Deciphering Carbohydrate Structure
From NMR Chemical Shifts to Conformational Analysis

Kevin M. Dorst
Deciphering Carbohydrate Structure
From NMR Chemical Shifts to Conformational Analysis

Kevin M. Dorst

Abstract
Carbohydrates are ubiquitous in nature and exhibit a multitude of roles. Besides nucleic and amino acids, they can be regarded as the third alphabet of life. They are used as energy source to fuel the cells, as structural building blocks and play a key role in cellular recognition processes. Compared to the other two groups of biomacromolecules, carbohydrates display a higher level of structural complexity by virtue of the number of individual monosaccharide building blocks, as well as the greater number of possibilities of connecting them and additional modifications. This renders a high information content and a good understanding of the structure-function relationship of glycans is important, since the presence or absence of specific structures can make the difference between health and disease.

Carbohydrate structures can be characterized and studied by NMR spectroscopy at the atomic level. This process is time-consuming and error-prone, due to the narrow spectral window, in which most carbohydrate resonances are located leading to severe spectral overlap. Computer programs have been developed, aiding this process. This thesis investigates the quality of prediction of NMR chemical shifts of glycopeptides, highly branched oligosaccharide structures and those bearing a non-natural organic aglycone at the reducing end, as well as the automated determination of primary carbohydrate structures from unassigned NMR spectroscopic data thereof. Novel developments of the CASPER program are highlighted. The three-dimensional structure of carbohydrates plays an important role during carbohydrate-protein interactions. This thesis investigates the conformational preferences and dynamics of glycan structures ranging from di- to tetrasaccharides. A particular focus lies on the measurement of transglycosidic $^3J_{	ext{CH}}$ coupling constants by NMR. Furthermore, the experimental spectroscopic data is compared to results from MD simulations.

Synthetic carbohydrate chemistry has a strong focus on stereoselective C−O bond formation for the synthesis of oligo- and polysaccharides. Each glycosylation reaction can produce two stereoisomeric structures. To date, the mechanistic pathway of glycosylation reactions is still not fully understood, since many different parameters influence the stereoselectivity. A combined experimental and computational study exploring the role of the solvent is presented and a linear correlation of the selectivity with a solvatochromic parameter for the polarizability of the solvent was found.

Keywords: NMR chemical shift prediction, Structural elucidation, Conformational studies, NMR spectroscopy, MD simulations.

Stockholm 2023
http://urn.kb.se/resolve?urn=urn:nbn:se:diva-223154

ISBN 978-91-8014-566-4

Department of Organic Chemistry
Stockholm University, 106 91 Stockholm
Deciphering Carbohydrate Structure
From NMR Chemical Shifts to Conformational Analysis

Kevin M. Dorst
Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. It is our light not our darkness that most frightens us.

- Marianne Williamson
Abstract

Carbohydrates are ubiquitous in nature and exhibit a multitude of roles. Besides nucleic and amino acids, they can be regarded as the third alphabet of life. They are used as energy source to fuel the cells, as structural building blocks and play a key role in cellular recognition processes. Compared to the other two groups of biomacromolecules, carbohydrates display a higher level of structural complexity by virtue of the number of individual monosaccharide building blocks, as well as the greater number of possibilities of connecting them and additional modifications. This renders a high information content and a good understanding of the structure-function relationship of glycans is important, since the presence or absence of specific structures can make the difference between health and disease.

Carbohydrate structures can be characterized and studied by NMR spectroscopy at the atomic level. This process is time-consuming and error-prone, due to the narrow spectral window, in which most carbohydrate resonances are located leading to severe spectral overlap. Computer programs have been developed, aiding this process. This thesis investigates the quality of prediction of NMR chemical shifts of glycopeptides, highly branched oligosaccharide structures and those bearing a non-natural organic aglycone at the reducing end, as well as the automated determination of primary carbohydrate structures from unassigned NMR spectroscopic data thereof. Novel developments of the CASPER program are highlighted.

The three-dimensional structure of carbohydrates plays an important role during carbohydrate-protein interactions. This thesis investigates the conformational preferences and dynamics of glycan structures ranging from di- to tetrasaccharides. A particular focus lies on the determination of transglycosidic $^3J_{\text{CH}}$ coupling constants by NMR. Furthermore, the experimental spectroscopic data is compared to results from MD simulations.

Synthetic carbohydrate chemistry has a strong focus on stereoselective C–O bond formation for the synthesis of oligo- and polysaccharides. Each glycosylation reaction can produce two stereoisomeric structures. To date, the mechanistic pathway of glycosylation reactions is still not fully understood, since many different parameters influence the stereoselectivity. A combined experimental and computational study exploring the role of the solvent is presented and a linear correlation of the selectivity with a solvatochromic parameter for the polarizability of the solvent was found.
Populärvetenskaplig sammanfattning


En metod för bestämning av kolhydratstrukturer på atomär nivå är kärnemagnetisk resonansspektroskopi (NMR). Tolkningen av NMR-spektra av socker är en utmanande uppgift, eftersom ett stort antal karakteristiska signaler finns i ett begränsat område. För att förenkla och automatisera denna uppgift har datorprogram såsom CASPER utvecklats. Vidare utveckling av CASPER med hänsyn till mer kompliceradare strukturer, samt föreningar modifierade vid den reducerande änden har skett och detta illustreras med hjälp av exempel.

Dessutom kan NMR-spektroskopi användas för att studera organiska föringenarnas tre-dimensionella struktur. Denna är av yttersta betydelse för kolhydrater vid interaktionen med proteiner och konformationsstudier av oligosackarider diskuteras.

Kolhydratkemi är specialiserad inom stereoselektiv syntes av C—O bindningar, där det alltid finns möjlighet att skapa två olika föreningar. Olika faktorer, bland annat lösningsmedel, har stort inflytande på resultaten. En kombinerad experimentell och teoretisk studie om inflytandet av lösningsmedlen på selektiviteten av glykosyleringsreaktioner presenteras.


List of Publications and Projects

This thesis is based on the following work, referred to in the text by their Roman numerals.

I. **NMR chemical shift prediction of glycopeptides and glycoproteins aided by the computer program CASPER**
   K. M. Dorst and G. Widmalm
   *submitted*

II. **NMR chemical shift prediction and structural elucidation of linker containing oligo- and polysaccharides using the computer program CASPER**
    K. M. Dorst and G. Widmalm
    DOI: doi.org/10.1016/j.carres.2023.108937

III. **Conformational preferences at the glycosidic linkage of disaccharides in solution as deduced from NMR experiments and MD simulations: comparison to crystal structures**
    K. M. Dorst and G. Widmalm
    *submitted*

IV. **Conformational studies of oligosaccharides related to bacterial polysaccharides by NMR spectroscopy**
    K. M. Dorst, T. Angles d’Ortolli, O. Engström and G. Widmalm
    *Appendix B*
On the Influence of Solvent on the Stereoselectivity of Glycosylation Reactions

Submitted

Related work by the author not included in this thesis:

Peroxy Intermediate Drives Carbon Bond Activation in the Dioxygenase AsqJ
D. Auman, F. Ecker, S. L. Mader, K. M. Dorst, A. Bräuer, G. Widmalm, M. Groll, V. R. I. Kaila
DOI: doi.org/10.1021/jacs.2c05650

Direct N-alkylation of Unprotected Aminosugars with Alcohols. Fast Access to Highly Functionalized Building Blocks and Biodegradable Surfactants
A. Bermejo-Lopéz, B. Saavedra, K. M. Dorst, M. Obieta, P. Maguire, G. Widmalm, B. Martín-Matute
Manuscript
Previous documents based on this work

This thesis is partly-based on the author’s half-time report titled: "Synthesis and Conformational Analysis of Glycans and Stories of CASPER" (presented on December 17, 2021, Stockholm).

The introduction and methods sections (Chapters 1 and 2) have been modified and amended to give a more detailed background. The results discussed in Chapter 3 have in part been previously presented. The chapter has been rewritten and references have been updated. Chapters 4 and 5 were not included in the half-time report and have been written for the purpose of this thesis.
Abbreviations

Abbreviations and acronyms in this work are in agreement with the standards of the American Chemical Society guidelines. Additional non-conventional:

**1DLR** one-dimensional long range

**CASPER** Computer Assisted Spectrum Evaluation of Regular Polysaccharides

**CIP** constant in-plane effect

**CPS** Capsular Polysaccharide

**DFT** Density functional theory

**EPS** Exopolysaccharide

**FF** Force Field

**HK** Hohenberg-Kohn

**HSQMBC** Heteronuclear single quantum multiple bond correlation

**IOS** inner oxygen substituent

**IPAP** in-phase/anti-phase

**ISPA** isolated spin pair approximation

**Kdo** 3-deoxy-d-manno-oct-2-ulosonic acid

**KS** Kohn-Sham

**LPS** Lipopolysaccharide

**MD** Molecular Dynamics

**ML** Machine Learning

**MM** Molecular Mechanics

**NOE** Nuclear Overhauser Effect

**NOESY** Nuclear Overhauser Effect Spectroscopy
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PANIC</td>
<td>peak amplitude normalization for improved cross-relaxation correction</td>
</tr>
<tr>
<td>QM</td>
<td>quantum mechanical</td>
</tr>
<tr>
<td>RMSD</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>rmse</td>
<td>root mean square error</td>
</tr>
<tr>
<td>RU</td>
<td>repeating unit</td>
</tr>
<tr>
<td>sel</td>
<td>selective</td>
</tr>
<tr>
<td>SNFG</td>
<td>symbol nomenclature for glycans</td>
</tr>
<tr>
<td>VIP</td>
<td>variable in-plane effect</td>
</tr>
</tbody>
</table>
# Table of Contents

Abstract ..... i

Populärvetenskaplig sammanfattning ..... ii

Populärwissenschaftliche Zusammenfassung ..... iii

List of Publications and Projects ..... iv

Previous documents based on this work ..... vi

Abbreviations ..... vii

1 Introduction 1
  1.1 Carbohydrates 1
  1.2 Chemical Synthesis of Carbohydrates 4
  1.3 Chemoenzymatic Synthesis 6
  1.4 Bacterial Glycans 6
  1.5 Aims of the Thesis 7

2 Methods 8
  2.1 Nuclear Magnetic Resonance Spectroscopy 8
  2.2 Molecular Modelling 12

3 NMR Chemical Shift Prediction and Structural Elucidation of Glycan Structures (Paper I & II) 16
  3.1 Background 16
  3.2 NMR chemical shift prediction of glycopeptides and glycoproteins aided by the computer program CASPER (Paper I) 18
  3.3 NMR Chemical Shift Prediction and Structural Elucidation of Linker Containing Oligo- and Polysaccharides Using the Computer Program CASPER (Paper II) 24
  3.4 Structural Elucidation of Permethylated Polysaccharides 28
1. Introduction

1.1 Carbohydrates

Besides polynucleotides (DNA and RNA), polypeptides (proteins) and lipids, carbohydrates are the fourth major class of biomacromolecules. They are ubiquitous in nature and are the most abundant of the aforementioned groups. They play critical roles as structural components, energy suppliants, mediators of cell-cell, cell-matrix and cell-molecule interactions.

