to R=4.95 Å. In this

simulation the galactose ring was getting close to a parallel alignment with the indole ring of Trp68 while still remaining in the expected distance from the ring system. In the course of the simulation some of the resulting ligand poses allowed the transfer of saturation from protein to the *N*-acetyl group of the Neu5Ac residue.

Table 3.15: The torsion angles of the glycosidic bonds in SiaLac over 2 ns MD simulation and expected⁹⁰ values for comparison.

	Φ NeuAc''-Gal' [°]	Ψ NeuAc''-Gal' [°]	Φ Gal'-Glc [°]	Ψ Gal'-Glc [°]
expected	-156/-74	-24/0	40	15
input	-92.9	27.1	54.7	-37.8
MD avg.	-70.34	-39.87	64.48	116.80
MD dev.	14.04	18.47	14.53	117.82

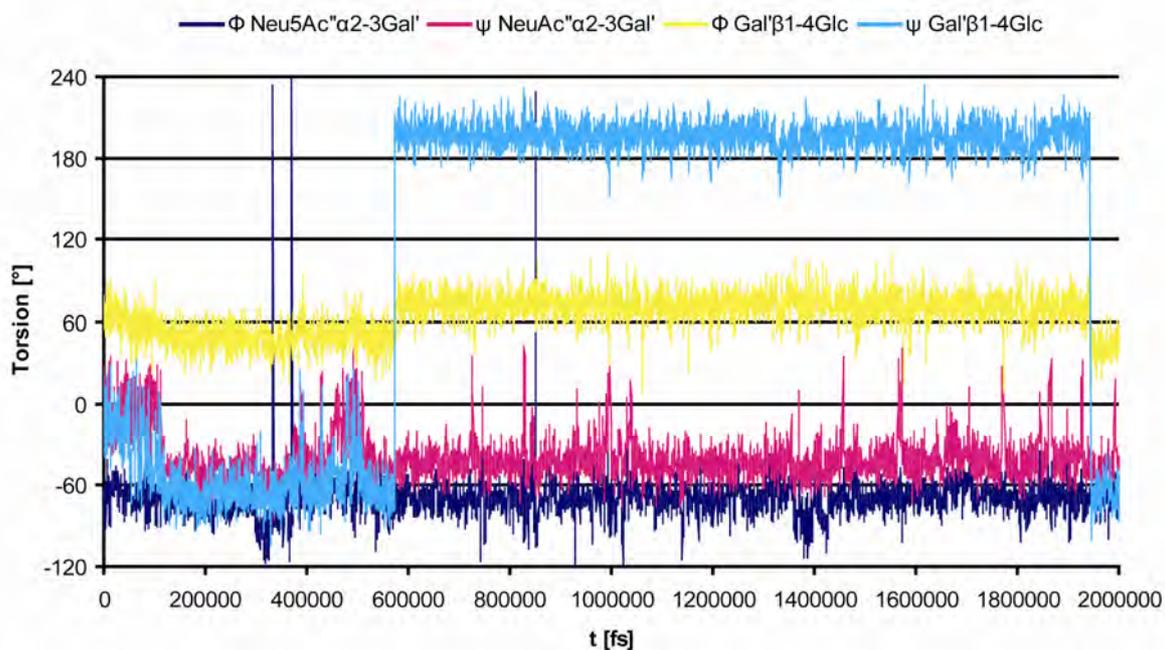


Figure 3.82: Trajectories of the glycosidic torsion angles in SiaLac over 2 ns MD simulation of SiaLac bound to galectin-1.

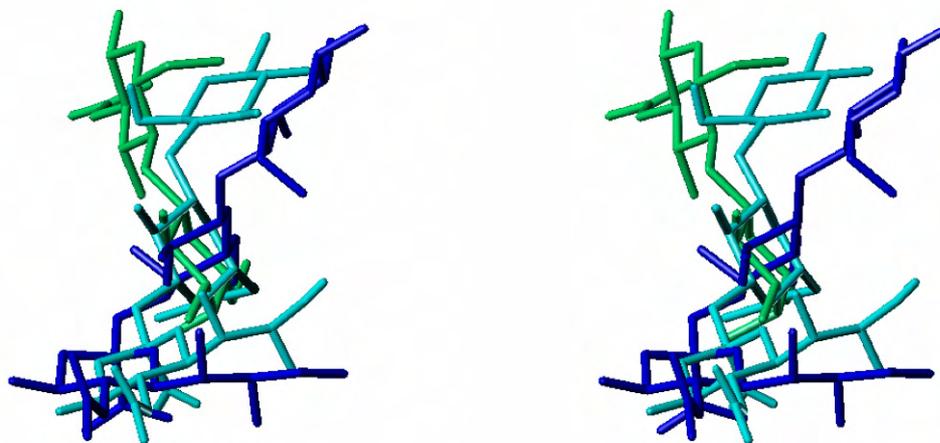


Figure 3.83: Comparison of the poses of LacNAc from X-ray structure 1slt (green), the average structure SiaLac after 2 ns MD simulation (cyan) and the docking result of SiaLac (blue).

Two clusters of results were generated from the MD ensemble of SiaLac in complex with galectin-1 as described for lacto-*N*-tetraose (p92). Poses of the ligand were extracted from the trajectories and stored in a multi molecule file in the same fashion as the docking results. These poses could be processed in the same way as the docking results from docking studies.

The first cluster of MD results was chosen following a mixed energy based and random selection protocol. The first structure in the cluster was chosen to be the energy minimized average structure of the 2 ns simulation time. The next poses were 15 structures of the MD with the lowest potential energy calculated by the force field during the MD simulation. The last four structures had the highest calculated potential energy respectively. The result cluster was supplemented by ten more structures, chosen at random from the remaining snapshots of the simulation, yielding an ensemble of 30 structures in total. Most of these poses were able to transfer saturation from the protein to the *N*-acetyl group of Neu5Ac'' including all poses with low potential energy. The second cluster consisted of eleven ligand structures extracted from the MD simulation in constant time frames. The first structure was the result after 500 fs simulation time as the starting structure was already included into the first result cluster. The second structure was taken after 200 ps simulation time and in 200 ps intervals from here on. In all of these poses transfer of magnetization from protein to the *N*-acetyl group of Neu5Ac'' is possible.

Comparison of the calculated STD effects and experimentally determined STD effects already revealed good agreement of prediction and experiment for most of the ligand protons. However, experimental STD NMR data indicated contacts of H9a and H9b along with contacts of the *N*-acetyl group of the Neu5Ac residue that could not be explained by docking experiments.

The averaged calculated STD effect determined for the second cluster of structures displayed intensity for the protons in 9-position and for the *N*-acetyl group of the Neu5Ac residue. The calculated STD effects of the *N*-acetyl group were a result of interactions with the side chains of Cys2 and Leu31 of galectin-1. The intensities calculated for H9a'' and H9b'' were a result of contacts with the side chain of Trp68. However, calculated intensities did not match those experimentally determined. This could be a result of insufficient conformational sampling based on the setup of the MD simulation. In the averaged structure the intensities calculated for H3a'' and H3e'' differ from the experimentally determined pattern. The average structure is in a low energy conformation (Table 3.15) but differs from the conformation of the docked carbohydrate (p32). H3a'' interacts with Ser29, Leu31 and His52, H3e'' additionally interacts with His44, Asn46, Arg48, Ala51 and Gly124.

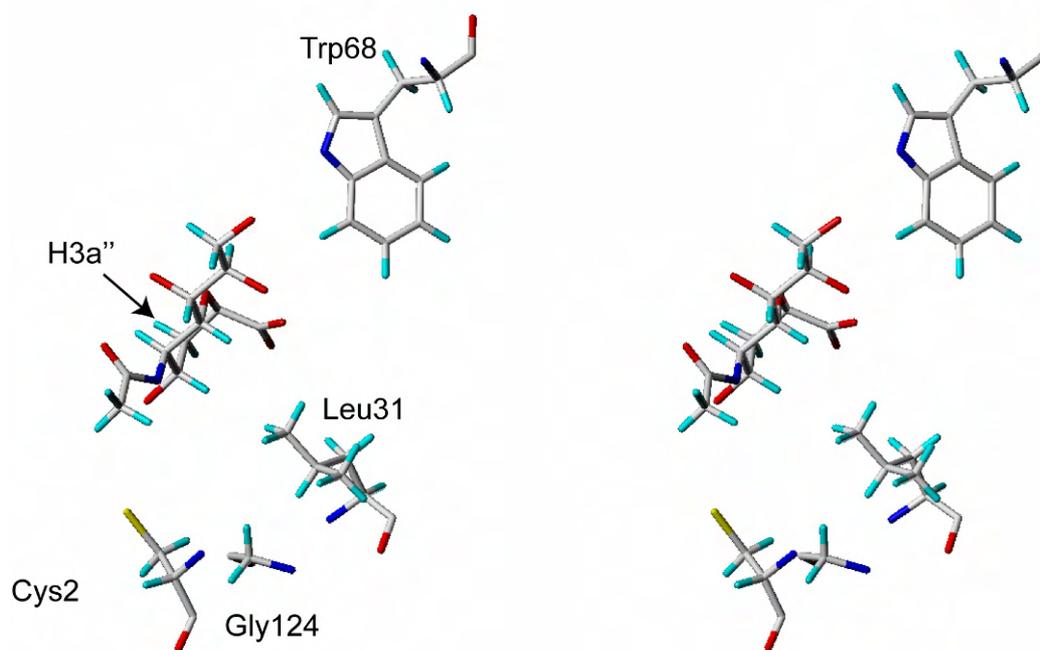


Figure 3.84: The interactions of the Neu5Ac'' residue glycerol side chain are happening with the H ζ of Trp68. The N-acetyl group interacts with Cys2, Leu31 and Gly124. Interactions of H3a'' and H3e'' are not highlighted for clarity. Hydrogen atoms are only shown if they can transfer saturation (protein) or be detected by NMR (Neu5Ac), Gal' and Glc residue omitted for clarity.

The strong calculated effect of H4'' is a result of a close contact with the methyl group in δ -position of Leu31. In the MD setup the receptor structure was kept rigid, thus the distances could be an artifact. The distance did not fall below 2 Å. In the averaged structure from the MD simulation the distance ranged from 2.6 Å to 3.2 Å causing the highest calculated effect.

Again, the setup of the MD simulation omitting explicit treatment of water molecules may have had an influence on the conformational sampling and did not allow investigation of possible saturation pathways through solvent molecules. Thus certain STD intensities determined experimentally could not be justified by structural ensembles proposed by modeling techniques. As in the case of LNT before, additional experimentally determined constraints on the conformation of SiaLac, bound to galectin-1 may give further helpful information for modeling approaches.

However, the analysis of conformational properties and resulting influences on the orientation of ligand receptor complexes by MD simulations did yield indications on the influence of the complexes flexibility on the STD NMR results.

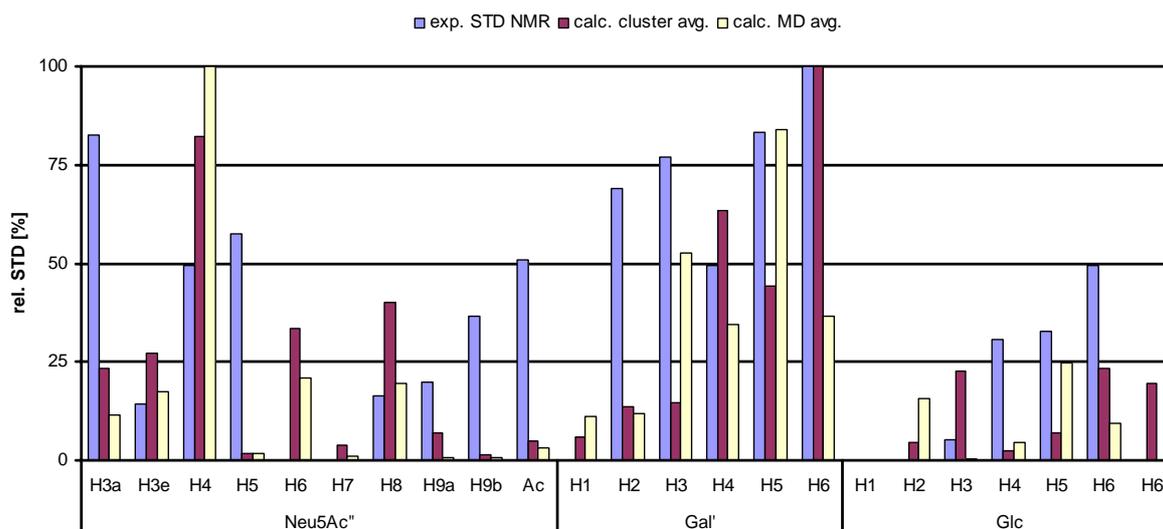


Figure 3.85: The comparison of the experimental STD effect (blue) with the average calculated STD effects for the cluster compiled from periodically chosen MD frames (red) and for the average structure of the MD simulation (yellow) displays broad agreement.

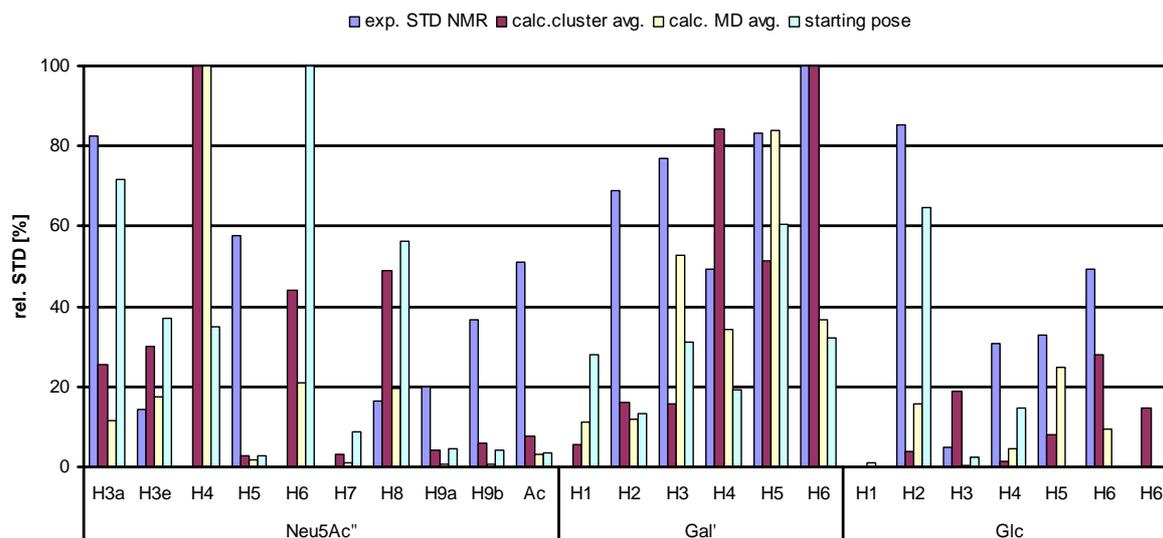


Figure 3.86: The comparison of the epitopes from STD NMR and calculated from structures selected from potential energy reveals some similarities. The epitope calculated for the averaged cluster of structures (red, cluster avg.) displays the common pattern of intensities on the Gal' residue but also show strong interaction on H4'' which was not found experimentally. The starting pose of the MD simulation is the structure with the lowest potential energy in the MD simulation.