In contrast to the template driven biosynthesis of proteins from DNA through RNA, the construction of carbohydrate structure solely relies on enzymatic reactions without any template. Carbohydrates – or saccharides (from Greek (sákkharon), meaning "sugar") – are historically referred to as molecules with the sum formula $C_n(H_2O)_m$ and coined by this term as this constitutes “hydrates of carbon”. It was shown later, that carbohydrate structures exist not fulfilling this prerequisite, e.g. 6-deoxy, 2-amino-2-deoxy sugars or uronic acids. The elucidation of the stereochemistry of this polyol bearing either an aldehyde (aldose) or keto group (ketose) dates back to the end of the 19th century, when pioneer Emil Fischer explored the theory of the asymmetric carbon as first described by Le Bel and Van’t Hoff and later was awarded the Nobel Prize in chemistry in 1912.

Fischer’s representation of carbohydrates is nowadays known as the Fischer projection illustrated in Figure 1.1. The molecule is drawn in a straight line with the aldehyde or keto group at the top and specifying the stereochemistry at each stereogenic center. The stereocenter furthest away from the aldehyde or keto group determines the absolute configuration of the molecule, D or L. These linear structures are less stable compared to their cyclic form (hemiacetals, -ketals), as already anticipated by Tollens in 1883. Therefore, different representations are used today. Figure 1.2 illustrates the most commonly used ones, describing the 3-dimensional structure more accurately, namely the Haworth projection and the chair representation. The depiction of larger and more complex carbohydrate structures can be difficult to interpret, when the Haworth projection or chair representation are used. Thus, the symbol nomenclature for glycans (SNFG) was developed. Each sugar has a specific shape and coloring. Moreover, linkages between monosaccharides and the anomic configuration are indicated, comprising an easily understandable representation.

Carbohydrates in their cyclic form can be present as two different stereoisomers, or more precisely epimers or anomers. The equilibrium process of in-
terconversion between different accessible cyclic hemiacetals/-keta\ls with the acyclic form as an intermediate is experimentally observable via change of optical rotation in solution. This phenomenon is called mutarotation. The equili\bration rates and the equilibrium distributions are specific for each sugar and are dependent on the environment, e.g. solvent and temperature (Figure 1.3).

In cyclic systems, especially in cyclohexane-derived structures, heteroatomic substituents in the vicinity of another heteroatom have the tendency to orient themselves in the sterically less favored axial position, instead of the equatorial position. This tendency is known as the Edward-Lemieux effect or anomeric effect. The driving force is the interaction of a filled lone pair orbital ($n_X$) of the heteroatom with the empty antibonding $\sigma^*$ orbital ($\sigma^*_{C-X}$) of a carbon-heteroatom bond, thus lowering the energy of the system (Figure 1.4 (b)). For carbohydrates two different cases can be distinguished: (i) the endo-anomeric effect (Figure 1.4 (a)), where the lone pair orbital of the ring oxygen...
interacts with the antibonding $\sigma^{*}_{C_1-O_1}$ orbital, thereby favoring an $\alpha$ configuration, (ii) in case of an anomeric linkage, a lone pair orbital of the oxygen connecting the two ring systems (O$_n$) can interact with the antibonding $\sigma^{*}_{C_1-O_5}$ orbital, thus having a strong impact on the $\phi$ torsion angle (Figure 1.4 (c)).

Hexoses do not solely occur in their chair conformation as depicted in Figure 1.2. According to IUPAC recommendations\textsuperscript{11} hexoses can exist in 38 canonical or idealized conformations, namely two chair (C), six boat (B), six skew (S), twelve half-chair (H) and twelve envelope (E) conformations. The recommendations suggest assignment of a reference plane constructed by four of the ring atoms. Atoms above and below the plane are indicated by their index number in superscript and subscript, respectively. D sugars for example most often adapt a $^{4}C_1$ conformation. Even though most hexoses adopt a chair-like conformation similar to cyclohexane - as it is the minimum energy conformation - this cannot be generalized. An increase of steric demand can force the 6-membered ring into higher energy conformers.

Monosaccharides can be linked together to form disaccharides and larger oligo- and polysaccharides. The linkage between two residues is called a glycosidic linkage and is formed between the anomeric carbon of one unit and a hydroxyl group of another or in special cases with the anomeric position of
the second residue. Carbohydrates are usually linked by an oxygen atom, but also nitrogen, sulphur and carbon are accessible by chemical methods. Due to the stereoelectronic properties of sugars, disaccharides exhibit specific conformations. These are characterized by the exocyclic torsion angles $\phi$ (H1'-C1'-On-Cn) and $\psi$ (C1'-On-Cn-Hn). The definition of the reference atoms differs between IUPAC,\textsuperscript{12} crystallography and NMR description. During the course of this work only the NMR notation was employed and it is exemplified for a schematic $\beta$-d-pyranosyl-(1$\rightarrow$4)-d-pyranose system (Figure 1.5).

![Illustration of the NMR definition of the glycosidic torsion angles $\phi$ and $\psi$ used in this thesis.](image)

The rotation around the C5-C6 bond is more flexible and an additional exocyclic torsion $\omega$ is defined for (1$\rightarrow$6)-linked saccharides. This torsion is denoted by the torsional angle of O5-C5-C6-O6 and depends on the stereoelectronic properties of the sugar. Three staggered rotamers can be identified - namely gauche-gauche (gg), gauche-trans (gt) and trans-gauche (tg) - based on the orientation of O6 towards O5 and C4 (Figure 1.6).

![Newman projection of the three staggered rotamers of the $\omega$ torsion (O5-C5-C6-O6).](image)

### 1.2 Chemical Synthesis of Carbohydrates

Naturally occurring oligo- and polysaccharides, mostly from pathogenic species, can stimulate an immune response and are thus used as vaccines.\textsuperscript{13} These
structures can be obtained by isolation from the corresponding organisms, e.g. gram-negative bacteria. However, a major drawback is the inhomogeneity of these isolates, as for instance chain length, composition and substitutions are directly influenced by biosynthetic aspects and environmental factors. In order to access well-defined structures, chemical synthesis can be employed.

The reaction of two monosaccharides furnishing a disaccharide is known as a glycosylation reaction. Hereby, a chemically modified monosaccharide bearing a leaving group (LG) at the anomeric position – the donor – is activated by an activator/promoter system, before reacting with a nucleophile – the acceptor. In order to achieve this regioselectively, carbohydrate chemistry strongly relies on the use of protecting groups. Additionally, when forming a glycosidic linkage a new stereogenic center is established, which can be of either α- or β-configuration, as discussed above.

Despite numerous attempts to elucidate the mechanism of chemical glycosylations, the nature of this reaction type is still strongly debated and the stereoselectivity is often hard to predict. Parameters like temperature, solvent, activator/promoter system, concentration, as well as the intrinsic reactivity of the donor – the ability to stabilize a developing positive charge at the anomeric position upon extrusion of the leaving group - and the nucleophilicity of the acceptor can change the selectivity drastically. In classical organic chemistry textbooks, the reader encounters the mechanism described purely as an $S_N^1$-type reaction with an oxocarbenium ion as a key intermediate. The existence of this intermediate in solution under the reaction conditions is highly controversial and was to date just characterized in the gas phase$^{14}$ and in super-acidic medium.$^{15}$ A more complete picture was drawn by Crich, who described the mechanism as a continuum between $S_N^1$ and $S_N^2$ mechanisms, featuring characteristics of both extremes (Figure 1.7).$^{16}$

![Figure 1.7: Reaction mechanism continuum for glycosylation reactions (CIP = contact ion pair, SSIP = solvent separated ion pair). Adapted and reproduced with permission from reference 17.](image)

Chemical synthesis of target structures is a very time-consuming task with
purification of the material after each glycosylation and deprotection step. In the past decade successful attempts have been made to automate the synthesis of oligo- and polysaccharides, in order to facilitate rapid synthesis and purification. The proposed and successfully applied methods range from batch-type synthesis reactors, similar to the ones used for peptide and nucleotide synthesis to solid-phase flow approaches.

1.3 Chemoenzymatic Synthesis

Avoiding the laborious task of protecting group installation and nevertheless having the ability to perform stereo- and regioselective glycosylation reactions is nearly impossible with classical carbohydrate chemistry methods. On the other hand, nature has its own machinery to assemble carbohydrate structures and does this without the help of protecting groups in a highly selective fashion. Therefore, over the last decades glycoscientists started exploiting enzymes to construct desirable target structures. These enzymes are known as glycosyltransferases (GTs), which use nucleotides as donors; the formation of a glycosidic linkage is promoted by phosphate hydrolysis. Even though it was shown in recent years that unnatural substrates can be used to exploit the promiscuity of enzymes, a major drawback of chemoenzymatic synthesis of carbohydrate structures is the limited number of known GTs. Therefore, the introduction of unnatural carbohydrates by this methodology still remains a significant challenge. With the advances in directed evolution during the past decades, this problem can potentially be solved.

Purification of unprotected carbohydrate structures presents a great difficulty. In order to solve this issue, different protocols were developed ranging from automated chemoenzymatic solid-phase synthesis to introduction of a fluorine affinity tag.

1.4 Bacterial Glycans

Bacterial glycans display a greater level of diversity than mammalian and phytoglycans, due to the occurrence of a greater number of different individual monosaccharides as well as different substituents and modifications. The outermost part of the bacterial cell is covered by highly dense glycan layers helping the bacterium to withstand its surroundings. These glycan structures are the major cause of pathogenicity. Based on the staining method by Gram, bacteria can be divided into two distinct categories/families, Gram-positive and Gram-negative. The former have a cell wall with a thick peptidoglycan layer to which different types of polysaccharides such as EPS, CPS and techoic acids
are attached. In contrast, Gram-negative bacteria have a cell wall encapsulated by a thinner peptidoglycan layer, on top of which an outer membrane is covered by a dense forest of lipopolysaccharides with a distinct O-antigen repeating unit (RU) by which the different strains are serotyped.

1.5 Aims of the Thesis

The aims of this thesis are the studies of different aspects of glycans using methods, ranging from synthesis to NMR spectrometry, molecular modelling and computer aided approaches for characterization.

The structural characterization of carbohydrate systems based on NMR spectroscopic methods is a highly time-consuming task, usually carried out by specialists. Further difficulties are met, due to the narrow spectral window in which most carbohydrate associated NMR resonances reside. The automation of this task is a desirable goal. Therefore, further development and application of CASPER, a computer program facilitating the NMR spectral analysis of carbohydrate-based structures, is among others shifted towards highly branched structures and implementation of synthetically relevant aglycones.

Knowledge about the 3-dimensional conformations and dynamics of carbohydrate structures is important as these are the main determinants for glycan-enzyme interactions. A broader understanding of these should be developed from conformational studies with particular emphasis on the measurement of transglycosidic long-range scalar coupling constants as sensitive conformational indicators. Models, which usually stem from MD simulations, are needed to interpret the experimental data. Comparison of the experimental data with the simulation will be beneficial for the further development of force fields.

The selectivity of glycosylation reactions can be affected by multiple factors, some of which have been studied experimentally. A better understanding of the influence of the individual factors will be valuable for synthetic methodology development. The influence of the solvent on the stereoselectivity of glycosylation reactions is studied using a combined experimental and computational approach aiming to decipher its role.
2. Methods

2.1 Nuclear Magnetic Resonance Spectroscopy

2.1.1 General Theory

Since the first observations of nuclear magnetic resonance (NMR) by Rabi and followed by pioneering studies by Bloch and Purcell, NMR spectroscopy evolved as a tool for scientists to investigate molecules at a molecular level in the liquid and solid state, e.g. to elucidate or verify a structure, investigate dynamical processes like conformational changes and interactions between different molecules. The phenomenon relies on the nucleus’ ability to interact with electromagnetic radiation and the observable signal is very much reliant on the chemical surroundings of the nucleus under observation.