4 Conclusions

The accuracy of binding mode predictions of ligands in contact with a specific receptor protein still poses an obstacle in the application of docking strategies in the drug development process²⁷. Improvements originating from optimized computer programs or analysis protocols take considerable time to become established and require thorough testing and considerable computing capacities.

In this work it was shown that results from established docking programs can be checked and improved if experimental data on the binding mode can be incorporated into the setup and analysis of the docking experiment. Information on the binding mode was taken from Saturation Transfer Difference (STD) NMR experiments. STD NMR and docking experiments were performed with galectin-1, a mammalian lectin specifically binding to β -D-galactoside containing carbohydrates, and various natural and unnatural mono-, di- and oligosaccharides. Docking experiments were performed using DOCK.

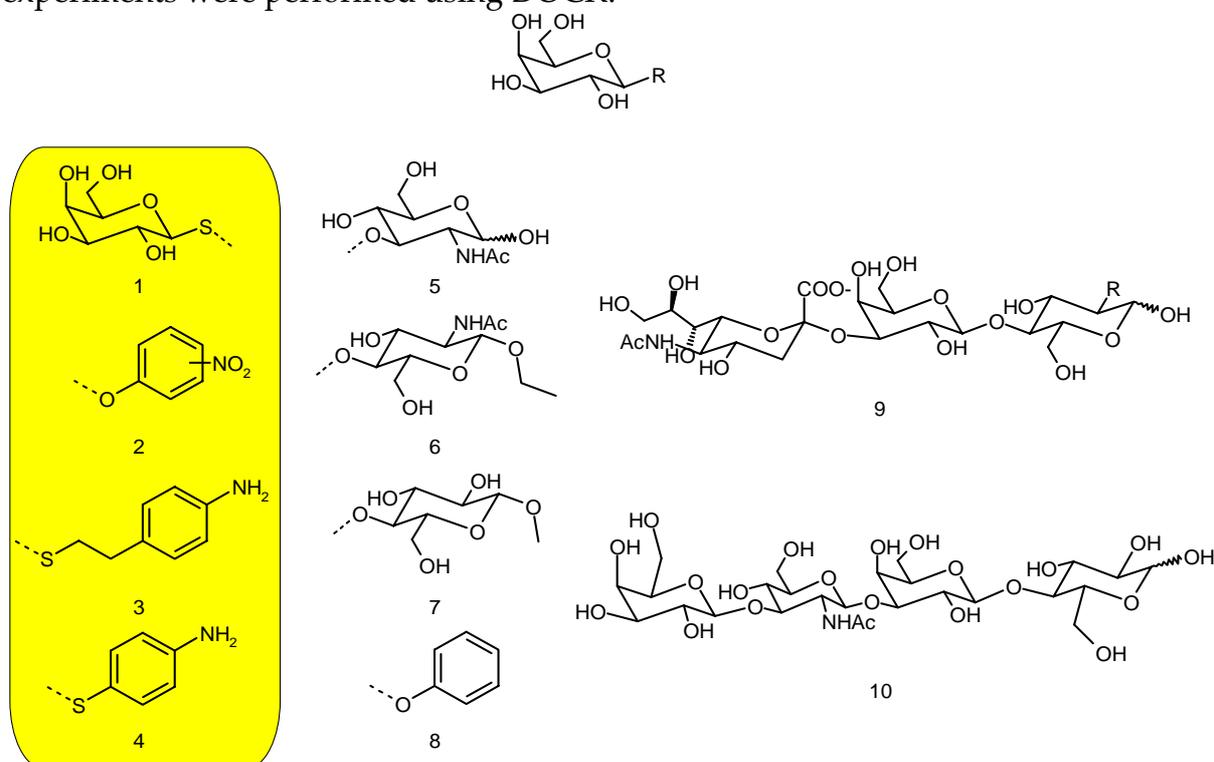


Figure 4.1: ligands examined in this study. Structures highlighted with yellow background were only studied by modeling procedures. 1: TDG, 2: *p*-nitro-PheGal and *o*-nitro-PheGal, 3: 2-(*p*-amino-phenyl)-ethyl thio-galactoside, 4: *p*-amino thio-galactoside, 5: LNB, 6: LacNAcOEt, 7: LacOMe, 8: PheGal, 9: SiaLac, 10: LNT (R: structures 1-8)

STD NMR experiments revealed that the strongest contacts occurred for protons of the galactose residues of various ligands, as was expected for a galactose specific receptor. The strongest interaction was found on the protons in 6-position of galactose in the disaccharides *N*-acetyl lactosamine (type II) and lacto-*N*-biose and the trisaccharide sialyllactose. Additionally the 4-position of the galactose residues showed medium to strong intensity and the 5-position was weaker than the 4-position.

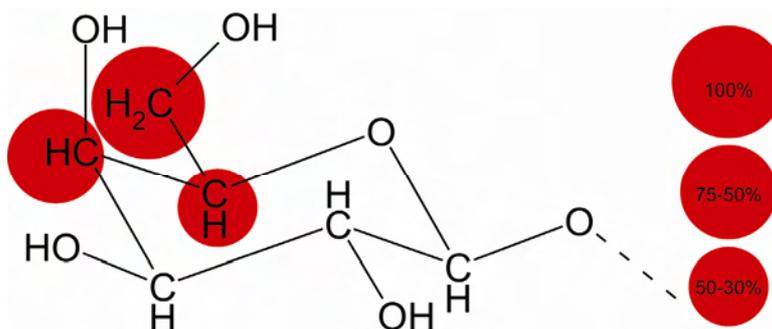


Figure 4.2: The typical pattern of intensities in STD NMR spectra revealed the strongest intensities on H6 of β -D-galactose residues followed by H4 and H5.

In the course of this work the kinetics of the binding event between carbohydrates and galectin-1 proved to be disadvantageous for STD NMR experiments. Due to slow kinetics only low turnover in the sample was achieved, which in turn resulted in low signal to noise ratio. This led to difficulties in accurate determination of STD intensities.

Based on the preference of galectin-1 for β -galactosides and the strong interactions of galactose residues with the protein found in STD NMR experiments a docking protocol was developed that was able to produce poses of the ligand in the expected binding mode. The docking procedure started with placing the galactosyl residue of a given ligand in the binding site in an optimized orientation. Remaining parts of the ligand were consecutively attached to the galactosyl residue in various orientations to allow ligand flexibility in the docking process. The final ligand structure is minimized to optimize orientation and conformation of the ligand. This docking protocol proved successful in the generation of binding modes for various ligands. With the improvements gained from the incremental construction approach the docking protocol was capable of handling ligands like tri- and tetrasaccharides that would have been difficult to handle otherwise, due to their size and 'undruglike' flexibility. However, docking performance got worse with increasing size and flexibility of the ligand docked.

Thus, some results achieved with the docking protocol were unsatisfactory. These unsatisfying results displayed high structural variability and disagreed with the binding mode expected from published X-ray data⁷⁴ and STD NMR data^{77,78}. This was shown in the case of the tetrasaccharide lacto-*N*-tetraose (LNT). The setup that was capable to dock various disaccharides to galectin-1 was not able to create consistent resulting structures for the binding mode of this ligand. The consistence of the docked structures was improved by creating a cluster of spheres matched to the non reducing end of the tetrasaccharide on the basis of a binding mode in accordance with STD NMR data of lacto-*N*-biose in complex with galectin-1. However, the resulting structures of LNT were not able to explain the results obtained from STD NMR experiments. The results for the trisaccharide sialyllactose were also structurally inconsistent but three poses obtained by the docking procedure were in agreement with the binding mode expected from X-ray data. These three poses were also largely in agreement with the data on the binding mode from STD NMR with the exception of the orientation of the *N*-acetyl group and the protons in 9-position of the neuraminic acid residue. The results from STD NMR experiments indicated a

contact of the *N*-acetyl group with galectin-1 but data from docking did not display any contact.

In order to assess the validity of a set of structurally diverse binding modes created by DOCK a methodology was developed to match the epitope determined from STD NMR experiments with the binding mode of the ligand in complex with the receptor. The intensity of a proton resonance in the STD spectrum was correlated with the distance of the respective proton to protons of the receptor protein. The effectiveness of saturation transfer from protein proton to ligand proton is proportional to r^{-6} and is not detectable at distances (r) exceeding 5 to 6 Å. By mapping the distances of each proton of the ligand to protons of the receptor within 6 Å and summing up the r^{-6} values a measure for the expected intensity in the STD NMR was achieved.

Comparison of STD effects determined for type II *N*-acetyl lactosamine binding to galectin-1 and calculated for the crystal structures of *N*-acetyl lactosamine in the binding site revealed good accordance. The methodology was applied to evaluate poses generated from docking experiments. To evaluate more than one pose in short time a computer program was developed capable of handling docking results containing several poses. The methodology allowed the selection of single poses proposed by the docking program that were in accordance with the STD NMR results and the assumed binding mode derived from the binding mode in the crystal structure.

The flexibility of the ligand-receptor complex and its influence on results from STD NMR were assessed by molecular dynamics simulations. The disaccharide *N*-acetyl lactosamine proved to be treated easily by a MD simulation with fully flexible ligand and receptor in a water box with periodic boundary conditions applied. However, larger and more flexible ligands required simplification of the approach to keep computational time within reasonable limits. Thus explicit treatment of water and full receptor flexibility were omitted for sake of computing time. The tetrasaccharide lacto-*N*-tetraose was difficult to handle. The simulation required constraints defined for the glycosidic torsion angles to keep them in low energy conformation. The tetrasaccharide did not remain in the binding site of galectin-1 in the course of the simulation. It became obvious that proper handling of a highly flexible ligand as lacto-*N*-tetraose required a more sophisticated approach in modeling. This was not within the scope of this thesis where fast and reliable prediction of binding modes was the main goal.

The use of simple MD approaches in the assessment of binding modes was demonstrated in the case of sialyllactose in complex with galectin-1. In this simple approach with rigid receptor and with no torsion angle constraints applied the ligand stayed in the binding site of the receptor in low energy conformation. The course of the simulation revealed binding modes of the ligand that allowed saturation of the *N*-acetyl group of the Neu5Ac'' residue. Additionally, contacts of the protons in 9-position of the Neu5Ac'' residue to the protein were found. The *N*-acetyl group and the protons in 9-position both displayed intensity in the STD NMR spectra but the binding modes from docking experiments ruled out transfer of saturation to these groups.

5 Zusammenfassung

Die Vorhersage des Bindungsmodus von Liganden, die in Wechselwirkung mit einem Rezeptor treten, stellt immer noch ein Problem bei der Verwendung von *docking* Strategien bei der Entwicklung neuer Medikamente dar. Fortschritte, welche aus der Optimierung der verwendeten Computerprogramme oder Analyseprotokolle herrühren, benötigen mitunter beträchtliche Rechenzeit.

In dieser Arbeit konnte gezeigt werden, dass die Ergebnisse eines etablierten *docking* Programms mit experimentellen Daten überprüft und verbessert werden können, wenn die experimentellen Daten in den Aufbau des *dockings* und Analyse des Ergebnisses miteinbezogen werden. Informationen über den Bindungsmodus wurden mit der *Saturation Transfer Difference* (STD) NMR Spektroskopie erhalten. STD NMR Experimente wurden mit Galectin-1, einem in Säugetieren vorkommenden Lektin, welches spezifisch β -Galaktoside bindet, und verschiedenen natürlichen und unnatürlichen Mono-, Di- und Oligosacchariden durchgeführt (Figure 4.1).

In den STD NMR Experimenten zeigten sich die stärksten Interaktionen mit Protonen der Galaktoseeinheit verschiedener Liganden, was für einen Galaktose spezifischen Rezeptor auch erwartet wurde. Die stärksten Interaktionen wurden für die Protonen in 6-Position der Galaktoseeinheit in den Disacchariden *N*-Acetyllactosamin (Typ II) und Lacto-*N*-biose gefunden. Zusätzlich wurde eine mäßig starke Interaktion für die 4-Position der Galaktose und schwächere Interaktionen für die 5-Position gefunden (Figure 4.2).

Im Verlauf dieser Arbeit stellte sich die Kinetik der Bindung von Kohlenhydraten and Galectin-1 als nachteilig für STD NMR Messungen heraus. Durch die langsame Kinetik konnte nur ein geringer Austausch von Liganden in der Bindungstasche erzielt werden, welcher in einem schlechten Signal zu Rausch Verhältnis resultierte. Dies führte zu Problemen in der exakten Bestimmung der Intensitäten im STD Spektrum.

Basierend auf der Spezifität des Galectin-1 für β -Galaktoside und den starken Interaktionen der Galaktoseeinheiten mit dem Protein, welche im STD NMR Experiment gefunden wurden, wurde ein *docking* Protokoll entwickelt, welches Strukturen der Liganden in der erwarteten Orientierung generierte. Am Beginn des Vorgangs wurde die Galaktoseeinheit eines gegeben Liganden in der Bindungstasche platziert und deren Orientierung optimiert. Verbleibende Teile des Liganden wurden nacheinander in verschiedener Orientierung angebaut, um die Flexibilität des Liganden zu simulieren. Die resultierende Struktur des Liganden wurde erneut minimiert, um die Lage und Konformation des Liganden zu optimieren. Diese Vorgehensweise wurde erfolgreich zur Generierung der Bindungsmodi verschiedener Liganden eingesetzt. Mit der Verbesserung der Leistung, welche aus dem inkrementellen Aufbau des Liganden resultierte, konnte dieses Verfahren auf Liganden wie Tri- und Tetrasaccharide angewandt werden. Liganden dieser Art sind aufgrund ihrer Größe und Flexibilität schwer zu handhaben. Trotz der Verbesserungen nahm die Qualität der Ergebnisse mit zunehmender Größe und Flexibilität des Liganden ab.

Daher waren einige Ergebnisse des *docking* Protokolls unbefriedigend. Diese zeigten eine hohe strukturelle Variabilität und stimmten nicht mit den Bindungsmodi aus Röntgenstrukturuntersuchungen und den STD NMR Experimenten dieser Arbeit überein. Dies war der Fall für das Tetrasaccharid Lacto-*N*-tetraose (LNT). Das *docking* Protokoll, welches verschiedene Disaccharide erfolgreich an Galectin-1 docken konnte, erwies sich nicht in der Lage, konsistente Ergebnisse für den Bindungsmodus dieses Liganden zu liefern. Das Ergebnis konnte verbessert werden, indem die Koordinaten des in Übereinstimmung mit Daten aus STD NMR Experimenten an Galectin-1 gebundenen Disaccharids Lacto-*N*-biose als Basis für den Bindungsmodus herangezogen wurden. Dieses Disaccharid entspricht dem nicht reduzierenden Ende von LNT. Dennoch waren die Ergebnisse des *docking* Vorgangs nicht in der Lage, die experimentellen STD NMR Daten zu erklären. Die *docking* Ergebnisse für das Trisaccharid Sialyllactose waren ebenfalls inkonsistent in ihrer Orientierung. Dennoch waren drei der Strukturen in Übereinstimmung mit dem aus der Röntgenstruktur erwarteten Bindungsmodus. Diese drei Strukturen waren auch größtenteils mit den Daten der STD NMR Experimente in Übereinstimmung. Die Ausnahme bildeten die *N*-Acetyl Gruppe und die 9-Position des Neuraminsäurerestes. In den STD NMR Experimenten zeigte sich ein Kontakt für die *N*-Acetyl Gruppe, welcher jedoch nicht mit den strukturellen Daten aus dem *docking* unterstützt werden konnte.

Um die Plausibilität eines Bindungsmodus aus einem strukturell diversen *docking* Ergebnis zu prüfen wurde eine Methode entwickelt, um das Bindungszeptop aus den STD NMR Experimenten mit dem Bindungsmodus aus *docking* Prozeduren zu vergleichen. Die Intensität eines Signals für ein Proton im STD Spektrum wurde mit der Distanz dieses Protons zu Protonen im Rezeptor korreliert. Die Effizienz des Sättigungstransfers ist proportional zu r^{-6} und ist bei Entfernungen (r) über 5 bis 6 Å sehr gering. Durch Vermessung aller Distanzen von Protonen des Liganden zu Protonen des Rezeptors innerhalb 6 Å Abstand und Summierung von r^{-6} konnte ein Maß für die Intensität im STD NMR Spektrum berechnet werden.

Der Vergleich von STD NMR Intensitäten von *N*-Acetyllactosamin, welches an Galectin-1 bindet und aus der Kristallstruktur berechneten Intensitäten ergab eine gute Übereinstimmung. Diese Methode wurde zur Bewertung von *docking* Ergebnissen angewandt. Um mehrere Strukturen in kurzer Zeit verarbeiten zu können, wurde ein Computerprogramm entwickelt, welches *docking* Ergebnisse mit mehreren Strukturen verarbeiten kann. Diese Methode erlaubte die Selektion bestimmter Bindungsmodi aus der *docking* Prozedur, welche mit den STD NMR Daten übereinstimmten und dem Bindungsmodus aus der Kristallstruktur entsprachen.

Die Flexibilität des Komplexes aus Ligand und Rezeptor und deren Einfluss auf die STD NMR Ergebnisse wurde mittels Molekül Dynamik (MD) Simulationen überprüft. Das Disaccharid *N*-Acetyllactosamin konnte in einer MD Simulation mit vollständig flexiblem Liganden und Rezeptor in einer Wasserbox simuliert werden. Größere, flexiblere Moleküle machten eine Vereinfachung des Systems nötig um Rechenzeit zu sparen. Daher wurde auf die explizite Berücksichtigung von Wasser und vollständige Flexibilität des Rezeptors verzichtet. Die Simulation des Tetrasaccharids Lacto-*N*-tetraose erwies sich als problematisch. Die MD Simulation

erforderte die Definition von Beschränkungen (*constraints*) für die Torsionswinkel der glycosidischen Bindungen, um eine Konformation mit niedriger Energie zu gewährleisten. Im Verlaufe der MD Simulation blieb das Tetrasaccharid nicht in der Bindungstasche. Es stellte sich heraus, dass die Betrachtung dieses komplexen Systems einen anspruchsvolleren Ansatz zur Modellierung benötigt hätte. Dies war im Rahmen dieser Arbeit, die die schnelle und zuverlässige Vorhersage des Bindungsmodus zum Ziel hatte, nicht angemessen und wurde daher nicht durchgeführt.

Der Nutzen einfacher MD Simulationen konnte anhand der Beurteilung des Bindungsmodus von Sialyllactose im Komplex mit Galectin-1 gezeigt werden. In einem einfachen Ansatz mit starrem Rezeptor und ohne Beschränkung der Flexibilität der glycosidischen Bindung verblieb der Ligand in der Bindungstasche in einer günstigen Konformation. Im Verlaufe der Simulation traten Strukturen, auf die den Transfer von Sättigung auf die *N*-Acetyl Gruppe und die Protonen in 9-Position des Neuraminsäurerestes erlaubten. Diese Gruppen zeigten Intensität im STD NMR Spektrum, die Bindungsmodi aus dem *docking* schlossen jedoch einen Sättigungstransfer aus.

6 Experimental section

6.1 Chemicals

Table 6.1: Sources of the substances used.

Phenyl- β -D-galactoside, Methyl- β -D-galactoside, Methyl- β -lactoside, Galaptin (<i>bos Taurus</i> , Gal1), <i>N</i> -acetyl-lactosamine	Sigma
2-Phenylethyl- β -D-thiogalactoside	Fluka
Lacto- <i>N</i> -biose, Lacto- <i>N</i> -tetraose,	Calbiochem, Merck Biosciences
D ₂ O	Deutero
TRIS-d11, DTT-d10	Eurisotop
Di-sodium-hydrogen-phosphate, Sodium-di-hydrogen-phosphate, Sodium Azide	Merck
Sodium chloride	Baker

6.2 Toxicology

Table 6.2: Risk phrases and safety precaution for harmful chemicals used in this work.

	Hazard symbol	risk phrases	safety precautions
sodium azide	T+, N	28-32-50/53	28.1-45-60-61
hydro chloric acid	C	34-37	26-36/37/39-45
sodium hydroxide	C	35	26-37/39-45
tris(hydroxymethyl)methylamine-d11	Xi	33/37/38	26-36
<i>N</i> -acetyl-D-lactosamine	Xi	36/37/38	26-23/37
DL-dithiothreitol-d10	Xi	22-36/37/38	26-36

6.3 Buffers

Buffer solutions were prepared at room temperature. The solvents and prepared solutions were degassed by exposure to ultrasonic waves for five to ten minutes. Buffers and stock solutions were degassed by this procedure before usage in NMR experiments.

6.3.1 TRIS buffer

The buffer solution contained 10 mM TRIS-D11, 150 mM NaCl, 4 mM NaN₃ and 2 mM DTT-D₁₀ (8 fold excess referring to 135 nmol galectin-1, containing 6 free cysteins in 3 mL of buffer) in D₂O. The pH was adjusted to 7.0 to 7.5 using DCl and NaOD.

6.3.2 PBS buffer

The buffer solution contained 4 mM NaN₃, 145 mM NaCl, 8 mM Na₂HPO₄ and 2 mM NaH₂PO₄ in D₂O. Usually the buffer was in the desired pH range of 7.0 to 7.5.

6.3.3 Preparation of galectin-1

Galectin-1 was purchased from Sigma as a solid in a TRIS/DTT salt mix (content 10%). The solid was dissolved in 4 mL of a TRIS/PBS mixture (1:1) in D₂O. 2 mL of the slightly hazy solution were transferred to a pre-rinsed Amicon spintube with 5000 g/mol mol weight cut off and spinned down for 40 min at 4000 rpm in an Eppendorf 2k15 centrifuge. This led to approximately 500 μ L of residue in on the membrane. The remaining 2 mL of solution were added and spinned down likewise. The original phial was washed out two times with TRIS-D₁₀ buffer which was added to the residue in the spin tube. Finally the residue in the spin tube was five times rinsed with 1 mL of deuterio TRIS buffer solution (35 min spin time at 4000 rpm for 1 mL). The residue was spinned down to 500 μ L and transferred to a plastic phial. The membrane was washed with 1 mL of TRIS-D₁₀ buffer two times and the combined liquids were stocked up to 3 mL volume with TRIS-D₁₀ buffer. This yielded a 45 μ M stock solution of galectin-1. The galectin-1 stock solution was degassed and kept at temperatures < 10° C in a refrigerator.

6.3.4 Preparation of ligand stock solutions

Carbohydrate ligands were dissolved in defined volumes of PBS-D₂O buffer to produce stock solutions of 5 mM or 10 mM concentration.

6.3.5 NMR samples

NMR experiments were carried out in Wilmad 5 mm NMR tubes, type 542-PP at 300 K.

NMR samples contained 225 μ L of galectin-1 stock solution (10 nmol of protein). The volume of carbohydrate stock solution was determined by the desired excess of carbohydrate in the sample. The sample was stocked up to a total volume of 540 μ L. Titrations were carried out by mixing the start sample and adding carbohydrate stock solution to the sample according to the excess desired. All samples were gently shaken after composition.

6.4 NMR recording and processing

NMR spectra were recorded on an Bruker DRX500 with inverse triple resonance 5 mm probe head and a Bruker AMX700 with inverse triple resonance 5 mm probe head with cryogenic cooling. Water suppression was achieved by applying the WATERGATE sequence. Protein resonances were suppressed using a T_{1 ρ} filter with a spin lock pulse length of 30 ms. STD spectra were recorded in a pseudo 2D style, where on and *off resonance* FIDs were stored in a serial file using an alternating phase cycle that in turns writes on and *off resonance* FIDs. 2D spectra were recorded phase sensitive in the F1 domain using TPPI.

NMR spectra were processed on SGI Octane computers (R14000 and R12000 CPU's) with XWINNMR 3.1. 2D spectra were processed with double zero filling in F1 and F2 direction. If applicable FIDs were Fourier transformed using an exponential window function, line broadening was set according to the digital resolution. Phase correction was achieved using automated phase correction. STD spectra were processed by splitting the serial file into separate FIDs for on and *off resonance*. 1D FIDs were processed as described above. In order to minimize subtraction artifacts line broadening was set to 0.5 to 2 Hz. The primary phase (phc0) of the *off resonance* spectrum was inverted by 180° and the FIDs were added up. Signals were defined and integrated in the *off resonance* spectrum, the same definitions were applied to the difference spectrum and the integrals scaled to the *off resonance* spectrum.

2D homonuclear spectra often required processing with a sine or q-sine window function in F1 direction. Plots of NMR spectra were generated either with Bruker XWinplot 3.1 or Bruker Topspin 1.3.

6.5 NMR assignments of carbohydrates

6.5.1 Assignment for LNT

Table 6.3: ^1H and ^{13}C shift assignments for LNT. All chemical shift assignments are referenced to the resonance of HDO at 4.71ppm at 300 K. Chemical shifts were assigned using HSQC, HMBC, TOCSY and COSY spectra, acquired on a 700 MHz spectrometer all values given in ppm. The structure of LNT is given in Figure 3.56.

	Glc(α/β)		Gal'		GlcNAc''		Gal'''	
	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C
H1	4.591 5.149	95.64 91.73	4.371	102.86	4.6609	102.51	4.376	103.36
H2	3.208 3.570	73.75 73.77	3.519	69.96	3.829	54.70	3.451	70.68
H3	3.569 3.874	74.29 70.06	3.668	81.93	3.748	82.08	3.754	71.41
H4	3.578 3.578	78.29	4.082	68.29	3.498	68.41	3.843	68.52
H5	3.530 3.564	74.73 74.34	3.641	75.09	3.413	75.14	3.573	72.42
H6a					3.819	60.31		
H6b					3.721			
NHAc	--	--	--	--	1.957	22.07	--	--

6.5.2 Assignment for 3'-sialyllactose

Table 6.4: ^1H and ^{13}C shift assignments for Sialyllactose. All chemical shifts are referenced to the signal of HDO at 4.715ppm at 300 K. The resonance of H1 from αGlc was the only resonance of the α -anomer visible in the spectrum and was discarded. Chemical shifts were determined using HSQC and COSY spectra on a 700 MHz spectrometer, all values given in ppm. The structure of sialyllactose is given in Figure 3.43.(nd: not determined)

	Glc		Gal'		Neu5Ac''	
	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C
H1	4.624	96.019	4.491	102.867	--	nd
H2	3.245	74.040	3.533	69.708	--	nd
H3a	3.603	73.198	4.072	75.856	1.759	nd
H3e					2.717	nd
H4	3.642	78.554	3.926	67.843	3.650	68.865
H5	3.561	75.023	3.673	75.525	3.804	52.024
H6a	3.921	60.244	3.708	61.385	3.606	74.702
H6b	3.850					
H7	--	--	--	--	3.559	68.444
H8	--	--	--	--	3.851	72.120
H9a	--	--	--	--	3.824	62.820
H9b	--	--	--	--	3.602	
Ac	--	--	--	--	1.997	22.485

6.5.3 Assignment for ethyl- β -D-lactosamine (LacNAcOEt)

Table 6.5: ^1H and ^{13}C shift assignments for LacNAcOEt. All chemical shifts were referenced to the signal of HDO at 4.72ppm at 300 K. Chemical shifts were determined using HSQC and COSY spectra on a 500 MHz spectrometer, all values given in ppm. In the ethyl group the 1-position is the methylene group and the 2-position is the methyl group.

	GlcNAc		Gal	
	^1H	^{13}C	^1H	^{13}C
Ethyl H1a	3.867	66.746	--	--
Ethyl H1b	3.645		--	--
Ethyl H2	1.133	14.702	--	--
H1	4.529	101.02	4.44	103.28
H2	3.681	55.48	3.506	71.381
H3	3.567	75.244	3.637	72.938
H4	3.678	78.797	3.895	69.064
H5	3.673	72.982	3.698	75.719
H6a	3.952	60.531	3.739	61.541
H6b	3.795	60.531	3.708	
Ac	2.005	22.61	--	--

6.5.4 Assignment for lacto-*N*-biose (LNB)

Table 6.6: ^1H and ^{13}C shift assignments for LNB. All ^1H chemical were referenced to the signal HDO at 4.700ppm at 300 K on a 700 MHz. Chemical shift were determined using HSQC and COSY spectra, all values are given in ppm.

	$\alpha\text{GlcNAc}[\text{ppm}]$		$\beta\text{GlcNAc}[\text{ppm}]$		Gal[ppm]	
	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C
H1	4.931	91.155	4.509	94.970	4.227	103.712
H2	3.825	53.241	3.602	nd	3.279	70.875
H3	3.702	80.294	nd	nd	3.412	72.687
H4	3.345	68.858	3.7	nd	3.674	68.744
H5	3.666	71.446	3.350	nd	3.481	75.361
H6	nd	nd	60.648	3.595	3.522	61.264
Ac	nd	nd	22.378	1.798	--	--

6.5.5 Assignment for methyl- β -D-lactoside (LacOMe)

Table 6.7: ^1H and ^{13}C shift assignment for LacOMe. All ^1H chemical shifts were referenced to the signal of HDO at 4.700ppm at 300 K. Chemical shifts were determined using a HSQC spectrum, all values are given in ppm.

	GlcNAc		Gal	
	^1H	^{13}C	^1H	^{13}C
H1	4.163	103.040	4.131	103.201
H2	3.253	71.025	3.017	72.909
H3	3.383	72.552	3.364	74.544
H4	3.370	78.229	3.634	68.636
H5	3.322	74.925	3.438	75.523
H6a	3.710	60.022	3.495	61.217
H6b	3.530	60.037	3.458	61.217
OMe	3.291	57.355	--	--

6.5.6 Assignment for phenyl- β -D- galactoside (PheGal)

Table 6.8: ^1H and ^{13}C shift assignment for LacOMe. All ^1H chemical shifts were referenced to the signal of HDO at 4.700 ppm at 285 K on a 500 MHz. Chemical shifts were determined using HSQC and COSY spectra, all values are given in ppm.

	Gal	
	^1H	^{13}C
H1	4.848	100.786
H2	3.575	70.742
H3	3.549	72.706
H4	3.772	68.631
H5	3.653	75.632
H6a	3.548	60.963
Phe-o	7.156	130.052
Phe-m	6.906	123.301
Phe-p	6.913	116.494

6.6 Molecular modeling and computational chemistry

Molecular modeling was performed using Sybyl 7.2 (Tripos)⁸⁹ on SGI Octane computers with R14000 and R12000 CPUs. Analysis of ligand receptor proton distances was performed on a HP zx6000 workstation with two Itanium 2 processors.