Nuclei possess spin quantum number I and are NMR-active, if they do not have both an even number of protons and neutrons. In the case that the sum of neutrons and protons is an odd number I is a half-integer (e.g. I = 1/2, 3/2, and so on). This applies to nuclei such as $^1$H, $^{13}$C, $^{15}$N, $^{19}$F and $^{31}$P. I is an integer when the number of protons and the number of neutrons are both odd (e.g. I = 1, 2, 3 and so on). Examples of nuclei with an integer spin quantum number are $^2$H, $^{10}$B and $^{14}$N. Nuclei, which are not NMR active are termed “silent”. Examples are $^2$He, $^{12}$C and $^{16}$O. These nuclei possess both an even number of neutrons and protons and therefore I = 0. Nuclei with spin quantum number I > 1/2 have an asymmetric charge distribution, giving rise to an electric quadrupole moment besides the magnetic moment and are thus termed “quadrupolar” nuclei. These embody more than two-thirds of all NMR active nuclei. NMR spectroscopy observing these types of nuclei can be difficult as longitudinal relaxation times ($T_1$) are very short, hence resulting in broadened signals, if observable at all. Fortunately for organic and biochemists, most elements found in living organisms, e.g. H, C, N, P, have an isotope with I = 1/2, making them very suitable for NMR spectroscopy with only natural abundance of certain nuclei posing a problem.

The spinning nucleus can be seen as a spinning charge with angular momentum $\vec{P}$ giving rise to a magnetic moment $\vec{\mu}$, which is dependent on the magnetogyric ratio $\gamma$ of the nucleus (Eq. 2.1).

$$\vec{\mu} = \gamma \vec{P}$$  \hspace{1cm} (2.1)

As both the magnetic moment and the angular momentum are vector quantities, they possess a direction and magnitude. There exist 2I+1 quantized spin
states with discrete energies for a nucleus with spin quantum number I. These spin states are degenerate without the influence of a strong external magnetic field. When placed into an external magnetic field the spin states become non-degenerate and split into 2I+1 discrete states, caused by the alignment of the microscopic magnetic moments within the field. For a spin-1/2 nucleus such as $^1\text{H}$ two discrete states, $+1/2$ and $-1/2$, are observable. A popular analogy for these kinds of nuclei is a bar magnet, which can align itself with a magnetic field either in direction of the field ($\alpha$ state) or against it ($\beta$ state) and the latter thus lying higher in energy. The energy difference of the non-degenerate spin states is dependent on the strength of the external magnetic field (Figure 2.1).

![Figure 2.1: Nuclear spin states as a function of the external magnetic field exemplified for a I=1/2 nucleus.](image)

As the static external field interacts with the magnetic moment of a nucleus, a torque is imposed on the magnetic moment, which in turn starts to precess on a circular path around the axis of the magnetic field. The rate of precession, known as the Larmor frequency, is defined as:

$$\omega = -\gamma B_0 \quad \text{or} \quad \nu = -\frac{\gamma B_0}{2\pi} \quad (2.2)$$

with the magnetogyric ratio $\gamma$ and the external magnetic field $B_0$. The phenomenon of NMR can be observed, when spin state transitions happen. This can be achieved by absorption of energy via application of electromagnetic radiation. The frequency of the electromagnetic signal must match the Larmor frequency, in order to satisfy the resonance condition (Eq. 2.3)

$$\Delta E = h\nu = \frac{h\nu B_0}{2\pi} = h\gamma B_0 \quad (2.3)$$

with the Planck constant $h$, the frequency $\nu$ of the electromagnetic radiation, the magnetogyric ratio $\gamma$ and the external magnetic field $B_0$. At equilibrium the
population of the lower lying $\alpha$ spin state is slightly higher compared to the $\beta$ spin state. This can be described by a Boltzmann distribution (Eq. 2.4):

$$\frac{N_\alpha}{N_\beta} = e^{\Delta E/k_B T}$$  \hspace{1cm} (2.4)

with the number of nuclei in a specific spin orientation $N_{\alpha,\beta}$, the Boltzmann constant $k_B$ and the temperature $T$. A reason why NMR spectroscopy is much less sensitive compared to other spectroscopic methods, e.g. IR and UV, is the small energy difference between the different spin states. In turn, also the population difference is rather small, with only 1 in $10^4$. Fortunately, this small difference is enough to be observable. Modest changes in the chemical environment influence the Larmor frequency of the nucleus. This is experimentally observed by a frequency shift of one signal compared to another, giving rise to a distinct chemical shift difference. As the absolute resonance frequencies of nuclei are different depending on the external magnetic field, an arbitrary referencing system with respect to a common standard was introduced. The unit of this reference scale is given as parts per million (ppm) and is defined as follows:

$$\delta = \frac{V_{\text{Signal}} - V_{\text{Standard}}}{V_{\text{Spectrometer}}} \times 10^6$$  \hspace{1cm} (2.5)

### 2.1.2 Scalar Couplings

Different NMR-active nuclei can interact with each other, giving rise to a fine structure, so called scalar couplings, denoted as $J$ and expressed in hertz (Hz). Scalar couplings can be observed for homo- and heteronuclear interactions and thus can give detailed information about the chemical surrounding of the nuclei under observation. The interactions are dependent on stereoelectronic effects. $^3J_{\text{HH}}$ coupling constants can be used to determine ring conformations and anomeric configurations of carbohydrates. The anomeric configuration can be inferred from these homonuclear couplings, e.g. $^3J_{\text{H1-H2}} > 7$ Hz is an indication of an antiperiplanar orientation between the two protons and $^3J_{\text{H1-H2}} < 4$ Hz for a synperiplanar orientation in e.g. glucopyranose. In cases where the homonuclear coupling is not necessarily indicative, e.g. mannopyranose, $^1J_{\text{CH}}$ heteronuclear coupling constants can be employed or used for additional information. $^1J_{\text{CH}} > 169$ Hz is characteristic for an $\alpha$-configuration and $^1J_{\text{CH}} < 169$ Hz for a $\beta$-configuration, respectively. The torsion angles $\phi$ and $\psi$ of a glycosidic linkage between two sugars can be investigated by addressing $^1\text{H},^{13}\text{C}$ long-range couplings between the different ring systems as shown in Figure 1.5. These long-range couplings can be described by a Karplus-type relationship (Eq. 2.6) of the form:
where $\theta$ is the torsional angle and $A$, $B$, $C$ are constants, depending on the system under consideration.

\[ J(\theta) = A \cos^2(\theta) + B \cos(\theta) + C \]  (2.6)

\[ J(\theta) = A \cos^2(\theta) + B \cos(\theta) + C \]

2.1.3 Nuclear Overhauser Effect

Besides scalar couplings, through space interactions, known as dipolar couplings, between two nuclei can be employed to infer the stereochemistry of organic compounds. The magnitude of these couplings can not readily be determined in common NMR solvents, since molecular tumbling will average these to zero. Using a high viscosity alignment medium, thereby restricting the molecular motion, these coupling constants can yield information on the 3-dimensional alignment of the involved spins. Albeit rotational motion renders the acquisition of dipolar coupling constants impossible, the effect of the interaction can be utilized. After selective perturbation of one spin, which is dipolar coupled to a second spin, the relaxation of the latter can be observed indirectly via the nuclear Overhauser effect, giving rise to a change in the intensity of the observable NMR resonance. This change in intensity originates in the fact, that upon perturbation of a spin via electromagnetic radiation, the system will try to return to the equilibrium state via different relaxation pathways as illustrated for a two spin-\(\frac{1}{2}\) system (Figure 2.2), namely the single-quantum ($\Delta M = 1$) transitions $W_{1I}$ and $W_{1S}$, the zero-quantum ($\Delta M = 0$) transition $W_0$ between the $\beta \alpha$ and $\alpha \beta$ states as well as the double-quantum ($\Delta M = 2$) transition between the $\beta \beta$ and $\alpha \alpha$ states.

According to the Boltzmann distribution the population of the different energy levels at equilibrium, considering a homonuclear system, will be approximately $\alpha \alpha \geq \beta \alpha \approx \alpha \beta \geq \beta \beta$. By selective saturation of one of the spins, e.g. spin $S$, thus forcing the population difference across the $S$ transitions ($W_{1S}$) to zero, the system will try to reestablish the equilibrium distribution. On the other hand, the $W_{11}$ transitions will remain at equilibrium rate and will therefore not contribute in restoring equilibrium populations. The cross-relaxation pathways $W_0$ and $W_2$ in turn are responsible for the return to equilibrium populations across the $W_{1S}$ transitions. These two pathways operate in tandem and produce the observable NOE. Depending on the rates of both pathways the NOE can be either positive or negative.\[34\]
The cross-relaxation rate $\sigma_{IS}$ between two spins can be determined by acquisition of NOESY experiments with different mixing times $\tau_{\text{mix}}$ producing a NOE build-up curve. Common ways to extract the cross-relaxation rate are the approach by Dixon et al.\textsuperscript{35} and the PANIC approach,\textsuperscript{36,37} both of which are used in this thesis. For the former the intensities of two spins $I$ and $S$ at a specific mixing time are normalized by division and further divided by the mixing time. This yields a horizontal line, where the intersection with the ordinate is equal to the cross-relaxation rate $\sigma_{IS}$ between spins $I$ and $S$. Another possibility is to average the extracted cross-relaxation rates at different mixing times. The latter relies on correlation of the normalized intensities with the mixing time. The cross-relaxation rate is determined as the slope from a linear regression fit of the normalized intensities as a function of the mixing time.

Using the isolated spin pair approximation\textsuperscript{38} (ISPA, Eq. 2.7), the interatomic distance $r_{IS}$ between spin $I$ and $S$ can be determined from a reference distance $r_{\text{ref}}$, usually originating from molecular modelling or crystal structures, and the corresponding cross-relaxation rates $\sigma_{\text{ref}}$ and $\sigma_{IS}$.

$$r_{IS} = r_{\text{ref}} \left( \frac{\sigma_{\text{ref}}}{\sigma_{IS}} \right)^{\frac{1}{6}}$$  \hspace{1cm} (2.7)

### 2.2 Molecular Modelling

#### 2.2.1 Molecular Mechanics

Molecular mechanics (MM) is a computational modelling approach describing the motions and interactions of atoms and molecules using classical Newtonian mechanics. The potential energy of the system is defined by a force field (FF). The FF consists of a number of parameterized equations. Parameterization
can be derived from empirical data or from quantum mechanical (QM) calculations. The quality and accuracy of a molecular mechanical simulation is dependent on the accuracy of the FF.

### 2.2.2 Force Fields

The force field describes the potential energy of a system and consists of multiple terms describing both internal and external interactions. Force fields can be divided into different categories based on design purpose:

- Classical all-atom force fields
- Coarse-grained force fields
- Polarizable force fields
- Reactive force fields

Classical all-atom force fields treat each atom as a non-elastic sphere, but do not include polarization effects. Coarse-grained force fields sacrifice atomic resolution by mapping several atoms into one interaction site, thereby decreasing the computational cost. Polarizable force fields include electronic polarizability leading to a physically more accurate model, though increasing the computational cost. Furthermore, reactive force fields include bond breaking and formation and therefore allows for the modeling of reactions.

CHARMM36 is a commonly used classical all-atom FF and has the following form for the energy function\(^{39}\) (Eq. 2.8):

\[
\nu(r^N) = \sum_{\text{bonds}} k_i (l_i - l_{i,0})^2 + \sum_{UB} k_{UB} (S_i - S_{i,0})^2 + \sum_{\text{angles}} k_\theta (\theta_i - \theta_{i,0})^2 + \sum_{\text{dihedrals}} k_\omega (\cos(n\omega - \delta) + 1) + \sum_{\text{impropers}} k_{imp} (\phi - \phi_0)^2 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}
\]

where \(k_i, k_{UB}, k_\theta, k_\omega\) and \(k_{imp}\) are the force constants for the bond stretching, Urey-Bradley 1-3 interactions, angles, dihedral angles and improper dihedral angles. \(l_i, S_i, \theta_i, \omega\) and \(\phi\) denote the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, improper dihedral angle, with subscript
zero indicating the equilibrium value. The last term describes non-bonding interactions such as van der Waals interactions as well as electrostatic interaction, according to the Lennard-Jones potential and Coulomb law.

### 2.2.3 Molecular Dynamics

Molecular dynamics (MD) is a computer simulation technique describing the movement and interactions of molecules, ranging from small molecules to complete biological membrane systems and materials over time. Their motion is calculated according to Newton’s law of motion (Eq. 2.9). Solving of equation 2.9 is performed by different algorithms such as Verlet, \(40\) Velocity Verlet \(41\) and Leapfrog algorithm. \(42\)

\[
\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \quad (2.9)
\]

Restraints are generally applied for the simulation of biological systems. The micro-canonical ensemble (NVE) represents a system with particle number \(N\), constant volume \(V\) and constant energy \(E\). Commonly used in simulations are the isothermal-isobaric (NPT) ensemble with constant pressure \(P\) and constant temperature \(T\), as well as the canonical (NVT) ensemble with constant volume \(V\) and constant temperature \(T\). Using MD, atomistic phenomena from pico- to microsecond time scale can be investigated. \(43\) Sampling the atomic motions for a sufficiently long time leads to a representative ensemble of macroscopic properties, which can be described by the average of the microscopic trajectories (Eq. 2.10).

\[
\langle A \rangle = \lim_{t \to \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(\tau) \, d\tau \quad (2.10)
\]

The time step \(\delta t\) is dependent on the fastest motion of the system, corresponding the vibration of bonds to hydrogen atoms. If the time step is larger than the vibrational frequency, inaccuracies to the calculated atomic positions can arise and thus yield erroneous potential energies. Different algorithms such as LINCS \(44\) and SHAKE \(45\) using rigid bond lengths to hydrogen atoms, or hydrogen mass repartitioning \(46\) employing a reassignment of mass from heavier atoms to hydrogen atoms, can be used to lower these vibrational frequencies and in turn permit for larger time steps. Such techniques offer a large decrease in computational time.

### 2.2.4 Density Functional Theory

Commonly used computational methods for the description and characterization of chemical compounds are widely based on quantum mechanics. Fur-
thermore, these methods can be broadly divided into two groups. The wave-function based methods and the density functional theory (DFT) based methods. The latter, instead of describing the system as wave function, utilize the electron density of the system, so that the energy is expressed as a functional of the electron density. DFT is based on the Hohenberg-Kohm (HK) theorems. The first theorem states, that a unique electron density describes the properties of a system in the ground state, thereby simplifying a many-body problem to a density reliant on three spatial coordinates. Whereas the second theorem states, that the variational principle applies to DFT. Therefore, a guessed electron density $\rho_{\text{guess}}$, which is different from the exact electron density $\rho_{\text{exact}}$, represents an upper bound to the exact ground state energy (Eq. 2.11) and thus the true electron density of a given system can be found by minimization of the total energy.

$$E (\rho_{\text{guess}}) \geq E (\rho_{\text{exact}})$$

(2.11)

### 2.2.5 Functional

The energy functional can be divided into several distinct terms (Eq. 2.12). Assuming non-interacting electrons, the energy of a system is given by the kinetic energy of the non-interacting system - known as the KS system - $T_{\text{ni}}$, nuclei-electron attractive interactions $V_{\text{ne}}$, electron-electron repulsion $V_{\text{ee}}$, in form of a coulomb term and the so called exchange-correlation term $E_{\text{XC}}$. The latter consists of a correction for the kinetic energy of an interacting system and a non-classical correction for the repulsive electron-electron interactions. The first three terms are known, whereas technically all functionals differ in the implementation of the exchange-correlation term.

$$E (\rho) = T_{\text{ni}} (\rho) + V_{\text{ne}} (\rho) + V_{\text{ee}} (\rho) + E_{\text{XC}} (\rho)$$

(2.12)

with: $E_{\text{XC}} (\rho) = \Delta T (\rho) + \delta V_{\text{ee}} (\rho)$
3. NMR Chemical Shift Prediction and Structural Elucidation of Glycan Structures (Paper I & II)

3.1 Background

Structural elucidation of polysaccharides is a time-consuming task, usually carried out by specialists. Due to the high structural similarity of the individual monosaccharides, spectral evaluation can become a tedious exercise with severe signal overlap. In order to facilitate and potentially automate this process, computer software has been developed. At first CASPER (Computer Assisted Spectrum Evaluation of Regular Polysaccharides) was able to provide suggestions of possible structures matching the input of unassigned $^{13}$C NMR chemical shifts. Distinguishing similar structures solely based on $^{13}$C NMR chemical shift data is challenging. Therefore, further developments led to the introduction of $^1$H and 2D NMR data as input, increasing the quality of the predictions.

CASPER is working based on incrementations (Figure 3.1). As a base value the NMR chemical shifts of the monosaccharide itself are used (I). For a disaccharide a glycosylation shift is applied, as the position experiencing a glycosylation is exhibiting a downfield displacement of the $^{13}$C NMR chemical shift of ca. 10 ppm with small upfield shifts for the neighbouring positions. The disaccharide correction is calculated by subtracting the chemical shifts of the monosaccharide from the disaccharide subunit (II and III). In the case of vicinally substituted trisaccharides, a trisaccharide correction is added for steric and electronic differences not accounted for by the glycosylation shifts alone. Here the glycosylation corrections for both disaccharide subunits and chemical shifts of the monosaccharide are subtracted from the trisaccharide chemical shifts (IV).

Today, three different modules are incorporated into CASPER. Firstly, the chemical shift prediction tool enables the user to build a certain structure and the program predicts the $^1$H and $^{13}$C NMR chemical shifts. The predicted values are calculated as described before. Additionally, $^1$H and $^{13}$C chemical shifts can be provided as input. Those are assigned to the closest matches of the predicted values. Secondly, the sequence determination tool facilitates the au-
Automated structural elucidation based on unassigned 1D and 2D NMR data and specifications on how many constituting monosaccharides are present. The algorithm will permutate all possible structures based on the input, predict their respective chemical shifts and compare them to the input. The output is the ten most likely structures ranked according to chemical shift similarity between input and prediction. The process can be accelerated by restricting the number of permutations, e.g. by providing linkage information. Lastly, CASPER is able to perform component analysis of hydrolysed and derivatized polysaccharide samples using 1D and 2D NMR data. In recent years model compounds were synthesized and resonance assignments were performed; expanding the database and thus increasing the quality of the chemical shift predictions.52–54

Figure 3.1: Example of how CASPER calculates the predicted NMR chemical shifts of a residue.

In the following sections new features and additions to the CASPER program are highlighted and showcased. Besides the extension of available monosaccharide building blocks, such as Kdo (3-deoxy-D-manno-oct-2-ulosonic acid), this includes the addition of new glycopeptides such as tyrosine glycosides and introduction of non-natural organic spacer molecules, commonly used for immobilization on microarrays for further biological activity testing or conjugation to a protein for vaccine development. The chosen examples demonstrate the abilities of CASPER with respect to the developed extensions.
3.2 NMR chemical shift prediction of glycopeptides and glycoproteins aided by the computer program CASPER (Paper I)

3.2.1 Introduction

Glycosylation is one of the most abundant post-translational modifications of proteins, playing key roles in the proper folding and maturation of glycoproteins. A covalent linkage is established between the anomeric carbon of the carbohydrate moiety and the side chain of an amino acid. Based on the side chain atom different classes of protein glycosylation can be distinguished:

<table>
<thead>
<tr>
<th>Class</th>
<th>Amino acid(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-glycosylation</td>
<td>Serine, Threonine, Tyrosine</td>
</tr>
<tr>
<td>N-glycosylation</td>
<td>Aspargine</td>
</tr>
<tr>
<td>C-glycosylation</td>
<td>Tryptophane</td>
</tr>
</tbody>
</table>

Multiple glycoforms can exist based on the availability of glycosyltransferases during the assembly of the glycan structure, yielding structural heterogeneity. Among the different functions protein glycosylation has are, i.a., aiding proper protein folding and protein stability against proteases, as well as mediation of signal transduction. Research has shown that malfunctioning or altered expression of glycosyltransferases can cause autoimmune diseases, congenital disorder and can be linked to cancer. Nonetheless, glycopeptides can be beneficial and be utilized as antibiotics.

NMR spectroscopy finds use in the field of glycoscience, i.e. for the structural verification of synthesized material, protein-carbohydrate interaction studies and the structural elucidation of isolated unknown structures. Due to the narrow spectral window where most carbohydrate NMR resonances are present, $\delta_H \sim 4-3$ ppm and $\delta_C \sim 70-60$ ppm, severe spectral overlap can complicate the interpretation of NMR spectra and be a source of errors in the resonance assignment process. Therefore, automated approaches using computer programs, such as CASPER, can guide and simplify this time-consuming procedure.

In the following section the versatility, usefulness and the latest implementations of CASPER for the prediction of NMR chemical shifts with regard to glycopeptides and glycoproteins is demonstrated with selected examples.
3.2.2 Results and Discussion

Tentative Resonance Assignment of $\beta$-D-Gal-O-L-Tyr

CASPER is able to approximate the NMR chemical shifts for structures not deposited in the database. This can be illustrated for the case of $\beta$-D-Gal-(1$\rightarrow$3)-$\alpha$-D-Glc-OMe. This disaccharide is currently not present in the database. In order to calculate the glycosylation shift correction, i.e., the differences of chemical shifts between the disaccharide and the two respective monosaccharides a suitable surrogate structure needs to be found in the database. On the other hand, the disaccharide $\beta$-D-Glc-(1$\rightarrow$3)-$\alpha$-D-Glc-OMe is part of the database and since the difference in structure is the stereogenic center in position 4 of the non-reducing end saccharide and is situated further away from the glycosidic linkage the influence on the glycosylation shift will be small.

As part of a recent update to the database of CASPER the assigned $^1$H and $^{13}$C NMR chemical shifts for $\beta$-D-Glc-O-L-Tyr$^1$ were added. In a subsequent study $^7$ $\beta$-D-Gal-O-L-Tyr 1 (Figure 3.2) was synthesized and its biological activity evaluated. Unfortunately, the published NMR data did not contain resonance assignments. Thus, the NMR chemical shifts of the tyrosine galactoside can be approximated from the glucoside analogue, which is handled by CASPER in an automated fashion. The predicted NMR chemical shifts for $\beta$-D-Gal-O-L-Tyr 1 by CASPER are summarized in Table 3.2.