6.6.1 Preparation of X-ray structures

Procedures for structure preparation prior to docking were taken from the manual of Dock v4.0. PDB files were stripped of crystallographic water molecules, metal ions and counter ions. Ligand molecules were extracted and processed separately. If obvious violations of angles, bond lengths and atom distances were detected by the protein preparation tool from Sybyl corrections were applied. Connolly surfaces were computed with the ms program from Quantum Chemical Program Exchange (<http://www.qcpe.indiana.edu/>, QCPE No. 429) with a probe radius of 1.4 Å and surface density of 5.0. Residues more than 8 Å away from the ligand in the crystal structure were discarded for surface calculation. Spheres for the determination of the binding site were created using SPHGEN. Sphere clusters were displayed with the ligand from X-ray structure to select valid clusters and discard spheres out of range. Alternatively coordinates of the ligand were taken as coordinates for spheres in a cluster. Afterwards hydrogen atoms were added to the protein structure and Kollman^{23,24} charges were applied for all atoms. The resulting structure was allowed relaxation of bonds with a short minimization of 500 steps. The grid was computed using the GRID program. Grids were calculated for energy, contact and chemical scoring. SPHGEN and GRID versions were used as distributed with DOCK v4.0.1.

Ligands from crystal structures were prepared by adding hydrogen after extraction from the protein and assignment of Gasteiger charges^{116,117}. The structure was allowed to relax bond angles and torsion during a minimization of 200 steps.

6.6.2 General setup for docking experiments

Analogous to the setup of Grid, distance dielectric with a dielectric constant of 4.5 was chosen. The exponents for the attractive and repulsive terms in the Lennar-Jones-potential were chose as before, 6 and 12 respectively. No additional scaling was applied to van-der-Waals or electrostatic interactions. The energy cut off distance was set to 10 Å or 12 Å.

7 References

1. Böhm H.-J., Klebe G., Kubinyi H., Wirkstoffdesign, Spektrum Akademischer Verlag, Heidelberg, 1996.
2. van de Waterbeemd H., Gifford E., ADMET in silico modelling: towards prediction paradise?, *Nat Rev Drug Discov*, 2003, 2, (3), 192-204.
3. Wolcke J., Ullmann D., Miniaturized HTS technologies - uHTS, *Drug Discov Today*, 2001, 6, (12), 637-46.
4. Bleicher K. H., Bohm H. J., Muller K., Alanine A. I., Hit and lead generation: beyond high-throughput screening, *Nat Rev Drug Discov*, 2003, 2, (5), 369-78.
5. Preziosi P., Science, pharmacoeconomics and ethics in drug R&D: a sustainable future scenario?, *Nat Rev Drug Discov*, 2004, 3, (6), 521-6.
6. Vogtherr M., Fiebig K., NMR-based screening methods for lead discovery, *EXS*, 2003, (93), 183-202.
7. Reichert J. M., Trends in development and approval times for new therapeutics in the United States, *Nat Rev Drug Discov*, 2003, 2, (9), 695-702.
8. Fitzgerald G. A., Opinion: anticipating change in drug development: the emerging era of translational medicine and therapeutics, *Nat Rev Drug Discov*, 2005, 4, (10), 815-8.
9. Sottriffer C., Klebe G., Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design, *Farmaco*, 2002, 57, (3), 243-51.
10. Schmitt S., Kuhn D., Klebe G., A new method to detect related function among proteins independent of sequence and fold homology, *J Mol Biol*, 2002, 323, (2), 387-406.
11. Wieman H., Tondel K., Anderssen E., Drablos F., Homology-based modelling of targets for rational drug design, *Mini Rev Med Chem*, 2004, 4, (7), 793-804.
12. Leahy D. E., Progress in simulation modelling for pharmacokinetics, *Curr Top Med Chem*, 2003, 3, (11), 1257-68.
13. Levinthal C., Wodak S. J., Kahn P., Dadivanian A. K., Hemoglobin interaction in sickle cell fibers. I: Theoretical approaches to the molecular contacts, *Proc Natl Acad Sci U S A*, 1975, 72, (4), 1330-4.
14. Salemme F. R., An hypothetical structure for an intermolecular electron transfer complex of cytochromes c and b5, *J Mol Biol*, 1976, 102, (3), 563-8.
15. Goodsell D. S., Olson A. J., Automated docking of substrates to proteins by simulated annealing, *Proteins*, 1990, 8, (3), 195-202.
16. Goodford P. J., A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, *J Med Chem*, 1985, 28, (7), 849-57.
17. Meng E. C., Shoichet B. K., Kuntz I. D., Automated docking with grid-based energy evaluation., *J Comp Chem*, 1992, 13, 505-24.
18. Kuntz I. D., Blaney J. M., Oatley S. J., Langridge R., Ferrin T. E., A geometric approach to macromolecule-ligand interactions, *J Mol Biol*, 1982, 161, (2), 269-88.
19. Brooijmans N., Kuntz I. D., Molecular recognition and docking algorithms, *Annu Rev Biophys Biomol Struct*, 2003, 32, 335-73.
20. Desjarlais R. L., Sheridan R. P., Dixon J. S., Kuntz I. D., Venkataraghavan R., Docking flexible ligands to macromolecular receptors by molecular shape, *J Med Chem*, 1986, 29, (11), 2149-53.
21. Ewing T. J., Makino S., Skillman A. G., Kuntz I. D., DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, *J Comput Aided Mol Des*, 2001, 15, (5), 411-28.
22. Rarey M., Kramer B., Lengauer T., Klebe G., A fast flexible docking method using an incremental construction algorithm, *J Mol Biol*, 1996, 261, (3), 470-89.
23. Weiner S. J., Kollman P. A., Case D. A., Singh U. C., Ghio C., Alagona G., Profeta S., Weiner P., A new force field for molecular mechanical simulation of nucleic acids and proteins., *J Am Chem Soc*, 1984, 106, 765-84.
24. Weiner S. J., Kollman P. A., Nguyen D. T., Case D. A., An all atom force field for simulations of proteins and nucleic acids., *J Comp Chem*, 1986, 7, 230-52.

25. Jorgensen W. L. , Tirado-Rives J. J., The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *J Am Chem Soc*, **1988**, 110, (6), 1657-66.
26. Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S., Karplus M. J., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations , *J Comput Chem*, **1983**, 4, (2), 187-217.
27. Erickson J. A., Jalaie M., Robertson D. H., Lewis R. A., Vieth M., Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy, *J Med Chem*, **2004**, 47, (1), 45-55.
28. Taylor R. D., Jewsbury P. J., Essex J. W., A review of protein-small molecule docking methods, *J Comput Aided Mol Des*, **2002**, 16, (3), 151-66.
29. Kollman P. A., Free energy calculations: applications to chemical and biochemical phenomena., *Chem Rev*, **1993**, 93, 2395-417.
30. Kramer B., Rarey M., Lengauer T. , Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking, *Proteins*, **1999**, 37, (2), 228-41.
31. Pang Y. P., Perola E., Xu K., Prendergast F. G., EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases, *J Comput Chem*, **2001**, 22, (15), 1750-71.
32. Jones G., Willett P., Glen R. C., Leach A. R., Taylor R., Development and validation of a genetic algorithm for flexible docking, *J Mol Biol*, **1997**, 267, (3), 727-48.
33. Kick E. K., Roe D. C., Skillman A. G., Liu G., Ewing T. J., Sun Y., Kuntz I. D., Ellman J. A., Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D, *Chem Biol*, **1997**, 4, (4), 297-307.
34. Ha S., Andreani R., Robbins A., Muegge I., Evaluation of docking/scoring approaches: a comparative study based on MMP3 inhibitors, *J Comput Aided Mol Des*, **2000**, 14, (5), 435-48.
35. Lamb M. L., Burdick K. W., Toba S., Young M. M., Skillman A. G., Zou X., Arnold J. R., Kuntz I. D., Design, docking, and evaluation of multiple libraries against multiple targets, *Proteins*, **2001**, 42, (3), 296-318.
36. Kroemer R. T., Molecular modelling probes: docking and scoring, *Biochem Soc Trans*, **2003**, 31, (Pt 5), 980-4.
37. Pellecchia M., Sem D. S., Wuthrich K., NMR in drug discovery, *Nat Rev Drug Discov*, **2002**, 1, (3), 211-9.
38. Meyer B., Peters T., NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors, *Angew Chem Int Ed Engl*, **2003**, 42, (8), 864-90.
39. Fejzo J., Lepre C., Xie X., Application of NMR screening in drug discovery, *Curr Top Med Chem*, **2003**, 3, (1), 81-97.
40. Hajduk P. J. , Augeri D. J. , Mack J. , Mendoza R. , Yang J. , Betz S. F. , Fesik S. W., NMR-based Screening of Proteins Containing ¹³C-Labeled Methyl Groups , *J Am Chem Soc*, **2000**, 122 , 7898-904.
41. Meyer B., Weimar T., Peters T., Screening mixtures for biological activity by NMR, *Eur J Biochem*, **1997**, 246, (3), 705-9.
42. Hajduk P. J., Olejniczak E. T., Fesik S. W., One-Dimensional Relaxation- and Diffusion Edited NMR Methods for Screening Compounds that Bind to Macromolecules, *J Am Chem Soc*, **1997**, 119, 12257-61.
43. Chen A., Shapiro M. J., NOE Pumping: A Novel NMR Technique for Identification of Compound with Binding Affinity to Macromolecules, *J Am Chem Soc*, **1998**, 120, 10258-9.
44. Chen A., Shapiro M. J., NOE Pumping. 2. A High-Throughput Method to Determine Compounds with Binding Affinity to Macromolecules, *J Am Chem Soc*, **2000**, 122, 414-5.
45. Carr H. Y., Purcell E. M., Effects of diffusion on free precession in nuclear magnetic resonance experiments, *Phys Rev*, **1954**, 94, 630-8.
46. Jahnke W., Perez L. B., Paris C. G., Strauss A., Fendrich G., Nalin C. M., Secon Site NMR Screening with a Spin-Labeled First Ligand , *J Am Chem Soc*, **2000**, 122, 7394-5.
47. Otting G., NMR studies of water bound to biological molecules., *Prog NMR spectrosc*, **1997**, 31, 259-85.
48. Dalvit C., Fogliatto G., Stewart A., Veronesi M., Stockman B., WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability, *J Biomol NMR*, **2001**, 21, (4), 349-59.
49. Mayer M. , Meyer B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectra, *Angew Chem Int Ed*, **1999**, 38, 1784-8.
50. Klein J. , Meinecke R. , Mayer M. , Meyer B., Detecting Binding Affinity to Immobilized Receptor Proteins in Compound Libraries by HR-MAS STD NMR, *J Am Chem Soc*, **1999**, 121, 5336-7.

51. Mayer M., Meyer B., Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, *J Am Chem Soc*, **2001**, 123, (25), 6108-17.
52. Meinecke R., Meyer B., Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin alphaIIb beta3, *J Med Chem*, **2001**, 44, (19), 3059-65.
53. Claasen B., Axmann M., Meinecke R., Meyer B., Direct observation of ligand binding to membrane proteins in living cells by a saturation transfer double difference (STDD) NMR spectroscopy method shows a significantly higher affinity of integrin alpha(IIb)beta3 in native platelets than in liposomes, *J Am Chem Soc*, **2005**, 127, (3), 916-9.
54. Hajduk P. J., Mack J. C., Olejniczak E. T., Park C., Dandliker P. J., Beutel B. A., SOS-NMR: a saturation transfer NMR-based method for determining the structures of protein-ligand complexes, *J Am Chem Soc*, **2004**, 126, (8), 2390-8.
55. Jayalakshmi V., Krishna N. R., Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes, *J Magn Reson*, **2002**, 155, (1), 106-18.
56. Wen X., Yuan Y., Kuntz D. A., Rose D. R., Pinto B. M., A combined STD-NMR/molecular modeling protocol for predicting the binding modes of the glycosidase inhibitors kifunensine and salacinol to Golgi alpha-mannosidase II, *Biochemistry*, **2005**, 44, (18), 6729-37.
57. Jayalakshmi V., Biet T., Peters T., Krishna N. R., Refinement of the conformation of UDP-galactose bound to galactosyltransferase using the STD NMR intensity-restrained CORCEMA optimization, *J Am Chem Soc*, **2004**, 126, (28), 8610-1.
58. Jayalakshmi V., Rama Krishna N., CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities, *J Magn Reson*, **2004**, 168, (1), 36-45.
59. Cooper D. N., Barondes S. H., God must love galectins; he made so many of them, *Glycobiology*, **1999**, 9, (10), 979-84.
60. Hirabayashi J., Hashidate T., Arata Y., Nishi N., Nakamura T., Hirashima M., Urashima T., Oka T., Futai M., Muller W. E., Yagi F., Kasai K., Oligosaccharide specificity of galectins: a search by frontal affinity chromatography, *Biochim Biophys Acta*, **2002**, 1572, (2-3), 232-54.
61. Leffler H., Carlsson S., Hedlund M., Qian Y., Poirier F., Introduction to galectins, *Glycoconj J*, **2004**, 19, (7-9), 433-40.
62. Barondes S. H., Cooper D. N., Gitt M. A., Leffler H., Galectins. Structure and function of a large family of animal lectins, *J Biol Chem*, **1994**, 269, (33), 20807-10.
63. Rabinovich G. A., Galectin-1 as a potential cancer target, *Br J Cancer*, **2005**, 92, (7), 1188-92.
64. Elola M. T., Chiesa M. E., Alberti A. F., Mordoh J., Fink N. E., Galectin-1 receptors in different cell types, *J Biomed Sci*, **2005**, 12, (1), 13-29.
65. Lanteri M., Giordanengo V., Hiraoka N., Fuzibet J. G., Auberger P., Fukuda M., Baum L. G., Lefebvre J. C., Altered T cell surface glycosylation in HIV-1 infection results in increased susceptibility to galectin-1-induced cell death, *Glycobiology*, **2003**, 13, (12), 909-18.
66. Miura T., Takahashi M., Horie H., Kurushima H., Tsuchimoto D., Sakumi K., Nakabeppu Y., Galectin-1beta, a natural monomeric form of galectin-1 lacking its six amino-terminal residues promotes axonal regeneration but not cell death, *Cell Death Differ*, **2004**, 11, (10), 1076-83.
67. Vyakarnam A., Dagher S. F., Wang J. L., Patterson R. J., Evidence for a role for galectin-1 in pre-mRNA splicing, *Mol Cell Biol*, **1997**, 17, (8), 4730-7.
68. Dagher S. F., Wang J. L., Patterson R. J., Identification of galectin-3 as a factor in pre-mRNA splicing, *Proc Natl Acad Sci U S A*, **1995**, 92, (4), 1213-7.
69. Nickel W., The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes, *Eur J Biochem*, **2003**, 270, (10), 2109-19.
70. Rabinovich G. A., Baum L. G., Tinari N., Paganelli R., Natoli C., Liu F. T., Iacobelli S., Galectins and their ligands: amplifiers, silencers or tuners of the inflammatory response?, *Trends Immunol*, **2002**, 23, (6), 313-20.
71. Sacchettini J. C., Baum L. G., Brewer C. F., Multivalent protein-carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction, *Biochemistry*, **2001**, 40, (10), 3009-15.
72. Ozeki Y., Matsui T., Yamamoto Y., Funahashi M., Hamako J., Titani K., Tissue fibronectin is an endogenous ligand for galectin-1, *Glycobiology*, **1995**, 5, (2), 255-61.
73. Yang R. Y., Liu F. T., Galectins in cell growth and apoptosis, *Cell Mol Life Sci*, **2003**, 60, (2), 267-76.