![SNFG representation of $\beta$-D-Gal-O-L-Tyr 1.](attachment:image.png)

Table 3.2: Predicted NMR chemical shift for $\beta$-D-Gal-O-L-Tyr 1.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$7)-Tyr</td>
<td>174.00</td>
<td>56.30</td>
<td>36.20</td>
<td>130.30</td>
<td>131.70</td>
<td>117.90</td>
<td>156.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.06</td>
<td>3.12</td>
<td>3.26</td>
<td>7.28</td>
<td>7.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$-D-Galp-(1$\rightarrow$7)-Tyr</td>
<td>102.53</td>
<td>71.56</td>
<td>73.42</td>
<td>69.28</td>
<td>76.07</td>
<td>61.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.01</td>
<td>3.76</td>
<td>3.70</td>
<td>3.96</td>
<td>3.82</td>
<td>3.67</td>
<td>3.74</td>
<td></td>
</tr>
</tbody>
</table>

Based on the predicted NMR chemical shifts a simulated $^1$H,$^{13}$C-HSQC spectrum can be generated via an in-house python script (Figure 3.3). The observable cross-peaks are in good agreement with the experimentally reported.
Figure 3.3: Simulated $^1$H,$^{13}$C-HSQC spectrum of $\beta$-d-Gal-O-L-Tyr 1. The NMR chemical shifts correspond to the predicted ones from CASPER.

Additionally, if experimental NMR chemical shifts are used as input, CASPER will try to assign the resonances to individual atoms. The process of matching predicted and experimental shift as well as assigning the latter to a specific position is done employing the Kuhn-Munkres algorithm\textsuperscript{72} with respect to minimizing the difference between experimental and predicted NMR chemical shift. In this specific case, when supplying the experimental $^{13}$C and the resolved $^1$H NMR chemical shift from the characterization, CASPER provides a tentative assignment of the supplied chemical shift data, which is summarized in Table 3.3.

Table 3.3: Tentatively assigned experimental NMR chemical shifts for $\beta$-d-Gal-O-L-Tyr 1.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$ 7)Tyr</td>
<td>174.76</td>
<td>56.76</td>
<td>36.36</td>
<td>130.36</td>
<td>131.56</td>
<td>117.76</td>
<td>156.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n.d.</td>
<td>3.09</td>
<td>3.25</td>
<td>7.28</td>
<td>7.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$-d-Gal(1$\rightarrow$</td>
<td>101.46</td>
<td>71.36</td>
<td>73.36</td>
<td>69.26</td>
<td>76.16</td>
<td>61.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n.d. not determined.

With the latest update to the CASPER web-interface, it is now possible to include data from $^1$H,$^{13}$C-HSQC experiments in the module 'Predict NMR chemical shifts', ensuring that the respective coupled nuclei pair is assigned as
a pair and not split during the assignment process.

**NMR Chemical Shift Prediction of Glycan Portion of O-Glycosylated Conopeptide CcTx from *Conus consors* Venom**

Compared to tyrosine glycosides, O-glycosylation of serine and threonine is more common. For GalNAc O-glycosides different core structures have been described\(^7\) and are summarized in Table 3.4.

<table>
<thead>
<tr>
<th>Core</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core 1 or T antigen</td>
<td>β-\text{D-Gal-(1→3)}-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
<tr>
<td>Core 2</td>
<td>β-\text{D-Gal-(1→3)}[β-\text{D-GlcNAc(1→6)}]-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
<tr>
<td>Core 3</td>
<td>β-\text{D-GlcNAc-(1→3)}-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
<tr>
<td>Core 4</td>
<td>β-\text{D-GlcNAc-(1→3)}[β-\text{D-GlcNAc(1→6)}]-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
<tr>
<td>Core 5</td>
<td>α-\text{D-GalNAc-(1→3)}-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
<tr>
<td>Core 6</td>
<td>β-\text{D-GlcNAc-(1→6)}-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
<tr>
<td>Core 7</td>
<td>α-\text{D-GalNAc-(1→6)}-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
<tr>
<td>Core 8</td>
<td>α-\text{D-Gal-(1→3)}-α-\text{D-GalNAc(1→O)-Ser/Thr}</td>
</tr>
</tbody>
</table>

Hocking *et al.* describe the structural elucidation of a pentasaccharide moiety (Figure 3.4 top) attached to serine 7 of conopeptide CcTx from *Conus consors* 2.\(^7\) This pentasaccharide features several unique characteristics, such as the occurrence of the rare carbohydrate L-Gal, as well as the non-canonical core structure β-\text{D-Gal-(1→3)}[α-\text{D-GlcNAc(1→6)}]-α-\text{D-GalNAc(1→O)-Ser}, termed core 9 by the authors. Furthermore, all NMR resonances were assigned and thus enabled an atom-to-atom based comparison of experimental and predicted NMR chemical shifts. Suitable approximations need to be found for the L-Gal containing linkages, since no NMR data for L-Gal containing disaccharides is present in the CASPER database. This is readily automated and the software identifies L-Fuc as a surrogate. This presents a valid approximation, since the difference between the two monomers is the deoxygenation at position 6, which is located further away from the glycosidic linkage. Presently, the disaccharides α-\text{L-Fuc-(1→4)-D-Glc-OMe} and α-\text{L-Fuc-(1→2)-β-D-Gal} are used for the calculation of the respective glycosylation shifts. The agreement between experimental and predicted \(^1\)H and \(^13\)C NMR chemical shifts is good, with deviations of \(^13\)C NMR resonances of \(\sim 2 \text{ ppm}\) for the GalNAc residue (Figure 3.4). It should be noted, that the NMR chemical shift prediction using D-Gal instead of L-Gal is showing only small differences in chemical shifts (Figure 3.4 bottom). Automated resonance assignment of the experimental \(^1\)H,\(^13\)C-HSQC data results in a root mean squared error (rmse) of 0.41 and 0.40, rendering the determination of the absolute configuration of the terminal galactose residues solely based on NMR data difficult and is judged as not viable in this particular case.
Figure 3.4: SNFG representation of O-glycosylated conopeptide CcTx 2 from Conus consors venom (top); Comparison of (a) $^1$H and (b) $^{13}$C NMR chemical shifts predicted by CASPER versus assigned from NMR experiments for 2 (middle); Simulated $^1$H,$^{13}$C-HSQC spectra of predicted serine linked pentasaccharide with terminal L-Gal (red) and terminal D-Gal residues (black) (bottom).

NMR Chemical Shift Prediction of Highly-Branched Nonasaccharide from Paramecium bursaria Chlorella Virus 1

In mammals N-linked glycans present a highly conserved core structure, i.e. $[\alpha$-d-Man-(1→6)]-\alpha$-d-Man-(1→3)-\beta$-d-Man-(1→4)-\beta$-d-GlcNAc-(1→4)-\beta$-d-GlcNAc-(1→N)-Asn. In comparison, organisms like Paramecium bursaria
chlorella virus 1 (PBCV-1) evolved a vastly different glycosylation machinery. The N-linked oligosaccharide attached to the major capsid protein of PBCV-1 features an uncommon $\beta$-D-Glc-(1$\rightarrow$N)-Asn linkage, which connects the glycan portion to the protein, as well as a highly branched glycosylation pattern around a central L-Fuc, together with the presence of Rha in both the D- and L-configuration. All these features create a very unique structure. The $^1$H and $^{13}$C NMR chemical shifts of a nonasaccharide N-glycan 3 (Figure 3.5) were assigned in a study by De Castro et al.,75 thus allowing an atom-by-atom based analysis of the data.

![Figure 3.5: SNFG representation of a nonasaccharide N-linked glycan 3 from Paramecium bursaria chlorella virus 1 (PBCV-1) (top) and comparison of experimental and predicted $^1$H and $^{13}$C NMR chemical shifts (bottom).](image)

In this glycan the unusual structural element $\alpha$-D-Rha-(1$\rightarrow$3)-$\alpha$-L-Fuc is found. For this disaccharide sub-unit currently there is no suitable approximation found in the CASPER database. Therefore, glycosylation shifts were derived from $\alpha$-L-Fuc-(1$\rightarrow$3)-$\alpha$-D-Gal, which is almost a mirror image of the former. With this suitable approximation in hand, the experimental and predicted NMR chemical shifts were compared. Generally, the two sets of data showed good agreement with RMSD values of 0.10 and 1.61 ppm for $^1$H and $^{13}$C, respectively. When inspecting the difference between experimental and predicted NMR chemical shift more closely, it becomes apparent, that
with this type of hyper-branching accurate predictions, especially in the area of highest congestion is very difficult to achieve, viz. the data point at $\delta_C$ CASPER $\sim 80$. This signal corresponds to position 2 in the central fucose residue, which presents an unusually low experimental $^{13}$C NMR chemical shift of ca. 72 ppm. This difference of 8 ppm between experiment and prediction originates from the steric congestion around the highly substituted fucose moiety not accounted for by CASPER. Potentially this could also be an indication for the prevalence of the so called $\gamma$-gauche effect and highlights the need for further development of CASPER, in order to predict the NMR chemical shifts of such heavily branched structures more accurately.

### 3.3 NMR Chemical Shift Prediction and Structural Elucidation of Linker Containing Oligo- and Polysaccharides Using the Computer Program CASPER (Paper II)

#### 3.3.1 Introduction

Understanding the structure-function relationship of carbohydrate structures is vital in unraveling their biological role. Since small changes in glycan patterns can induce physiological dysfunctions. Advances in carbohydrate chemistry allow for the synthesis of highly complex structures. Nowadays, the synthesized structures commonly bear a non-natural, organic linker moiety at the reducing end, readily equipped for further functionalization. Immobilization of such glycans onto e.g. micro arrays permits a facile approach to study carbohydrate-protein interactions. This higher degree of complexity results in intricate characterization by NMR spectroscopy, due to limited spectral dispersion. Resonance assignment is a prerequisite for further NMR-based interaction studies. Therefore, software assisted approaches like CASPER will mitigate and alleviate such complications. This work exemplifies the introduction of non-natural, organic linker moieties at the reducing end of oligo- and polysaccharides into CASPER and the use thereof in NMR chemical shift prediction and structural elucidation workflows. This study contains 16 carbohydrate structure shown below (Scheme 3.1).
3.3.2 Results and Discussion

Oligo- and Polysaccharides related to *Clostridium bolteae*

The repeating unit (RU) of the cell-wall polysaccharide from *C. bolteae* consists of a disaccharide with the unusual monosaccharide D-Rha besides D-Man. Compounds 8, 13 and 14 (Scheme 3.1) are methyl and 5-aminopentyl derivatives of the naturally occurring free reducing end polysaccharide. The absolute configuration of the Rha unit has recently been confirmed by NMR.
spectroscopic methods. Intrigued if CASPER could be employed to determine the absolute configuration of the Rha unit of compound 8, a structure determination was performed with the monosaccharide components set to unknown HexOMe, unknown 6-deoxy-Hex and unknown Hex, together with the NMR chemical shifts information from $^1$H, $^{13}$C-HSQC and correlations from the anomeric positions from a $^1$H,$^1$H-COSY spectrum. Additionally all possible linkage permutations were allowed. The top four ranked structures are shown in Table 3.5 with the normalized relative deviations. The first two, as well as the latter two, are the enantiomeric structures to one another. Unsurprisingly, CASPER cannot determine the absolute configurations, since enantiomers will produce identical NMR resonances.

Table 3.5: Results of structural elucidation of compound 8 from CASPER based on unknown monosaccharide units and experimental NMR data.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\delta_{\text{Rel}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $\alpha$-D-Man-(1→4)$\beta$-D-Rha-(1→3)$\alpha$-D-Man-OMe</td>
<td>1.00</td>
</tr>
<tr>
<td>2. $\alpha$-L-Man-(1→4)$\beta$-L-Rha-(1→3)$\alpha$-L-Man-OMe</td>
<td>1.00</td>
</tr>
<tr>
<td>3. $\alpha$-L-Man-(1→4)$\beta$-D-Rha-(1→3)$\alpha$-D-Man-OMe</td>
<td>35.37</td>
</tr>
<tr>
<td>4. $\alpha$-D-Man-(1→4)$\beta$-L-Rha-(1→3)$\alpha$-L-Man-OMe</td>
<td>35.37</td>
</tr>
</tbody>
</table>

On the other hand, as soon as the absolute configuration of one constituting monosaccharides is known, a clear difference is distinguishable as shown in Figure 3.6. The only difference being, that the reducing end residue is defined as D-Manp-OMe. Though the normalized relative ranking is closer between the top three proposed structures, a difference larger than $\sim 20\%$ ($\delta_{\text{Rel}} \geq 1.2$) is required for significant differentiation between structures. Therefore, when the absolute configuration of one monosaccharide is known, CASPER is able to correctly determine the absolute configuration of the remaining residues.

![Elucidation of structure 8](image)

Figure 3.6: Results of structural elucidation of compound 8 from CASPER based on unknown monosaccharide units and D-Manp-OMe.

With the advances in synthetic carbohydrate chemistry are not only more complex glycan structures obtainable, but also larger polymeric structures.
One such example is compound 14, an octadecamer glycan representing a nine RU long polysaccharide from *C. bolteae*. In order to accommodate the NMR chemical shift prediction of such larger structures and enabling tentative assignment of experimental NMR data, the maximum number of residues available for a single job was increased from 12 to 26 residues. The generated CASPER report of the NMR chemical shift prediction of octadecamer 14 is shown in the supporting information of the publication. This should facilitate the use of CASPER for the $^1$H and $^{13}$C NMR chemical shift prediction of a large number of glycan structures.

**Structural elucidation of regioisomers**

During the chemical synthesis of oligosaccharides, protecting group manipulations consume most of the time. Therefore, recent studies have aimed towards using only partly protected or even protecting group free glycosylation protocols. In the latter study, the authors describe the stereoselective glycosylation of unprotected carbohydrates using an oxazoline donor derived from D-GlcNAc. The glycosylations exclusively yielded $\beta$-D-GlcNAc-(1→6)-linked glycosides. When employing $p$-nitrophenyl $\beta$-maltoside as donor the reaction yielded a mixture of two reaction products, which were identified as structures 11 and 12 based on NMR spectroscopic analysis. Structural elucidation of both structures using CASPER based on NMR chemical shift data, especially from $^1$H, $^{13}$C-HMBC experiments and the three monosaccharides D-Glc-OpNP, D-Glc and D-GlcNAc with all possible linkage permutations allowed, except for the restriction of the D-Glc-(1→4)-D-Glc linkage, yielded the anticipated structures as the top-ranked ones, respectively (Figure 3.7).

![Structural elucidation of regioisomers](image)

**Figure 3.7:** Top ranked structures for the structural elucidation of $p$-nitrophenyl containing trisaccharides 11 (left) and 12 (right) from unassigned NMR data using CASPER.

For the branched structure 11 the relative difference towards the second ranked structure was small, whereas for the linear structure 12 a larger relative difference was obtained. Though, for one case the relative difference was
small, taken together the results provided by CASPER are highly valuable and support the structural assignment.

3.4 Structural Elucidation of Permethylated Polysaccharides

3.4.1 Introduction

At times, the characterization of polysaccharides by solution state NMR techniques can be hindered solely because of solubility issues in D$_2$O, a commonly used NMR solvent for carbohydrate samples, with the sample either not properly dissolving or forming a high-viscosity fluid. This behavior can often be observed for samples extracted from plant cell wall components. A possible solution to increasing the solubility of the sample is derivatization. In a study by Ndukwe et al., the authors describe the influence of permethylation of carbohydrate samples on the NMR chemical shifts. Due to the derivatization CDCl$_3$ was employed as solvent. The database of CASPER contains among others several entries for mono- and dimethylated monosaccharides with assigned $^1$H and $^{13}$C NMR chemical shift in D$_2$O. To study the applicability of CASPER for the structural elucidation of permethylated polysaccharides based on unassigned NMR data, the results from aforementioned study were evaluated.

3.4.2 Results and Discussion

Even though the database relies on data determined with D$_2$O as solvent, CASPER can aid in the structure determination of permethylated samples acquired in CDCl$_3$. Table 3.6 shows the results of the top ranked structure for the structural elucidation based on $^1$H, $^{13}$C-HSQC data, an unknown monosaccharide, together with unspecified linkage positions and three methyl substituents, whose positions, too, are not further specified, as input. The first row represents the experimentally assigned NMR chemical shifts, followed by prediction and the assigned NMR chemical shift from CASPER. The predicted values largely agree well with the experimental values, with the largest deviation for $^{13}$C being $\sim$ 2.5 ppm. More importantly, all experimental resonances were correctly assigned to their respective position, except one methyl substituent for the permethylated dextran sample.
Table 3.6: Comparison of experimental and predicted $^1$H and $^{13}$C NMR chemical shifts for permethylated homopolymers, together with the tentative assignment from CASPER.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow$-4)-2,3,6-tri-O-methyl-$\beta$-D-Glc(p-(1\rightarrow a)</td>
<td>103.2</td>
<td>83.5</td>
<td>85.0</td>
<td>77.4</td>
<td>74.8</td>
<td>70.2</td>
<td>60.5</td>
<td>60.3</td>
<td>59.1</td>
<td></td>
</tr>
<tr>
<td>Predicted$^b$</td>
<td>102.73</td>
<td>82.05</td>
<td>83.43</td>
<td>78.42</td>
<td>74.00</td>
<td>69.97</td>
<td>60.80</td>
<td>60.10</td>
<td>59.40</td>
<td></td>
</tr>
<tr>
<td>Assigned$^c$</td>
<td>103.20</td>
<td>83.50</td>
<td>85.00</td>
<td>77.40</td>
<td>74.80</td>
<td>70.20</td>
<td>60.50</td>
<td>60.30</td>
<td>59.10</td>
<td></td>
</tr>
<tr>
<td>$\rightarrow$-4)-2,3,6-tri-O-methyl-$\alpha$-D-Glc(p-(1\rightarrow a)</td>
<td>97.1</td>
<td>82.4</td>
<td>83.2</td>
<td>73.7</td>
<td>70.5</td>
<td>71.0</td>
<td>59.2</td>
<td>60.4</td>
<td>59.5</td>
<td></td>
</tr>
<tr>
<td>Predicted$^b$</td>
<td>96.27</td>
<td>80.75</td>
<td>84.46</td>
<td>74.59</td>
<td>70.47</td>
<td>71.05</td>
<td>58.30</td>
<td>60.36</td>
<td>59.40</td>
<td></td>
</tr>
<tr>
<td>Assigned$^c$</td>
<td>97.1</td>
<td>82.4</td>
<td>83.2</td>
<td>73.7</td>
<td>70.5</td>
<td>71.0</td>
<td>59.2</td>
<td>60.4</td>
<td>59.5</td>
<td></td>
</tr>
<tr>
<td>$\rightarrow$-6)-2,3,4-tri-O-methyl-$\alpha$-D-Glc(p-(1\rightarrow a)</td>
<td>96.6</td>
<td>82.0</td>
<td>83.5</td>
<td>79.8</td>
<td>70.4</td>
<td>66.2</td>
<td>58.4</td>
<td>60.8</td>
<td>60.6</td>
<td></td>
</tr>
<tr>
<td>Predicted$^b$</td>
<td>95.97</td>
<td>80.43</td>
<td>82.34</td>
<td>79.09</td>
<td>69.67</td>
<td>65.70</td>
<td>58.30</td>
<td>60.36</td>
<td>60.92</td>
<td></td>
</tr>
<tr>
<td>Assigned$^c$</td>
<td>96.60</td>
<td>82.00</td>
<td>83.50</td>
<td>79.80</td>
<td>70.40</td>
<td>66.20</td>
<td>58.40</td>
<td>60.60</td>
<td>60.80</td>
<td></td>
</tr>
<tr>
<td>$\rightarrow$-4)-2,3,6-tri-O-methyl-$\beta$-D-Gal(p-(1\rightarrow a)</td>
<td>102.3</td>
<td>81.9</td>
<td>83.6</td>
<td>69.0</td>
<td>72.2</td>
<td>72.3</td>
<td>60.6</td>
<td>58.8</td>
<td>58.1</td>
<td></td>
</tr>
<tr>
<td>Predicted$^b$</td>
<td>104.87</td>
<td>81.71</td>
<td>82.88</td>
<td>77.23</td>
<td>73.75</td>
<td>71.61</td>
<td>61.00</td>
<td>60.10</td>
<td>59.40</td>
<td></td>
</tr>
<tr>
<td>Assigned$^c$</td>
<td>102.30</td>
<td>81.90</td>
<td>83.60</td>
<td>69.00</td>
<td>72.20</td>
<td>72.30</td>
<td>60.60</td>
<td>58.80</td>
<td>58.10</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Data from Ndukwe et al.$^{82}$; $^b$ NMR chemical shifts predicted by CASPER; $^c$ NMR chemical shifts assigned by CASPER based on structural elucidation.
An example of slightly higher complexity is a mixed linkage glucan polysaccharide with the repeating unit structure:\textsuperscript{82}

\[ \rightarrow 4) - \beta - d - \text{Glc} \text{p} - (1 \rightarrow 3) - \beta - d - \text{Glc} \text{p} - (1 \rightarrow 4) - \beta - d - \text{Glc} \text{p} - (1 \rightarrow \]  

CASPER is in this case not able to determine this structure solely based on \( ^1H, ^{13}C \)-HSQC spectral data with three D-GlcP units, nine methyl substituents and unspecified linkage positions as input. The number of possible permutations is simply too large and the NMR chemical shift differences between the individual structures too small. On the other hand, methylation analysis via GC-MS is a common technique in analytical laboratories working with unknown carbohydrate samples. The relative quantities of differently linked monosaccharides can be obtained routinely. Under the assumption that methylation analysis data for this sample is available and suggesting a ratio of 2:1 for (1\( \rightarrow \)4)-linked to (1\( \rightarrow \)3)-linked subunits, CASPER is able to use this information by user assignment of the methyl substituents to specific positions of the individual components before job submission. In this way, the top ranked structure obtained from CASPER is in fact the mixed linkage glucan structure.