74. Lopez-Lucendo M. F., Solis D., Andre S., Hirabayashi J., Kasai K., Kaltner H., Gabius H. J., Romero A., Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding, *J Mol Biol*, **2004**, 343, (4), 957-70.
75. Leonidas D. D., Vatzaki E. H., Vorum H., Celis J. E., Madsen P., Acharya K. R., Structural basis for the recognition of carbohydrates by human galectin-7, *Biochemistry*, **1998**, 37, (40), 13930-40.
76. Seetharaman J., Kanigsberg A., Slaaby R., Leffler H., Barondes S. H., Rini J. M., X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution, *J Biol Chem*, **1998**, 273, (21), 13047-52.
77. Bourne Y., Bolgiano B., Liao D. I., Strecker G., Cantau P., Herzberg O., Feizi T., Cambillau C., Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides, *Nat Struct Biol*, **1994**, 1, (12), 863-70.
78. Liao D. I., Kapadia G., Ahmed H., Vasta G. R., Herzberg O., Structure of S-lectin, a developmentally regulated vertebrate beta-galactoside-binding protein, *Proc Natl Acad Sci U S A*, **1994**, 91, (4), 1428-32.
79. Hohenester E., Sasaki T., Timpl R., Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily, *Nat Struct Biol*, **1999**, 6, (3), 228-32.
80. Shimura K., Arata Y., Uchiyama N., Hirabayashi J., Kasai K., Determination of the affinity constants of recombinant human galectin-1 and -3 for simple saccharides by capillary affinity electrophoresis, *J Chromatogr B Analyt Technol Biomed Life Sci*, **2002**, 768, (1), 199-210.
81. Ahmad N., Gabius H. J., Kaltner H., Andre S., Kuwabara I., Liu F. T., Oscarson S., Norberg T., Brewer C. F., Thermodynamic binding studies of cell surface carbohydrate epitopes to galectins-1, -3, and -7: Evidence for differential binding specificities, *Can J Chem*, **2002**, 80, (8), 1096-104.
82. Abbott W. M., Hounsell E. F., Feizi T., Further studies of oligosaccharide recognition by the soluble 13 kDa lectin of bovine heart muscle. Ability to accommodate the blood-group-H and -B-related sequences, *Biochem J*, **1988**, 252, (1), 283-7.
83. Schwarz F. P., Ahmed H., Bianchet M. A., Amzel L. M., Vasta G. R., Thermodynamics of bovine spleen galectin-1 binding to disaccharides: correlation with structure and its effect on oligomerization at the denaturation temperature, *Biochemistry*, **1998**, 37, (17), 5867-77.
84. Dettmann W., Grandbois M., Andre S., Benoit M., Wehle A. K., Kaltner H., Gabius H. J., Gaub H. E., Differences in zero-force and force-driven kinetics of ligand dissociation from beta-galactoside-specific proteins (plant and animal lectins, immunoglobulin G) monitored by plasmon resonance and dynamic single molecule force microscopy, *Arch Biochem Biophys*, **2000**, 383, (2), 157-70.
85. Lobsanov Y. D., Gitt M. A., Leffler H., Barondes S. H., Rini J. M., X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-Å resolution, *J Biol Chem*, **1993**, 268, (36), 27034-8.
86. Murray C. W., Baxter C. A., Frenkel A. D., The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase, *J Comput Aided Mol Des*, **1999**, 13, (6), 547-62.
87. Leach A. R., Kuntz I. D., Conformational analysis of flexible ligands in macromolecular receptor sites., *J Comp Chem*, **1992**, 13, 730-48.
88. Ahmed H., Allen H. J., Sharma A., Matta K. L., Human splenic galactin: carbohydrate-binding specificity and characterization of the combining site, *Biochemistry*, **1990**, 29, (22), 5315-9.
89. Sybyl 7.2, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
90. Meyer B., Conformational Aspects of Oligosaccharides, *Top. Curr. Chem.*, **1990**, 154, 141-208.
91. Drickamer K., Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins, *Structure*, **1997**, 5, (4), 465-8.
92. Elgavish S., Shaanan B., Lectin-carbohydrate interactions: different folds, common recognition principles, *Trends Biochem Sci*, **1997**, 22, (12), 462-7.
93. Meyer J. E., Schulz G. E., Energy profile of maltooligosaccharide permeation through maltoporin as derived from the structure and from a statistical analysis of saccharide-protein interactions, *Protein Sci*, **1997**, 6, (5), 1084-91.
94. Quijcho F. A., Vyas N. K., Atomic interactions between proteins/enzymes and carbohydrates., *Bioorganic chemistry. Carbohydrates* 441-57, Oxford University Press, New York.
95. Rao V. S., Lam K., Qasba P. K., Architecture of the sugar binding sites in carbohydrate binding proteins--a computer modeling study, *Int J Biol Macromol*, **1998**, 23, (4), 295-307.
96. Spiwok V., Lipovova P., Skalova T., Buchtelova E., Hasek J., Kralova B., Role of CH/π interactions in substrate binding by *Escherichia coli* beta-galactosidase, *Carbohydr Res*, **2004**, 339, (13), 2275-80.

97. Sujatha M. S., Sasidhar Y. U., Balaji P. V., Insights into the role of the aromatic residue in galactose-binding sites: MP2/6-311G++** study on galactose- and glucose-aromatic residue analogue complexes, *Biochemistry*, **2005**, *44*, (23), 8554-62.
98. del Carmen Fernandez-Alonso M., Canada F. J., Jimenez-Barbero J., Cuevas G., Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions, *J Am Chem Soc*, **2005**, *127*, (20), 7379-86.
99. Wüthrich K., 6.4. Nuclear Overhauser Effects, ¹H-¹H Distances and Molecular Dynamics, In: Wüthrich K., *NMR of Proteins and Nucleic Acids*, 1 ed., 111-3, John Wiley & Sons, New York, **1986**.
100. Tripos Inc., Tripos mol2 file format, <http://www.tripos.com/data/support/mol2.pdf>.
101. Guha R., Howard M. T., Hutchison G. R., Murray-Rust P., Rzepa H., Steinbeck C., Wegner J., Willighagen E. L., The Blue Obelisk-interoperability in chemical informatics, *J Chem Inf Model*, **2006**, *46*, (3), 991-8.
102. The Open Babel Package 1.100, <http://openbabel.sourceforge.net/>
103. Banck, Michael, Hutchison, G. L., Murray-Rust, Peter, and Stahl, Matt, OpenBabel, http://openbabel.sourceforge.net/wiki/Main_Page.
104. Carugo O., Argos P., Correlation between side chain mobility and conformation in protein structures, *Protein Eng*, **1997**, *10*, (7), 777-87.
105. Sparrow C. P., Leffler H., Barondes S. H., Multiple soluble beta-galactoside-binding lectins from human lung, *J Biol Chem*, **1987**, *262*, (15), 7383-90.
106. Leffler H., Barondes S. H., Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian beta-galactosides, *J Biol Chem*, **1986**, *261*, (22), 10119-26.
107. Leppanen A., Stowell S., Blixt O., Cummings R. D., Dimeric galectin-1 binds with high affinity to alpha2,3-sialylated and non-sialylated terminal N-acetyllactosamine units on surface-bound extended glycans, *J Biol Chem*, **2005**, *280*, (7), 5549-62.
108. Ford M. G., Weimar T., Kohli T., Woods R. J., Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity, *Proteins*, **2003**, *53*, (2), 229-40.
109. May A. P., Robinson R. C., Vinson M., Crocker P. R., Jones E. Y., Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 Å resolution, *Mol Cell*, **1998**, *1*, (5), 719-28.
110. Sorme P., Arnoux P., Kahl-Knutsson B., Leffler H., Rini J. M., Nilsson U. J., Structural and thermodynamic studies on cation- π interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction, *J Am Chem Soc*, **2005**, *127*, (6), 1737-43.
111. Siebert H. C., Andre S., Lu S. Y., Frank M., Kaltner H., van Kuik J. A., Korchagina E. Y., Bovin N., Tajkhorshid E., Kaptein R., Vliegenthart J. F., von der Lieth C. W., Jimenez-Barbero J., Kopitz J., Gabius H. J., Unique conformer selection of human growth-regulatory lectin galectin-1 for ganglioside GM1 versus bacterial toxins, *Biochemistry*, **2003**, *42*, (50), 14762-73.
112. Brewer C. F., Thermodynamic binding studies of galectin-1, -3 and -7, *Glycoconj J*, **2004**, *19*, (7-9), 459-65.
113. Gilleron M., Siebert H. C., Kaltner H., von der Lieth C. W., Kozar T., Halkes K. M., Korchagina E. Y., Bovin N. V., Gabius H. J., Vliegenthart J. F., Conformer selection and differential restriction of ligand mobility by a plant lectin--conformational behaviour of Galbeta1-3GlcNAcbeta1-R, Galbeta1-3GalNAcbeta1-R and Galbeta1-2Galbeta1-R' in the free state and complexed with galactoside-specific mistletoe lectin as revealed by random-walk and conformational-clustering molecular-mechanics, *Eur J Biochem*, **1998**, *252*, (3), 416-27.
114. Siebert H. C., Gilleron M., Kaltner H., von der Lieth C. W., Kozar T., Bovin N., Korchagina E. Y., Vliegenthart J. F., Gabius H. J., NMR-based, molecular dynamics- and random walk molecular mechanics-supported study of conformational aspects of a carbohydrate ligand (Gal beta 1-2Gal beta 1-R) for an animal galectin in the free and in the bound state, *Biochem Biophys Res Commun*, **1996**, *219*, (1), 205-12.
115. Alonso H., Bliznyuk A. A., Gready J. E., Combining docking and molecular dynamic simulations in drug design, *Med Res Rev*, **2006**.
116. Gasteiger J., Marsili M., A new model for calculating atomic charges in molecules, *Tetrahedron Lett.*, **1978**, (34), 3181-4.
117. Gasteiger J., Marsili M., Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges, *Tetrahedron*, **1980**, *36*, (22), 3219-22.

8 Appendix

8.1 Pulse programs

In the following section the pulse programs used in this study are documented

8.1.1 STD-HSQC

This pulse program records a STD-HSQC spectrum. The on and *off resonance* spectra are stored in a single serial file. This file is split into the separate spectra and processed.

```

;stdiietgpsl.mul
;avance-version (02/05/09)
;STD-HSQC
;2D H-1/X correlation via double inept transfer
; using sensitivity improvement
;phase sensitive using Echo/Antiecho-TPPI gradient selection
;with decoupling during acquisition
;using trim pulses in inept transfer
;A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn.
; Reson. 93, 151-170 (1991)
;L.E. Kay, P. Keifer & T. Saarinen, J. Am. Chem. Soc. 114,
; 10663-5 (1992)
;J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletsky,
; S.J. Glaser, O.W. Sorensen & C. Griesinger, J. Biomol. NMR 4,
; 301-306 (1994)
;
;STD sequence on and off resonance recorded interleaved spectra.
;Frequency alternates after each increment
;Requires 2 frequency lists!!!
;See comments at the end of the file!!
;CHECK CAREFULLY YOUR SETTING FOR TD
;Written by A.J.Benie

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p2=p1*2"
"p4=p3*2"
"d0=3u"
"d4=1s/(cnst2*4)"
"d11=30m"
"d13=4u"

"DELTA=p16+d16+50u+p2+d0*2"
"DELTA1=d13+p16+d16+4u"

"l0=0"

1 ze
  d11 p112:f2
2 d1 do:f2
      ;when l0 is odd perform on resoance spectra
  if "l0 %2 == 1" goto 100
      ;perform off resoance spectra
  d12 fq1:f1
5 p11:sp1:f1
  d31
  lo to 5 times l6
  d12 fq1:f1
      ;jump past on resoance stuff
  goto 101
      ;perform on resoance spectra
100 d12 fq2:f1
6 p11:sp1:f1
  d31
  lo to 6 times l6
  d12 fq2:f1

```

```

101 d13
                                ;end of STD stuff

3 (p1 ph1)
  d4 p12:f2
  (p2 ph1) (p4 ph6):f2
  d4
; p28 ph1 KEINE TRIMPULSE AUF CRYOKOPF !!!!!!!!!!!!!!!
d13
  (p1 ph2) (p3 ph3):f2
  d0
  p2 ph7
  d0
  50u UNBLKGRAD
  p16:gp1*EA
  d16
  (p4 ph4):f2
  DELTA
  (p1 ph1) (p3 ph4):f2
  d24
  (p2 ph1) (p4 ph1):f2
  d24
  (p1 ph2) (p3 ph5):f2
  d4
  (p2 ph1) (p4 ph1):f2
  d4
  (p1 ph1)
  DELTA1
  (p2 ph1)
  d13
  p16:gp2
  d16 p112:f2
  4u BLKGRAD
  go=2 ph31 cpd2:f2
  d1 do:f2 mc #0 to 2
    FLI(iu0, 2)
    FLEA(igrad EA & ip5*2, id0 & ip3*2 & ip6*2 & ip31*2)
exit

ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=1 1 3 3
ph6=0
ph7=0 0 2 2
ph31=0 2 2 0

;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default)
;p112: f2 channel - power level for CPD/BB decoupling
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel - 90 degree high power pulse
;p4 : f2 channel - 180 degree high power pulse
;sp1: f1 channel - shaped pulse
;p11: f1 channel - 90 degree shaped pulse
;p16: homospoil/gradient pulse
;p28: f1 channel - trim pulse
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I/O [30 msec]
;d13: short delay [4 usec]
;d16: delay for homospoil/gradient recovery
;d24: 1/(4J)XH for XH
; 1/(8J)XH for all multiplicities
;d31: delay between pulses for saturation

```

```

;l6 : overall irradiation time: (p11+d31) * 16
;cnst2: = J(XH)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 1 * n
;DS: >= 16
;td1: total number of experiments
;      = number of experiments for each 2D*2
;FnMODE: echo-antiecho
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence
;use gradient ratio:  gp 1 : gp 2
;                      80 : 20.1   for C-13
;                      80 :  8.1   for N-15
;for z-only gradients:
;gpz1: 80%
;gpz2: 20.1% for C-13, 8.1% for N-15
;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100
;define FQ1LIST (list has to be stored in "/u/exp/stan/nmr/lists/fl")
;O spectrometer freq for 1H
;off resonance freq
;H2O freq
;define FQ2LIST (list has to be stored in "/u/exp/stan/nmr/lists/fl")
;O spectrometer freq for 1H
;on resonance freq
;H2O freq
;use AU-program splitinvnoe to separate On and Off resoance spectra into
;  different datasets
;use AU-program 2dstd.mul to subtract On and Off spectra from one another

```

8.1.2 1D STD NMR spectra with separate *on* and *off resonance* FID's

The following pulse programs were used for acquisition of 1D 1H STD spectra. The std.t4 pulse programs were used on the 700 MHz Avance spectrometer with digital signal generation unit (SGU). The pulse programs store *on* and *off resonance* spectra in a serial file. This file is split prior to processing using the *splitser* command.

The following program uses the 3-9-19 WATERGATE phase cycle for water suppression.

```

;std.t4_3.bc
;avance-version (03/10/10)
;1D sequence
;  for saturation transfer difference
;with shaped pulse train for saturation on f2 channel
;with spinlock to suppress protein signals
;alternating between on and off resonance
;  to be defined by fq2list
;
;M. Mayer & B. Meyer, Angew. Chem. Int. Ed. 38, 1784-1788 (1999)
;M. Mayer & B. Meyer, Angew. Chem. 111, 1902-1906 (1999).

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"d12=20u"
"p29=d29"
"15=d20/p13+0.5"
"d31=p13*15"

"DELTA1=d1-d31"

1 ze
  10u st0
2 30m
  4u BLKGRAD
  d12 fq2:f2 st

  50u UNBLKGRAD

```

```

4u p110:f1
(p17 ph2)
(p17*2 ph3)
4u
p30:gp1
10m p11:f1
4u BLKGRAD

DELTA1

3 (p13:sp13 ph4):f2
4u
lo to 3 times 15

p1 ph1

4u p129:f1
(p29 ph21)

50u UNBLKGRAD
p16:gp2
d16 p118:f1
p28*0.23077 ph5
d19*2
p28*0.69231 ph5
d19*2
p28*1.46154 ph5
d19*2
p28*1.46154 ph6
d19*2
p28*0.69231 ph6
d19*2
p0*0.23077 ph6
46u
p16:gp2
d16
go=2 ph31
30m wr #0 if #0
4u BLKGRAD
d31
exit

ph1=0 0 2 2
ph2=0
ph3=1
ph4=0
ph5=0 0 1 1 2 2 3 3
ph6=2 2 3 3 0 0 1 1
ph21=1 1 3 3
ph31=0 2 0 2

;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default) [120 dB]
;p110: f1 channel - power level for TOCSY-spinlock
;p129: f1 channel - power level as for TOCSY-spinlock
;sp13: f2 channel - shaped pulse for saturation [35 - 60 dB]
;p1 : f1 channel - high power pulse
;p13: f2 channel - shaped pulse for saturation [50 msec]
;p17: f1 channel - trim pulse [2.5 msec]
;p29: f1 channel - trim pulse
;p30: gradient pulse [5 msec]
;d1 : relaxation delay: 1-5 * T1
;d12: delay for power switching [20 usec]
;d20: saturation time
;d29: spinlock time [10 - 50 msec]
;d31: saturation time as executed
;l5: loop for saturation: (p11 + d20) * 15 = saturation time
;NS: 16 * n, total number of scans: NS * TD0
;DS: 16

```

```

;gp1: gradient for spinlock
;gp2: gradient for watersuppression
;define FQ2LIST (irradiation frequencies, alternating on/off resonance)
;      (list has to be stored in "/u/exp/stan/nmr/lists/fl")
;use gradient ratio:   gp 1
;                      40
;for z-only gradients:
;gpz1: 40%
;gpz2: 20%
;use gradient files:
;gpnam1: SINE.100
;$Id: $

```

The following pulse sequence uses the w5 WATERGATE phase cycle for water suppression.

```

;std.t4_w5.th
;avance-version (03/10/10)
;1D sequence
;  for saturation transfer difference
;with shaped pulse train for saturation on f2 channel
;with spinlock to suppress protein signals
;alternating between on and off resonance
;  to be defined by fq2list
;
;M. Mayer & B. Meyer, Angew. Chem. Int. Ed. 38, 1784-1788 (1999)
;M. Mayer & B. Meyer, Angew. Chem. 111, 1902-1906 (1999).

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"d12=20u"
"p29=d29"
"l5=d20/p13+0.5"
"d31=p13*15"

"DELTA1=d1-d31"

1 ze
  10u st0
2 30m
  4u BLKGRAD
  d12 fq2:f2 st

  50u UNBLKGRAD
  4u pl10:f1
  (p17 ph2)
  (p17*2 ph3)
  4u
  p30:gp1
  10m pl1:f1
  4u BLKGRAD

DELTA1

3 (p13:sp13 ph4):f2
  4u
  lo to 3 times l5

  p1 ph1

  4u pl29:f1
  (p29 ph21)

  50u UNBLKGRAD
  p16:gp2
  d16 pl18:f1
  p28*0.0867 ph5
  d19*2
  p28*0.2056 ph5
  d19*2

```

```

p28*0.4133 ph5
d19*2
p28*0.7822 ph5
d19*2
p28*1.4911 ph5
d19*2
p28*1.4911 ph6
d19*2
p28*0.7822 ph6
d19*2
p28*0.4133 ph6
d19*2
p28*0.2056 ph6
d19*2
p0*0.0867 ph6
46u
p16:gp2
d16
go=2 ph31
30m wr #0 if #0
4u BLKGRAD
d31
exit

```

```

ph1=0 0 2 2
ph2=0
ph3=1
ph4=0
ph5=0 0 1 1 2 2 3 3
ph6=2 2 3 3 0 0 1 1
ph21=1 1 3 3
ph31=0 2 0 2

```

```

;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default) [120 dB]
;p110: f1 channel - power level for TOCSY-spinlock
;p129: f1 channel - power level as for TOCSY-spinlock
;sp13: f2 channel - shaped pulse for saturation [35 - 60 dB]
;p1 : f1 channel - high power pulse
;p13: f2 channel - shaped pulse for saturation [50 msec]
;p17: f1 channel - trim pulse [2.5 msec]
;p29: f1 channel - trim pulse
;p30: gradient pulse [5 msec]
;d1 : relaxation delay: 1-5 * T1
;d12: delay for power switching [20 usec]
;d20: saturation time
;d29: spinlock time [10 - 50 msec]
;d31: saturation time as executed
;l5: loop for saturation: (p11 + d20) * 15 = saturation time
;NS: 16 * n, total number of scans: NS * TD0
;DS: 16
;gp1: gradient for spinlock
;gp2: gradient for watersuppression
;define FQ2LIST (irradiation frequencies, alternating on/off resonance)
; (list has to be stored in "/u/exp/stan/nmr/lists/fl")
;use gradient ratio: gp 1
; 40
;for z-only gradients:
;gpz1: 40%
;gpz2: 20%
;use gradient files:
;gpnaml: SINE.100
;$Id: $

```

The following pulse program uses a single low power pulse during acquisition as spinlock for suppression of protein resonances. Water suppression is achieved by the w5 WATERGATE sequence. This was used on the 500 MHz spectrometer with analogue signal generation unit.

```

;stdw5slsp2d.bc
;M. Mayer; B. Meyer, Department of Chemistry
;University of Hamburg, Germany
;email: bernd_meyer@sg11.chemie.uni-hamburg.de
;avance-version
;1D difference sequence with f2 presaturation defined by frequency list
;presaturation by shaped pulses
;frequency alternates after every scan, defined by fqllist
;using different memory buffers for on- and off-resonance irradiation
;spin lock for protein suppression
;water suppression by watergate with w5 pulse
;use w5gpsl.rm to optimize parameters
;define 1H on channel f2 in edasp

#include <Avance.incl>
#include <Grad.incl>

1 ze
  10u st0
2 4u BLKGRAD
  20u p11:f1
  d7 fq1:f2 st
3 p11:sp1:f2
  d11
  lo to 3 times 17
  p1 ph1
  20u p110:f1
  p10 ph2
  50u UNBLKGRAD
  p16:gp1
  d16 p118:f1
  p28*0.0867 ph3
  d19*2
  p28*0.2056 ph3
  d19*2
  p28*0.4133 ph3
  d19*2
  p28*0.7822 ph3
  d19*2
  p28*1.4911 ph3
  d19*2
  p28*1.4911 ph4
  d19*2
  p28*0.7822 ph4
  d19*2
  p28*0.4133 ph4
  d19*2
  p28*0.2056 ph4
  d19*2
  p0*0.0867 ph4
  46u
  p16:gp2
  d16
  go=2 ph31
  30m wr #0 if #0
  4u BLKGRAD
exit

ph1=0 2
ph3=0 0 1 1 2 2 3 3
ph4=2 2 3 3 0 0 1 1
ph31=0 0 2 2

;*****Power Level*****
;p11 : f1 channel - power level for pulse (default)
;p12 : 120dB (use sp1 for adjusting power of shaped pulse)
;p118: f1 channel - power level for w5-pulse (watergate 12dB)
;p110 : f1 channel - power level for spin lock pulse (10-15 dB)
;sp1 : f2 - channel - power level for shaped pulse
;between 50 - 60 dB depending on protein and ligand
;

```

```

;*****Pulse*****
;p1 : f1 channel - 90 degree high power pulse
;p10 : f1 channel - spin lock pulse for protein suppr. (10-30 ms, depending on the protein)
;p11 : f2 channel - presaturation shaped pulse (gauss ca. 50 msec)
;p28: f1 channel - 90 degree pulse at p118
;p0 : f1 channel - 90 degree pulse at p118
;
;          use for fine adjustment
;
;*****Delays*****
;d1 : relaxation delay; 1-5 * T1
;d7 : additional delay (if nessesary) for complete T1 relaxation [min 20usec]
;d11 : delay between shaped pulses [1msec]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;      d19 = (1/(2*d)), d = distance of next null (in Hz)
;      d19 should be around 150-220 usec.
;
;presaturation = (p11 + d11) * 17  (presaturation should be around 2 sec)
;
;fq1 : define frequencies for on and off resonance presaturation
;      0 499.87000 off resonance 1x(15-20000 HZ) on resonance 1x(xxx HZ)
;      on frequency list f1
;NBL = number of memory buffers with TD size = 2 for two irradiation frequencies
;td1 = NBL = 2
;this pulse program produces a ser file (PARMOD = 2D)
;NS" = NS*2+DS"
;DS" = DS/2 .
;use gradient ratio      gp1      :      gp2
;                          20          20

```

This program is used to determine saturation transfer to the protein. No spinlock pulse was employed during acquisition.

```

;stdw5sp.rm
;M. Mayer; B. Meyer, Department of Chemistry
;University of Hamburg, Germany
;email: bernd_meyer@sgil.chemie.uni-hamburg.de
;avance-version
;1D difference sequence with f2 presaturation defined by frequency list
;presaturation by shaped pulses
;frequency alternates after every scan, defined by fq1list
;using different memory buffers for on- and off-resonance irradiation
;water suppression by watergate with w5 pulse
;use w5gp.rm to optimize parameters
;define 1H on channel f2 in edasp

#include <Avance.incl>
#include <Grad.incl>

```

```

1 ze
  10u st0
2 4u BLKGRAD
  20u p11:f1
  d7 fq1:f2 st
3 p11:sp1:f2
  d11
  lo to 3 times 17
  p1 ph1
  50u UNBLKGRAD
  p16:gp1
  d16 p118:f1
  p28*0.0867 ph3
  d19*2
  p28*0.2056 ph3
  d19*2
  p28*0.4133 ph3
  d19*2
  p28*0.7822 ph3
  d19*2
  p28*1.4911 ph3
  d19*2

```

```

p28*1.4911 ph4
d19*2
p28*0.7822 ph4
d19*2
p28*0.4133 ph4
d19*2
p28*0.2056 ph4
d19*2
p0*0.0867 ph4
46u
p16:gp2
d16
go=2 ph31
30m wr #0 if #0
4u BLKGRAD
exit

ph1=0 2
ph3=0 0 1 1 2 2 3 3
ph4=2 2 3 3 0 0 1 1
ph31=0 0 2 2

;*****Power Level*****
;p11 : f1 channel - power level for pulse (default)
;p12 : 120dB (use spl for adjusting power of shaped pulse)
;p118: f1 channel - power level for 3-9-19-pulse (watergate 12dB)
;spl : f2 - channel - power level for shaped pulse
;between 50 - 60 dB depending on protein and ligand
;
;*****Pulse*****
;p1 : f1 channel - 90 degree high power pulse
;p11 : f2 channel - presaturation shaped pulse (gauss ca. 50 msec)
;p28: f1 channel - 90 degree pulse at p118
;p0 : f1 channel - 90 degree pulse at p118
;
;          use for fine adjustment
;
;*****Delays*****
;d1 : relaxation delay; 1-5 * T1
;d7 : additional delay (if nessesary) for complete T1 relaxation [min 20usec]
;d11 : delay between shaped pulses [1msec]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;      d19 = (1/(2*d)), d = distance of next null (in Hz)
;      d19 should be around 150-220 usec.
;
;presaturation = (p11 + d11) * 17 (presaturation should be around 2 sec)
;
;fq1 : define frequencies for on and off resonance presaturation
;      0 499.87000 off resonance 1x(15-20000 HZ) on resonance 1x(XXX HZ)
;      on frequency list f1.
;NBL = number of memory buffers with TD size = 2 for two irradiation frequencies
;td1 = NBL = 2
;this pulse program produces a ser file (PARMOD = 2D)
;NS" = NS*2+DS"
;DS" = DS/2
;use gradient ratio      gp1      :      gp2
;                          20          20

```

8.1.3 1D STD NMR spectra with internal subtraction

The following STD NMR pulse programs stores the FID of the STD spectrum only. On and *off resonance* FID are subtracted internally. Water suppression is achieved by 3-9-19 WATERGATE phase cycling. STD intensity is determined from comparison with standard WATERGATE NMR spectrum with the same settings.

This program employs a low power spinlock pulse during acquisition to suppress protein resonances.

```

;std19slsp
;M. Mayer; B. Meyer, Department of Chemistry

```

```

;University of Hamburg, Germany
;email: bernd_meyer@sgil.chemie.uni-hamburg.de
;avance-version
;1D difference sequence with f2 presaturation defined by frequency list
;presaturation by shaped pulses
;frequency alternates after every scan, defined by fqllist
;spin lock for protein suppression
;water suppression by watergate, use p3919gp to optimize parameters
;define 1H on channel f2 in edasp

#include <Avance.incl>
#include <Grad.incl>

1 ze
2 20u p11:f1
   d7 fq1:f2
3 p11:sp1:f2
   d11
   lo to 3 times 17
   p1 ph1
   20u p110:f1
   p10 ph2
   50u UNBLKGRAD
   p16:gp1
   d16 p118:f1
   p28*0.231 ph3
   d19*2
   p28*0.692 ph3
   d19*2
   p28*1.462 ph3
   d19*2
   p28*1.462 ph4
   d19*2
   p28*0.692 ph4
   d19*2
   p0*0.231 ph4
   46u
   p16:gp2
   d16
   4u BLKGRAD
   go=2 ph31
   wr #0
exit

ph1=0 2
ph2=1 3
ph3=0 0 1 1 2 2 3 3
ph4=2 2 3 3 0 0 1 1
ph31=0 0 2 2

;*****Power Level*****
;p11 : f1 channel - power level for pulse (default)
;p118: f1 channel - power level for 3-9-19-pulse (watergate 12dB)
;p110 : f1 channel - power level for spin lock pulse (10-15 dB)
;sp1 : f2 - channel - power level for shaped pulse
;between 50 - 60 dB depending on protein and ligand
;
;*****Pulse*****
;p0 : f1 channel - 90 degree pulse at p118
;
;      use for fine adjustment
;p1 : f1 channel - 90 degree high power pulse
;p10 : f1 channel - spin lock pulse for protein suppr. (10-30 ms, depending on the protein)
;p11 : f2 channel - presaturation shaped pulse (gauss ca. 50 msec)
;p28: f1 channel - 90 degree pulse at p118
;
;*****Delays*****
;d1 : relaxation delay; 1-5 * T1
;d7 : additional delay (if necessary) for complete T1 relaxation [min 20usec]
;d11 : delay between shaped pulses [1msec]
;d16: delay for homospoil/gradient recovery

```

```

;d19: delay for binomial water suppression
;   d19 = (1/(2*d)), d = distance of next null (in Hz)
;   d19 should be around 150-220 usec.
;
;presaturation = (p11 + d11) * 17   (presaturation should be around 2 sec)
;
;fq1 : define frequencies for on and off resonance presaturation
;   0 499.87000 off resonance 1x(15-20000 HZ) on resonance 1x(XXX HZ)
;   on frequency list f1.
;NS = 16*n
;DS = 16
;use gradient ratio   gp1   :   gp2
;                       20     20