3.5 Conclusions and Outlook

The evaluation and characterization of glycans by NMR spectroscopy is a time-consuming and error-prone task, by virtue of the small spectral window most resonances originating from carbohydrates are located. Nevertheless, resonance assignment is instrumental for further NMR spectroscopy based studies involving e.g. carbohydrate-protein interactions in solution. Though CASPER was originally intended for use with regular polysaccharides, constant developments such as discussed in the previous sections lead to a broader usability and application with respect to different glycan classes. Firstly, the application of CASPER in the chemical shift prediction of glycopeptides was illustrated with various examples. The ability of CASPER to approximate the \( ^1H \) and \( ^{13}C \) NMR chemical shifts of an hitherto unassigned structure such as \( \beta - d - \text{Galp-O-1-Tyr} \) based on a related structure highlights the versatility of CASPER without compromising the high quality of predicted values. Such approximations make it possible to predict the NMR chemical shifts of countless glycan structures. Other examples such as the N-glycan 3 from PBCV-1 featuring a sterically highly congested area around a central fucose moiety exemplify that the way in which the chemical shifts are predicted by CASPER is a good approximation. Albeit yielding mostly a good agreement between experimental and predicted values, further development of CASPER is needed to obtain an even higher accuracy for such highly branched structures. A possible solution could be the introduction of another correction term into the prediction
accounting for the deviations, which could be termed a hyper-branching correction.

The utility of CASPER in handling glycans bearing an aglycone at the reducing end was exemplified with different examples. The determination of the absolute configuration of monosaccharide entities of an oligosaccharide related to the cell-wall polysaccharide of *C. bolteae* based on NMR data further shows yet a different application of CASPER. The increasing amount of available monosaccharide constituents for the NMR chemical shift prediction from 12 to 26 allows for the comparison to spectral data of a large number of structures, as illustrated for the octadecasaccharide 14. Furthermore, the structural determination of regioisomeric structures 11 and 12 exemplifies the usefulness of CASPER in the structural evaluation of synthetically obtained structures. The introduction of non-natural, organic aglycones commonly used in the synthesis of oligo- and polysaccharides into the CASPER database will be of great use to the synthetic carbohydrate community for the characterization of chemically or enzymatically synthesized glycans and a possible approach for structural confirmation of obtained samples via the 'Determine Glycan Structure' module based on unassigned 1D and 2D NMR chemical shift data.

Elucidation of glycan structures isolated from natural sources can be hampered by poor dissolution of the sample in the NMR solvent of choice and instead forming a solution of high-viscosity. NMR spectroscopic analysis of such samples can be undertaken after derivatization as e.g. permethylated samples. Though the database of CASPER relies on NMR chemical shift data acquired with D2O as solvent, the prediction and structural elucidation of permethylated polysaccharide samples, whose NMR chemical shifts were assigned in CDCl3, using CASPER resulted in very good agreement and automated verification of the structure including the resonance assignment. The results from methylation analysis by GC-MS of polysaccharide samples can be used as input for CASPER, thus decreasing the number of possible permutations and aiding in the determination of the structure thereof.

A drawback of CASPER is, that for the accurate prediction of NMR chemical shifts of a glycan structure, suitable mono-, di- and trisaccharide data sets may be lacking in the database. In the case that a constituent monosaccharide is not part of the database, CASPER is not able to carry out the prediction. The software is reliant on high quality data sets. These can only be obtained from the synthesis and characterization of carbohydrates. Advances in the field of machine learning translated into the application towards the prediction of NMR properties of organic molecules.83,84 The ML models rely on DFT-based data sets. The calculation of NMR properties, especially NMR chemical shifts, of carbohydrates by DFT can yield very diverge results based on the functional and basis-set used. Developments of a ML model for the prediction
of glycan NMR chemical shifts based on the high quality CASPER database is a future goal and presents an attractive way to circumvent the cumbersome synthesis of new structures to be included in the database.
4. Conformational Studies of Glycan Structures (Paper III & IV)

4.1 Introduction

Carbohydrates, especially oligo- and polysaccharides, play a key role in numerous biological processes such as inflammation, cell-cell communication and bacterial adhesion.\textsuperscript{85–87} Conformational analysis of oligosaccharides employing NMR methods can prove difficult, due to flexibility around the glycosidic linkages. Particularly during carbohydrate-protein interactions oligosaccharides tend to adopt conformations different from minimum energy conformation.\textsuperscript{88,89} The flexibility around the glycosidic linkage can be observed employing NMR experiments. The torsion angles across the glycosidic linkage are denoted as $\phi_H (H_1'-C_1'-\text{On}-C_n)$, $\phi_C (C_2'-C_1'-\text{On}-C_n)$, $\psi_H (\text{C_1'-On}-C_n-H_n)$ and $\psi_C (\text{C_1'-On}-C_n-C_n \pm 1)$ (Figure 4.1).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4_1.png}
\caption{Definition of the transglycosidic torsional angles $\phi$ and $\psi$ of methyl $\beta$-cellobioside.}
\end{figure}

The torsion angle preference of $\phi_H$ is primarily determined by the exo-anomeric effect, as well as steric effects, whereas $\psi_H$ is only influenced by steric effects. Therefore, $\phi_H$ is mostly $-50^\circ$ for $\alpha$-$D$ and $\beta$-$L$, and $+50^\circ$ for $\beta$-$D$ and $\alpha$-$L$ configured sugars (Figure 4.2). Due to the lack of the exo-anomeric effect, a greater flexibility is observed for $\psi_H$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4_2.png}
\caption{The influence of the exo-anomeric effect on the $\phi_H$ torsion exemplified for $\beta$-celllobiose ($\beta$-$D$-Glc$\text{p}(1\rightarrow4)$-$D$-Glc$\text{p}$).}
\end{figure}

A $^3J$ scalar coupling is the spin-spin interactions over three bonds and can be described by the following torsional angle dependent Karplus-type relation-
ship (Eq. 4.1):

\[ J(\theta) = A \cos^2(\theta + \delta) + B \cos(\theta + \delta) + C \]  

(4.1)

where \( A, B \) and \( C \) are empirical constants, \( \theta \) the torsional angle and \( \delta \) a potential phase shift. NMR experiments or quantum chemical modelling of model compounds can be employed for parameterization of the equation. Equations derived from QM are not always sufficiently accurate, since a proper treatment of all electronic effects is not guaranteed. In order to interpret the NMR data as a 3D model the quality of the parameterization is important.

The magnitude of the coupling constant across a specific torsion is influenced by electronegative substituents, which are either permanently aligned (constant in-plane effect) or might align themselves through rotation around a bond (variable in-plane effect). For carbohydrate systems, in the case of involvement of \(^{13}\text{C}\) spins, studies showed that one can distinguish the influence of a terminal oxygen or outer oxygen substituent from an internal one (inner oxygen substituent (IOS)). For \( \phi \)-dependent couplings the ring oxygen is to be considered as an IOS, whereas \( \psi \)-dependent couplings do not experience the impact of an IOS, but solely from an outer oxygen substituent. Since the couplings related to the different torsions are governed by different stereoelectronic effects, different Karplus-type relationships can be formulated for the designated torsions.\(^{90-92}\)

For the glycosidic linkage of the disaccharide \( \alpha-\text{D-Manp-(1} \rightarrow 2)\)-\( \alpha-\text{D-Manp-OMe} \) 20 the Karplus-type relationships were found to be described more accurately by equations denoted as JCX/SU09 (Figure 4.3).\(^ {93}\) In a later study, the introduction of an additional phase-shift \( \Theta \) for the Karplus-type relationship describing \( \phi_H \) yielded better agreement between experimental and calculated transglycosidic coupling constants for a range of structures.\(^ {94}\)

**4.2 Conformational preferences at the glycosidic linkage of disaccharides in solution as deduced from NMR experiments and MD simulations: comparison to crystal structures (Paper III)**

**4.2.1 Introduction**

The three-dimensional structure of glycans is correlated to their primary and secondary structure.\(^ {95-97}\) Their shape and dynamics are directly affected by the stereochemistry, modifications such as deoxygenations and derivatizations of the constituent monosaccharides.\(^ {98,99}\) Better understanding of the influence
of stereochemistry and linkage position on the conformation and dynamics of disaccharides is beneficial for the prediction of conformations of larger glycan structures incorporating such disaccharide motifs.\textsuperscript{100}

Subtle stereochemical differences such as the orientation of the hydroxyl group in position 4 in α-\textit{L}-\textit{Fucp}(1→3)-β-\textit{D}-\textit{Glc}-\textit{p}-OMe 21 and α-\textit{L}-\textit{Fucp}(1→3)-α-\textit{D}-\textit{Galp}-OMe 22 (Figure 4.4) lead to drastic changes in the solid state structure of the two disaccharides. The former crystallizing with glycosidic torsion angles of $\phi = 51^\circ$ and $\psi = 26^\circ$,\textsuperscript{101} whereas the latter, though displaying a similar conformation in $\phi$ with $\phi = 55^\circ$, being a mirror image of the former about a syn-periplanar conformation with $\psi = -24^\circ$.\textsuperscript{102} The O-antigen polysaccharide from \textit{E. coli} O6 contains a structurally similar subunit to α-\textit{D}-\textit{Glc}(1→4)-α-\textit{D}-\textit{Galp}-OMe 23, being the N-acetyl derivative thereof. During conformational studies of the former a disagreement between experimentally observed values and simulation results from MD was found for this glycosidic linkage displaying a bimodal distribution of the $\psi$ torsion angle.\textsuperscript{103} Since both disaccharide structure share the same stereochemistry, the latter was studied as a model. These disaccharide entities are constituents of lipo-, exo- and capsular polysaccharides, as unveiled by a search of the Carbohydrate Structure Database, which retrieved between 70 and 400 entries for each substructure.\textsuperscript{104}

The solution state conformations and dynamics of disaccharides 21-23 were investigated using NMR experiments and MD simulations based on the knowledge of the solid state conformation of the α-(1→3)-linked disaccharides and the discrepancy of the preferred conformation between simulation and experiment in the case of the α-(1→4)-linked substructure.
4.2.2 Results and Discussion

In the conformational study of organic molecules the nuclear Overhauser effect can be exploited by virtue of the spatial proximity of protons in the molecule. In the case of flexible structures such as carbohydrates an ensemble of conformations needs to be considered due to motional averaging.\(^{105,106}\) NOE buildup curves were obtained from 1D \(^1\)H,\(^1\)H-NOESY experiment via selective excitation of the anomeric proton resonance of the terminal pyranoside over a range of mixing times. \(^1\)H,\(^1\)H cross relaxation rates \(\sigma\) were extracted from the data according to two different methods. Firstly, according to Dixon \(\text{et al.},^{35}\) in which the cross relaxation rates at individual mixing times are averaged, and secondly by the peak amplitude normalization for improved cross-relaxation correction (PANIC) approach, in which the cross relaxation rate corresponds to the slope of a linear regression of \(-\frac{I_j}{I_i}\) versus \(\tau_{\text{mix}}\) (Figure 4.5).\(^{36,37}\) The cross relaxation rates across the glycosidic linkage were higher compared to the intraresidue rates between \(H^{1'}\) and \(H^{2'}\) (Table 4.1), indicating a shorter distance and a preferred \(\text{exo-syn}\) conformation. The effective proton-proton distance \(r_{H^{1'},H^{2'}}\) from the MD trajectory was found to be \(~2.4\ \text{Å}\). According to the ISPA,\(^{38}\) the distances \(r_{H^{1'},H^{n}}\) were calculated from this reference distance and the corresponding \(^1\)H,\(^1\)H cross relaxation rates \(\sigma\).