```

This program does not include a spinlock pulse. It is used to determine saturation of proteins.

```

;std19sp
;M. Mayer; B. Meyer, Department of Chemistry
;University of Hamburg, Germany
;email: bernd_meyer@sgil.chemie.uni-hamburg.de
;avance-version
;1D difference sequence with f2 presaturation defined by frequency list
;presaturation by shaped pulses
;frequency alternates after every scan, defined by fq1list
;water suppression by watergate, use p3919gp to optimize parameters
;define 1H on channel f2 in edasp

#include <Avance.incl>
#include <Grad.incl>

1 ze
2 20u p11:f1
   d7 fq1:f2
3 p11:sp1:f2
   d11
   lo to 3 times 17
   p1 ph1
   50u UNBLKGRAD
   p16:gp1
   d16 p118:f1
   p28*0.231 ph3
   d19*2
   p28*0.692 ph3
   d19*2
   p28*1.462 ph3
   d19*2
   p28*1.462 ph4
   d19*2
   p28*0.692 ph4
   d19*2
   p0*0.231 ph4
   46u
   p16:gp2
   d16
   4u BLKGRAD
   go=2 ph31
   wr #0
exit

ph1=0 2
ph3=0 0 1 1 2 2 3 3
ph4=2 2 3 3 0 0 1 1
ph31=0 0 2 2

;*****Power Level*****
;p11 : f1 channel - power level for pulse (default)
;p118: f1 channel - power level for 3-9-19-pulse (watergate 12dB)
;sp1 : f2 - channel - power level for shaped pulse
;between 50 - 60 dB depending on protein and ligand
;
;*****Pulse*****

```

```

;p1 : f1 channel - 90 degree high power pulse
;p0 : f1 channel - 90 degree pulse at p118
;
; use for fine adjustment
;p28: f1 channel - 90 degree pulse at p118
;p11 : f2 channel - presaturation shaped pulse (gauss ca. 50 msec)
;
;*****Delays*****
;d1 : relaxation delay; 1-5 * T1
;d7 : additional delay (if nessesary) for complete T1 relaxation [min 20usec]
;d11 : delay between shaped pulses [1msec]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;   d19 = (1/(2*d)), d = distance of next null (in Hz)
;   d19 should be around 150-220 usec.
;
;presaturation = (p11 + d11) * 17   (presaturation should be around 2 sec)
;
;fq1 : define frequencies for on and off resonance presaturation
;   0 499.87000 off resonance 1x(15-20000 HZ) on resonance 1x(xxx HZ)
;   on frequency list f1.
;NS = 16*n
;DS = 16
;use gradient ratio   gp1   :   gp2
;                   20     :   20

```

8.2 Programs and scripts

This is the source code of the distance mapping program employed in this thesis. Running the program requires that the libopenbabel is installed and in the search path. The program requires a configuration file to be given as argument when starting. The interactive mode is not supported anymore and was implemented solely for debugging purposes in the early development stages. For smooth operation of the program it is advised that the set containing the relevant atoms of the receptor is the first set in the according section of the mol2 file.

```

//system headers
#include<iostream>
#include<sstream>
#include<string>
#include<fstream>
#include<vector>
#include<map>
#include<algorithm>
#include<math.h>

//OpenBabel headers
#include "molvector.h"
#include "mol.h"
#include "obutil.h"
#include "fileformat.h"

using namespace std;
using namespace OpenBabel;

//functions
bool ConfigRead(map<string, string> &m, char filename[80])
{
    ifstream ifs;
    map<string, string> config;

    ifs.open(filename);
    if(!ifs)
        {cout << "File Error!\n"; return false;}
    else
        {
            while (!ifs.eof())
                {
                    bool comment = false;

```

```

    string key, value, s, temp;

    getline(ifs, s);
    if (s.size() > 0)
    {
        char first = s.at(0);
        if( first == '#' )
        {
            comment = true;
        }
    }

    istringstream instr(s);
    vector<string> line;
    while(instr >> temp && comment == false)
    {
        line.push_back(temp);
    }
    if(line.size() > 2)
        {cout << "Too many words in line, please check: "<< line.at(0) << endl; return
false;}
    if(line.size() < 2 && line.size() != 0)
        {cout << "Missing argument for: " << line.at(0) << "." << endl;}
    if(line.size() == 2)
    {
        key = line.at(0);
        m[key] += line.at(1);
        line.clear();
    }
}
}
map<string, string>::iterator i;
cout << "These values were read in(value):\n";
for (i=m.begin(); i!=m.end(); i++)
{
    cout <<i->first << "\t" << i->second << endl;
}
return true;
}

void StringExplode(vector<string> &v , string &str)
{
    istringstream instr(str);
    string temp;
    while (instr >> temp)
    {
        v.push_back(temp);
    }
}

bool MoreLines(string &s)
{
    string temp;
    int pos;
    vector<string> tempvec;

    StringExplode(tempvec, s);
    pos = tempvec.size()-1;
    temp = tempvec.at(pos);

    if(temp == "\\")
    {
        return true;
    }
    else
    {
        return false;
    }
}

```

```
    }
}

void StoreAtoms(istream &ifc, vector<long> &v)
{
    long temp;
    string tempstring, s;
    bool morelines = true;
    vector<long>::iterator pos;

    while(morelines == true)
    {
        getline(ifc, s);
        tempstring = s;
        istringstream instr(s);

        while(instr >> temp)
        {
            v.push_back(temp);
        }
        morelines = MoreLines(tempstring);
    }
    int noa = v.at(0);
    pos = v.begin();
    v.erase(pos);
    cout << "Atoms in set (readout): " << noa << "\n", in vector (size): " << v.size() << endl;
}

bool Test4Type(vector<string> &v, string &KeyWrd)
{
    string temp;
    bool SetTypeTest;
    temp = v.at(1);
    if(temp == KeyWrd)
    {
        SetTypeTest = true;
        return true;
    }
    else
    {
        return false ;
    }
    temp = v.at(2);
    if(temp == "SUBSTS")
    {
        cout << "Attention, this set contains the IDs of substructures, NOT atoms" << endl;
        //warning, that the ids read in are subststs ids
    }
}

bool Test4Name(vector<string> &v, string &KeyWrd)
{
    string temp;
    bool SetNameTest;
    temp = v.at(0);
    if(temp == KeyWrd)
    {
        SetNameTest = true;
        cout << "Found set named: " << KeyWrd << endl;
        return true;
    }
    else
    {
        return false;
    }
}

void LoadEpitope(char file[], vector<long> &v, string &Sname, string &Stype)
```

```

{
  ifstream ifs;
  ifs.open(file);
  if(!ifs)
  {
    cout << "LOADEPITOPE: Error while opening " << file << ".\n"; // may give alternative
name??
  }
  else
  {
    cout << "File found." << endl;

    bool SETdefn = false;
    string line;

    while(!ifs.eof() && SETdefn == false)
    {
      getline(ifs, line);
      if(line == "@<TRIPOS>SET")
      {
        cout << "Set(s) found." << endl;

        SETdefn = true;
      }
    }

    vector<string> StatusLine;
    while(getline (ifs, line))
    {

      StringExplode(StatusLine, line);
      bool SETstatic = Test4Type(StatusLine, Stype);
      bool SETnametest = Test4Name(StatusLine, Sname);

      if(SETstatic == true && SETnametest == true)
      {
        break;
      }
      else
      {
        StatusLine.clear();
      }
    }

    if (Stype == "STATIC")
    {
      StoreAtoms(ifs, v);
      cout << "Number of Atoms in Epitope: " << v.size() << endl;
    }

    ifs.close();
  }
}

double Distance(float &firstX, float &firstY, float &firstZ, float &secX, float &secY, float
&secZ)
{
  double dist;
  dist = sqrt(pow((firstX-secX),2)+pow((firstY-secY),2)+pow((firstZ-secZ),2));
  return dist;
}

double RecpPow6(double &dist)
{
  //calculates the reciprocal of the distance to the power of 6
  //TSintensd: theoretical STD intensity by distance

  double TSintensd = 1/(pow(dist,6));
  return TSintensd;
}

```

```
}

void StoreHistogram(OBAtom &LIGatom, double &dist, map<long, double> &m)
{
    double temp;
    long index = LIGatom.GetIdx();
    map<long, double>::iterator key;
    key = m.find(index);
    if(key == m.end())
    {
        m[index] = RecpPow6(dist);
    }
    else
    {
        temp = m[index];
        m[index] = temp + RecpPow6(dist);
    }
}

void WriteResults(vector< map<long, double> > &v, ofstream &ofs)
{
    for (unsigned int i = 0; i < v.size(); i++)
    {
        ofs << "Pose number " << i << endl;

        map<long, double> m = v.at(i);
        map<long, double>::iterator it;
        for(it = m.begin(); it != m.end(); it++)
        {
            ofs << it->first << "\t" << it->second << endl;
        }
    }
    ofs.close();
}

void WriteResults(map<long, double> &m, ofstream &ofs)
{
    ofs << "This is the clustered result." << endl;
    map<long, double>::iterator it;
    for (it = m.begin(); it!=m.end(); it++)
    {
        ofs << it->first << "\t" << it->second << endl;
    }
    ofs.close();
}

int main (int argc, char **argv)
{
    int TEST = 0;
    long i;
    float LIGx, LIGy, LIGz, RECx, RECy, RECz, MinDist, MaxDist;
    // MinDist shouldnt be lower than 1.8A, MaxDist is usually around 6A

    bool configfile;

    string SETname, SETtype, LIGstring, NRGstring, RECstring, LigSetName, RecSetName, MiniDist,
    MaxiDist;
    string RECfilename, LIGfilename, NRGfilename, RECfile, LIGfile, NRGfile, OutFile;
    string FileExtension = ".mol2";
    string NRGsuffix = "_nrg";
    string ResultStyle; // c: cluster only, i: individual only, a: all
    string CommandLineError = "Wrong or too many arguments!\n Usage: distancemap -i: interactive
mode (reads from standard input)\n\t distancemap -f [filename]: reads commmands from config
file filename (defaults to mapconfig). ";
    // REC: filename for the receptor/protein
    // LIG: template file of ligand
    // NRG dock result file, usually with nrg score
    // file : no file extension to be given
```

```

char title[80]; //stores title of OBMol data

vector<string> StatusLine;
vector<long> Epitope;
vector<long> BindingSite;

map<string, string> Config;
// map<long, double> Histogram;
typedef map<long, double> Histogram; // definition of a map holding <Idx, RecPow6(dist)>
vector<Histogram> AllPoses; // definition of a vector holding histograms (here:
for all docked poses)
vector<Histogram> AllResults; // definition of a vector holding histograms (here:
all poses and clustered result)
Histogram Cluster; // definition of a simple map (here: only the
clustered result)

ifstream RECifs, NRGifs;

if (argc == 1)
    {configfile = false;}
if (argc == 2)
    {
        bool FailBit = ConfigRead(Config, argv[1]); //Config: map to store variable/value pairs
        if ( FailBit == false)
            {return 0;}
        else
            {configfile = true; cout <<"INNER TEST: " << configfile <<endl;}
    }

// cout << "TEST: " <<argc<< "\t" << configfile << "\t" << argv[1] <<endl;
if (configfile == true)
    {
        LIGfile = Config["LIGfile"];
        RECfile = Config["RECfile"];
        SETtype = Config["SETtype"];
        LigSetName = Config["LigSetName"];
        RecSetName = Config["RecSetName"];
        ResultStyle = Config["ResultStyle"];
        NRGsuffix = Config["NRGsuffix"];
        FileExtension = Config["FileExtension"];
        // stuff needed for string to float conversion
        string tempstring;
        map<string, string>::const_iterator key;
        key = Config.find("MiniDist");
        if(key == Config.end())
            {MinDist = 1.8;}
        else
            {
                {
                    tempstring = Config["MiniDist"];
                    istream tempstream(tempstring);
                    tempstream >> MinDist ;
                }
            }
        key = Config.find("MaxiDist");
        if(key == Config.end())
            {MaxDist = 6;}
        else
            {
                {
                    tempstring = Config["MaxiDist"];
                    istream tempstream(tempstring);
                    tempstream >> MaxiDist;
                }
            }
        // identifying the name for the output file
        key = Config.find("OutFile");
        if(key != Config.end())
            {
                {
                    OutFile = Config["OutFile"];
                }
            }

        LIGfilename = LIGfile + FileExtension;
        char* LIGfil = new char[LIGfilename.length()+1]; // need extra char for '\0'-char
    }

```

```

// character arrays needed for the
io-stream methods (dont
// work with strings)

LIGfilename.copy(LIGfil,string::npos);
LIGfil[LIGfilename.length()] = 0;

cout << "Name: " << LIGfile << " Filename: " << LIGfilename << endl;

LoadEpitope(LIGfil, Epitope, LigSetName, SETtype);

RECfilename = RECfile + FileExtension;
char* RECfil = new char[RECfilename.length()+1];
RECfilename.copy(RECfil,string::npos);
RECfil[RECfilename.length()] = 0;

LoadEpitope(RECfil, BindingSite, RecSetName, SETtype);

}

else
{

cout << "This program assumes the files to be read in to be .mol2 files (ligand.mol2).
\n";
cout << "The .mol2 extension is automatically attached to the name provided by the user.
\n";
cout << "The results from docking experiments are assumed to be stored in files of like
\n";
cout << "'ligand_nrg.mol2', the '_nrg.mol2' is attached to the name provided by the
user.\n ";
cout << "Thus only the name of the molecule (i.e.ligand) must be provided. The receptor
file is\n";
cout << "assumed to be in mol2 format and the mol2 extension is attached
automaically.\n\n ";

cout << "Please enter name of the ligand to be opened (without .mol2 extension, 128
chars max.): ";
getline(cin, LIGfile);

LIGfilename = LIGfile + FileExtension;
NRGfilename = LIGfile + NRGsuffix + FileExtension;

char* LIGfil = new char[LIGfilename.length()+1]; // need extra char for '\0'-char
// character arrays needed for the
io-stream methods (dont
// work with strings)

LIGfilename.copy(LIGfil,string::npos);
LIGfil[LIGfilename.length()] = 0;

cout << "Type in name of the set you want to read:\t";
getline(cin, SETname);

cout << "Type of set: 1) STATIC, 2)DYNAMIC (checks for existance only).";
cin >> i;
switch(i)
{
case 1:
SETtype = "STATIC";
break;
case 2:
SETtype = "DYNAMIC";
break;
default:
cout << "Invalid option, please type 1 or 2!\n";
return 0;
}
}

```

```

LoadEpitope(LIGfil, Epitope, SETname, SETtype);

cout << "Please enter name of the receptor to be opened (without .mol2 extension, 128
chars max.): ";
cin >> RECfile;

RECfilename = RECfile + FileExtension;
char* RECfil = new char[RECfilename.length()+1];
RECfilename.copy(RECfil,string::npos);
RECfil[RECfilename.length()] = 0;

cout << "Type in name of the set you want to read:\t";
cin >> SETname;

cout << "Type of set: 1) STATIC, 2)DYNAMIC (checks for existance only).";
cin >> i;
switch(i)
{
case 1:
    SETtype = "STATIC";
    break;
case 2:
    SETtype = "DYNAMIC";
    break;
default:
    cout << "Invalid option, please type 1 or 2!\n";
    return 0;
}

LoadEpitope(RECfil, BindingSite, SETname, SETtype);
}

RECfilename = RECfile + FileExtension;
char* RECfil = new char[RECfilename.length()+1];
RECfilename.copy(RECfil,string::npos);
RECfil[RECfilename.length()] = 0;

RECifs.open(RECfil);
OBMol Receptor;
Receptor.SetInputType(MOL2);
Receptor.SetOutputType(MOL2);
OBFileFormat::ReadMolecule(RECifs, Receptor, title);
OBAtom *RECatom;

NRGfilename = LIGfile + NRGsuffix + FileExtension;

cout <<"Filename: " << NRGfilename <<endl;

char* NRGfil = new char[NRGfilename.length()+1];
NRGfilename.copy(NRGfil,string::npos);
NRGfil[NRGfilename.length()] = 0;

NRGifs.open(NRGfil);

if(!NRGifs)
{
    cout << "SCORING: Error while opening " << NRGfil <<".\n"; // may give alternative
name??
    return 0;
}
else
{
    OBMolVector DockResult;
    DockResult.OBMolVector::Read(NRGifs, MOL2, MOL2, -1);

    cout << "There are " << DockResult.GetSize() << " molecules in the file " <<
NRGfilename << endl;

    OBMol Pose;

```

```

cout << "OBJECT CONTROL. \tSize of Receptor: " << sizeof Receptor << endl << "\t\t\tSize
of MolVector: " << sizeof DockResult << endl;

for (unsigned int j = 0; j <= DockResult.GetSize()-1; j++)
{
    Pose = *DockResult.OBMolVector::GetMol(j);
    OBAtom *LIGatom;

    double dist;
    Histogram Temp;

    for (unsigned int l = 0; l < BindingSite.size(); l++)
    {
        RECAtom = Receptor.GetAtom(BindingSite.at(l));
        RECx = RECAtom->GetX();
        RECy = RECAtom->GetY();
        RECz = RECAtom->GetZ();

        for (unsigned int k = 0; k < Epitope.size(); k++ )
        {
            LIGatom = Pose.GetAtom(Epitope.at(k));
            LIGx = LIGatom->GetX();
            LIGy = LIGatom->GetY();
            LIGz = LIGatom->GetZ();

            dist = Distance(LIGx, LIGy, LIGz, RECx, RECy, RECz);
            cout << "Distance is :" << dist << "\tCycle j - k: RECAtom - LIGatom: " << j
            << " - " << k << " : : " << BindingSite.at(l) << " - " << Epitope.at(k) << endl;
            if(dist <= MaxDist)
            {
                if(dist <= MinDist)
                {dist = MinDist;}
                char Style = ResultStyle.at(0);
                switch(Style)
                {
                    case 'c':
                        TEST++;
                        StoreHistogram(*LIGatom, dist, Cluster);
                        break;

                    case 'i':
                        TEST++;
                        StoreHistogram(*LIGatom, dist, Cluster);
                        break;

                    case 'a':
                        TEST++;
                        StoreHistogram(*LIGatom, dist, Cluster);
                        StoreHistogram(*LIGatom, dist, Temp);
                        break;
                }
            }
        }
    }
}

if (ResultStyle.at(0) == 'i')
{
    AllPoses.push_back(Cluster);
    Cluster.clear();
}

if (ResultStyle.at(0) == 'a')
{
    AllResults.push_back(Temp);
    Temp.clear();
}

}

if (ResultStyle == "a")
{
    AllResults.insert(AllResults.begin(), Cluster);
}

```

```

    }
    cout << "The programm could identify " << TEST << " contacts less than 6A." << endl;

    if(OutFile.size() > 0)
    {
        char* OUTfil = new char[OutFile.length()+1];    // need extra char for '\0'-char
                                                       // character arrays needed for the
io-stream methods (dont
                                                       // work with strings)

        OutFile.copy(OUTfil, string::npos);
        OUTfil[OutFile.length()] = 0;

        ofstream out(OUTfil, ios_base::out);

        out << "These results were generated by the following settings:(alpha num ordered)"
<< endl;
        map<string, string>::iterator key;
        for (key = Config.begin(); key !=Config.end(); key++)
        {
            out << key->first << "\t" << key->second << endl;
        }
        out << endl << endl;

        char Style = ResultStyle.at(0);
        switch (Style)
        {
            case 'c':
                out << "This file contains the clustered result for" << Config[LIGfile] << "
only. " << endl;
                WriteResults(Cluster, out);
                break;
            case 'i':
                out << "This file contains all individaul histograms, but no clustered histogram"
<< endl;
                WriteResults(AllPoses, out);
                break;
            case 'a':
                out << "This file contains all individual histograms and the cluster result as
the first histogram" << endl;
                WriteResults(AllResults, out);
                break;
        }
    }
    else
    {
        char Style = ResultStyle.at(0);
        if (Style == 'c')
        {
            cout << "Results: " << endl;
            map<long, double>::iterator it;
            for (it = Cluster.begin(); it!=Cluster.end(); it++)
            {
                cout << it->first << "\t" << it->second << endl;
            }
        }
        if(Style == 'i')
        {
            cout << "Results: " << endl;
            for (unsigned int i = 0; i <= AllPoses.size(); i++)
            {
                Histogram temp = AllPoses.at(i);
                map<long, double>::iterator it;
                for (it = temp.begin(); it!= temp.end(); it++)
                {
                    cout << it->first << "\t" << it->second << endl;
                }
            }
        }
        if (Style == 'a')
        {

```

```

cout << "Results: " << endl;
for (unsigned int i = 0; i <= AllResults.size(); i++)
{
    Histogram temp = AllResults.at(i);
    map<long, double>::iterator it;
    for (it = temp.begin(); it!= temp.end(); it++)
    {
        cout << it->first << "\t" << it->second << endl;
    }
}
}
}
}
return 0;
}

```

The following script was employed to dock multiple ligands to a given receptor. Depending on placing of comment lines (#) the random number used is the same for all ligands (line 4 active) or determined for each ligand independently (line15 active). In the given setup all ligands are docked with the same seed for the random number generator. The file filelist.txt contains the names of the ligand files to be used.

```

#!/bin/tcsh
#preliminary preparations

set rnd = `awk -f /usr/user11/jwester/scripts/rnd.awk`
rm -f filelist.temp
setenv workdir `pwd`
setenv moldir /usr/user11/jwester/carb-anchor-repository
setenv molllib `cat filelist.txt | wc -l`
echo "Number of molecules to be screened: " $molllib

cd $workdir
foreach i (`cat filelist.txt`)
    # set rnd = `awk -f /usr/user11/jwester/scripts/rnd.awk`
    echo $rnd
    setenv name `echo $i | sed s/.mol2//`
    echo "Name ist: "$name
    setenv infile dock_anchor"$name".in
    echo "Infile Ist: "$infile
    setenv outfile dock_anchor"$name".out
    echo "Outfile ist :"$outfile
    cat dock_anchor.in.temp1 | sed s/%RND/$rnd/ | sed s/%LIGAND/"$i"/ | sed
s/%RESULT/"$name"anchor_nrg.mol2/ >> $infile
    dock -i $infile -p -o $outfile
end

```

The following script repeatedly docks a single given molecule to a receptor with different random seeds for the random number generator. The ligand file is given as first argument when invoking the script. The second argument is the number of docking experiments (e.g. random numbers) to be tried.

```
#!/bin/tcsh

if (! $#) then
  echo "Usage: rndscreen molecule.mol2 No-of-rounds "
  exit
endif

#preliminary preparations

setenv workdir `pwd`
setenv moldir /usr/user11/jwester/carb-anchor-repository
setenv ligandfile $1
set i=1

while ( $i <= $2)
  set rnd = `awk -f /usr/user11/jwester/scripts/rnd.awk`
  echo $rnd
  setenv name `echo $ligandfile | sed s/.mol2//`
  echo "Name ist: "$name
  setenv infile dock_rndscreen2-"$i"_"$name".in
  echo "Infile Ist: "$infile
  setenv outfile dock_rndscreen2-"$i"_"$name".out
  echo "Outfile ist :"$outfile
  cat dock-rndscreen.in.templ | sed s/%RND/$rnd/ | sed s/%LIGAND/"$1"/ | sed
s/%RESULT/"$name"_rndscr2-"$i"_nrg2.mol2/ >> $infile
  dock -i $infile -p -o $outfile
  set i=`expr $i + 1`
end
```

The following script invokes the distance analysis tool (distancemapper) for a given list of molecules (files.txt).

```
#!/bin/bash
##
## serial analysis of several docking results
##
for datei in `cat files.txt `;
do
  name=`basename $datei .mol2`;
  configfile="$name".conf;
  logfile="$name".log;
  cat config.template | sed s/%LIGAND/$name/ | sed s/%OUT/"$name".dat/ >> $configfile;
  # time -p distancemapper $configfile | cat >> $logfile ;
  distancemapper $configfile >> $logfile;
  times >> $logfile;
  echo $name "has been finished.";
done
```

8.3 Template files and configuration files

The following file demonstrates the setup for docking with the anchor first approach. The file is read in by the multi-anchor-dock script. Values beginning with % are substituted with the according values by the script. The resulting file is taken as input file for the DOCK run.

```
flexible_ligand          yes
orient_ligand           yes
score_ligand            yes
minimize_ligand         yes
multiple_ligands        no
random_seed             %RND
anchor_search           yes
multiple_anchors        no
write_partial_structures no
torsion_drive           yes
clash_overlap           0.5
```

```

configurations_per_cycle      50
torsion_minimize              yes
reminimize_layer_number      2
minimize_anchor                yes
reminimize_anchor              yes
reminimize_ligand              yes
match_receptor_sites          yes
random_search                  yes
ligand_centers                 no
automated_matching             yes
maximum_orientations           2000
write_configurations           yes
write_configuration_total      20
intramolecular_score           yes
intermolecular_score           yes
gridded_score                  yes
grid_version                   4
bump_filter                    yes
bump_maximum                   2
contact_score                  no
chemical_score                 no
energy_score                   yes
energy_cutoff_distance         10
distance_dielectric            yes
dielectric_factor              50
attractive_exponent            6
repulsive_exponent            12
atom_model                    a
vdw_scale                      1
electrostatic_scale            1
energy_minimize                yes
initial_translation            1
initial_rotation               0.1
initial_torsion                10
maximum_iterations             100
energy_convergence             0.1
maximum_cycles                 1
ligand_atom_file               /usr/user11/jwester/chondroitin/%LIGAND
receptor_site_file            gallmcl1.sph
score_grid_prefix              grid
vdw_definition_file            /usr/user11/jwester/Galectin/dock/parameter/vdw.defn
flex_definition_file           /usr/user11/jwester/Galectin/dock/parameter/flex.defn
flex_drive_file                /usr/user11/jwester/Galectin/dock/parameter/flex_drive.tbl
ligand_energy_file             %RESULT

```

The following file demonstrates the setup for flexible docking without anchor first approach. The functionality is according to the aforementioned script.

```

flexible_ligand                yes
orient_ligand                  yes
score_ligand                    yes
minimize_ligand                 yes
multiple_ligands                no
random_seed                     %RND
anchor_search                   no
torsion_drive                   no
torsion_minimize                yes
match_receptor_sites            yes
random_search                    yes
ligand_centers                 no
automated_matching              yes
maximum_orientations            2000
write_orientations              yes
rank_orientations               yes
rank_orientation_total          20
intramolecular_score            yes
intermolecular_score            yes
gridded_score                   yes
grid_version                    4
bump_filter                      yes
bump_maximum                    2

```

```
contact_score          no
chemical_score         no
energy_score           yes
energy_cutoff_distance 10
distance_dielectric    yes
dielectric_factor      4
attractive_exponent    6
repulsive_exponent     12
atom_model             a
vdw_scale              1
electrostatic_scale    1
energy_minimize        yes
initial_translation    1
initial_rotation       0.1
initial_torsion        10
maximum_iterations     100
energy_convergence     0.1
maximum_cycles        1
ligand_atom_file       /usr/user11/jwester/carb-repository/%LIGAND
receptor_site_file     gallmcl1.sph
score_grid_prefix      grid
vdw_definition_file    /usr/user11/jwester/Galectin/dock/parameter/vdw.defn
flex_definition_file   /usr/user11/jwester/Galectin/dock/parameter/flex.defn
ligand_energy_file     %RESULT
```

The following file demonstrates the setup for the grid program. The setup has to be in accordance with the desired setup of the docking experiments.

```
compute_grids          yes
grid_spacing           0.3
output_molecule       no
contact_score          yes
contact_cutoff_distance 4.5
chemical_score         yes
energy_score           yes
energy_cutoff_distance 10
atom_model             a
attractive_exponent    6
repulsive_exponent     12
distance_dielectric    yes
dielectric_factor      50
bump_filter            yes
bump_overlap           0.75
receptor_file          gall.mol2
box_file               gallmcl1box.pdb
vdw_definition_file    /usr/user11/jwester/dock401/parameter/vdw.defn
chemical_definition_file /usr/user11/jwester/dock401/parameter/chem.defn
score_grid_prefix      grid2
```

The following file demonstrates the setup for the distance analysis. The file is read in by the multianalyse script. The script creates a configuration file for the distance analysis (distancemapper) for one ligand and invokes the distancemapper with the respective configuration file. Values beginning with % are substituted with the according values by the script.

```
# basic configuration file for distancemapper (Version 0.3, jcw, Sept.14th 2005)
# Comment lines begin with '#'.
# Empty lines are allowed.
# Format: variable value, PLEASE DO NOT EDIT THE NAME OF VARIABLES!
# For structure files, no .mol2 extension must be given, the suffix
# (default _nrg) can be set in this file (will be expanded to filename_nrg.mol2
# automatically

FileExtension .mol2
LIGfile %LIGAND
NRGsuffix _anchor_nrg
LigSetName LIGSTATIC
RECfile gall-crdprot
RecSetName CRDPROT
SETtype STATIC
OutFile %OUT
# ResultStyle: cluster: cluster only; individual: listing for all poses; all: cluster and
individual,
# with cluster as first result. First char is relevant
ResultStyle a
# MiniDist: minimum distance to be accounted for,
# string value will be converted to float
MiniDist 1.8
# MaxiDist: maximum distance to be accounted for;
# string value will be converted to float
MaxiDist 6
```


Danksagung

Folgenden Personen möchte ich besonders danken:

- Meinen Eltern, für all die Liebe und Unterstützung während meines Studiums
- Heiko & Thomas, für ihre Freundschaft und einen festen Bezugspunkt außerhalb des Studiums
- Boris, für die perfekte Zusammenarbeit an den Computern und so Manchem guten Einfall beim Programmieren
- Winrich, für die reibungslose Übernahme der Computer

Sowie:

Dirk, Katharina, Atilla, Britta, Christian, Kolja, Mirko, Dennis, Svenja, Lars, Thomas H., Ilona, Martin, Robin, Marco, Thomas K., Karolina, Vera und allen weiteren früheren und jetzigen Mitgliedern des Arbeitskreises von Bernd Meyer für eine tolle Zeit und gute Zusammenarbeit. Des Weiteren möchte ich meinen Praktikanten Florian, Elena, Maike, Oliver und Jan-Peter danken.

Hiermit erkläre ich an Eides Statt, dass die vorliegende Dissertationsschrift selbstständig und allein von mir unter den angegebenen Hilfsmitteln angefertigt wurde.

Hamburg, im Juli 2006