The conformational preferences of the glycosidic torsion angles can further be investigated via homo- and heteronuclear scalar couplings related to the \(\phi\) and \(\psi\) torsions. Transglycosidic \(^3J_{\text{CH}}\) heteronuclear coupling constants were determined using three different NMR experiments, namely IDLR,\(^{107,108}\) IPAP sel-HSQMBC\(^{109,110}\) and J-HMBC\(^{111}\) (Figure 4.6). In this way, at least two of the three experiments yielded coupling constants for the respective coupling pathways. The experimentally determined coupling constants from different experiments were found to be within 0.2 Hz of each other, which is the generally accepted experimental accuracy margin for these types of NMR experiments. The good agreement across different experiments strengthened the confidence in the determined \(^3J_{\text{CH}}\) values (Table 4.2).
Figure 4.5: Plots of $-I_j/I_i$ versus $\tau_{\text{mix}}$ for 1D $^1$H,$^1$H-NOESY NMR experiments at 500 MHz for disaccharides 21 - 23 (a-c), respectively. Transglycosidic proton-proton interactions are indicated by circles and intraresidue interaction by squares. Lines indicate the linear regression fit.

Table 4.1: NMR $^1$H,$^1$H cross-relaxation rates ($\sigma$) x10² in s⁻¹ at 500 MHz from 1D NOESY experiments determined with the method by Dixon et al.³⁵ and the PANIC approach.³⁶,³⁷ Standard deviations for $\sigma$ are given in parentheses and interproton distances in Å from MD simulations for disaccharides 21 – 23.

<table>
<thead>
<tr>
<th>#</th>
<th>Proton pair</th>
<th>$\sigma_{\text{DIXON}}$</th>
<th>$\sigma_{\text{PANIC}}$</th>
<th>$r_{\text{expt,DIXON}}$</th>
<th>$r_{\text{expt,PANIC}}$</th>
<th>$r_{\text{MD}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>H1'-H3</td>
<td>9.43 (0.24)</td>
<td>9.62 (0.15)</td>
<td>2.30 [0.06]</td>
<td>2.34 [0.02]</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>H1'-H2'</td>
<td>7.78 (0.53)</td>
<td>8.61 (0.20)</td>
<td></td>
<td></td>
<td>2.38</td>
</tr>
<tr>
<td>22</td>
<td>H1'-H3</td>
<td>12.8 (0.60)</td>
<td>11.4 (0.46)</td>
<td>2.25 [0.04]</td>
<td>2.29 [0.03]</td>
<td>2.26</td>
</tr>
<tr>
<td></td>
<td>H1'-H2'</td>
<td>9.01 (0.23)</td>
<td>8.97 (0.15)</td>
<td></td>
<td></td>
<td>2.38</td>
</tr>
<tr>
<td>23</td>
<td>H1'-H4</td>
<td>13.3 (0.50)</td>
<td>12.2 (0.57)</td>
<td>2.26 [0.06]</td>
<td>2.28 [0.08]</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>H1'-H2'</td>
<td>9.24 (0.70)</td>
<td>8.87 (0.79)</td>
<td></td>
<td></td>
<td>2.40</td>
</tr>
<tr>
<td></td>
<td>H1'-H6S</td>
<td>3.38 (0.45)</td>
<td></td>
<td>2.84 [0.10]</td>
<td></td>
<td>2.90</td>
</tr>
</tbody>
</table>

*Interproton distances calculated according to $r_{ij} = r_{\text{ref}}(\sigma_{\text{ref}}/\sigma_{ij})^{1/6}$ with propagated error in square brackets. Effective proton-proton distances averaged over MD trajectories according to $r = \langle r^{-6} \rangle^{1/6}$. Reference distance from MD simulation.
Figure 4.6: Figure 3. (a) $^1$H NMR spectral region of H5' (Glc) and H4 (Gal) in $\alpha$-D-GlcP-(1→4)-$\alpha$-D-Galp-OMe 23; (b) 1DLR spectrum after selective excitation of C1' in the glucosyl residue; (c) 1D slices at the $^{13}$C NMR chemical shift of C1' from 2D $^1$H,$^{13}$C IPAP sel-HSQMBC spectra after selective inversion of H4 ($\alpha$ and $\beta$ spectra are blue and red, respectively); (d) selected spectral region projected in the $F_1$-dimension from $^1$H,$^{13}$C J-HMBC NMR spectrum.

Table 4.2: Transglycosidic $^3 J_{CH}$ coupling constants (Hz) of disaccharides 21 – 23 measured by 1D LR, J-HMBC and IPAP sel-HSQMBC experiments and calculated from MD simulations.

<table>
<thead>
<tr>
<th>Method</th>
<th>21 $^3 J_{CH}(\phi)$</th>
<th>22 $^3 J_{CH}(\phi)$</th>
<th>23 $^3 J_{CH}(\phi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D LR</td>
<td>3.6</td>
<td>n.d.$^a$</td>
<td>n.d.$^a$</td>
</tr>
<tr>
<td>J-HMBC</td>
<td>3.7 (0.1)$^b$</td>
<td>3.7 (0.1)$^b$</td>
<td>4.1 (0.1)$^b$</td>
</tr>
<tr>
<td>IPAP-HSQMBC</td>
<td>3.7</td>
<td>3.7</td>
<td>4.3</td>
</tr>
<tr>
<td>MD (JCX/SU09)</td>
<td>4.3</td>
<td>3.6</td>
<td>4.9</td>
</tr>
<tr>
<td>MD (JCX/SU16)</td>
<td>4.7</td>
<td>4.2</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D LR</td>
<td>n.d.$^a$</td>
<td>5.1</td>
<td>5.3</td>
</tr>
<tr>
<td>J-HMBC</td>
<td>4.7 (0.1)$^b$</td>
<td>4.9 (0.1)$^b$</td>
<td>5.3 (0.2)$^b$</td>
</tr>
<tr>
<td>IPAP-HSQMBC</td>
<td>4.8</td>
<td>4.9</td>
<td>5.3</td>
</tr>
<tr>
<td>MD (JCX/SU09)</td>
<td>5.2</td>
<td>5.4</td>
<td>5.0</td>
</tr>
</tbody>
</table>

$^a$ not determined; $^b$ standard deviation calculated from 4 – 6 experiments; $^c$ The Karplus-type relationship has an additional term, i.e., a phase shift $\Theta = +6^\circ$. 

38
Subsequently, MD simulations with explicit water molecules of disaccharides 21 - 23 were carried out as five replicas of 100 ns simulation per structure, yielding in total 500 ns long trajectories. The conformational space populated by all glycans is the exo-syn conformation. Disaccharides 21 and 22 are characterized by a unimodal distribution, whereas 23 showed a bimodal distribution in $\psi$ (Figure 4.7). Furthermore, the average torsional angles and the root-mean-square-deviations (RMSD) thereof were calculated (Table 4.3).

**Figure 4.7:** Population density maps for the glycosidic torsion angles $\phi$ and $\psi$ in 21 - 23 (a – c) from MD simulations color-coded from 10%—90%. The outer boundary corresponds to 90% of the population.

**Table 4.3:** Averages of glycosidic torsion angles and root-mean-square-deviations (RMSD in parenthesis) of the sampled conformational space from the MD simulation.

<table>
<thead>
<tr>
<th>Disaccharide</th>
<th>$\phi$</th>
<th>$\psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>+39° (15)</td>
<td>+8° (30)</td>
</tr>
<tr>
<td>22</td>
<td>+47° (12)</td>
<td>−1° (24)</td>
</tr>
<tr>
<td>23</td>
<td>−29° (15)</td>
<td>+17° (22)</td>
</tr>
</tbody>
</table>

The experimentally determined proton-proton distances of the (1→3)-linked disaccharides agreed exceptionally well with the ones obtained from the MD simulation (Table 4.1). For the (1→4)-linked disaccharide too, a good agreement between experiment and simulation was found, with a difference between the respective distances within 0.1 Å. The MD simulation predicts a bimodal distribution of the $\psi$ torsion angle for disaccharide 23 with a local minimum
around $\psi \sim 8^\circ$. The effective proton-proton distances $H1',H4$ for the full trajectory was found to be 2.18 Å, whereas for the two underlying $\psi$ conformational states the distance was calculated to 2.24 Å for $\psi^- < 8^\circ$ and 2.15 Å for $\psi^+ > 8^\circ$ and thus indicating a better agreement between experimental and simulation value could be obtained, if the $\psi^-$ state would be the higher populated state. Generally, the distances calculated from the cross relaxation obtained from the PANIC approach were found to be $\sim 2\%$ longer compared to the ones obtained from the method by Dixon et al.

Further comparison between experiment and simulation was undertaken in form of the transglycosidic coupling constants. $^3J_{C,H}$ scalar coupling constants were calculated from the MD trajectories using the Karplus-type equations described by Säwén et al. $^93$ (JCX/SU09) and in case of the $\phi$ related torsion angle also with an equation described by Yang et al. $^94$ (JCX/SU16), which incorporates an additional phase shift $\Theta = +6^\circ$. Better agreement was found with the former for the (1→3)-linked disaccharides, though for disaccharide 23 the opposite was observed. The experimentally determined $^3J_{C,H}(\psi)$ values differed slightly between the three structures, which is coherent with the small differences in the $\psi$ distribution from the simulations. Nonetheless, on the one hand in the case of 21 and 22 the values were overestimated by MD, on the other hand for 23 the value was underestimated.

From the MD trajectories the distribution of $\omega$ torsion angles was obtained related to the conformations of the exo-cyclic hydroxymethyl groups and was compared to the respective monosaccharide (Table 4.4). In disaccharides 21 and 22 the distributions followed the same trends as for the corresponding monosaccharides, with $gt > gg > tg$ for the former and $gt > tg > gg$ for the latter. These are in agreement with the anticipated distributions for the $\omega$ distributions based on the gauche effect $^{112}$ and the Hassel-Otar effect. $^{113,114}$ For disaccharide 23 the distribution, especially for the $\omega'$ torsion angle distribution in Glcp, differed considerably from the one of the monosaccharide 25, in that the $gt$ conformation is favored over the $gg$ conformation. In order to gather more information on this distribution NMR experiments were performed. Originally, the NMR chemical shifts of 23 were assigned at 343 K and due to small chemical shift differences of the $^1H$ NMR chemical shifts of the hydroxymethyl protons only approximate chemical shifts were reported. $^{115}$ $^1H$ NMR spectra of compound 3 were acquired at different temperatures ranging from 300 K to 330 K, in the hope to alleviate the degeneracy. The $^1H$ NMR chemical shifts and $^nJ_{HH}$ coupling constants were refined by NMR spin-simulation using the software Cosmic Truth $^{116}$ from spectra acquired at 300 and 330 K. The $^1H$ NMR chemical shifts of the hydroxymethyl protons were assigned based on observations, that $^3J_{H5,H6pro-R} > ^3J_{H5,H6pro-S}$ for methyl glycosides having gluco- or galacto-configuration. $^{117,118}$ The refined $^3J_{H5,H6}$ coupling constants
were then correlated to the corresponding \( \omega \) distributions via Karplus-type relationships developed by Stenutz et al.\textsuperscript{119} Contrary to the MD simulation this revealed, that the major conformer of \( \omega' \) was the \( gg \) conformation and for \( \omega \) the \( gt \) conformation. Furthermore, this temperature study showed conspicuous temperature dependencies of especially the H1' and H5' \(^1\)H NMR chemical shifts (Figure 4.8).

**Table 4.4:** Rotamer distribution (%) of the hydroxymethyl groups from NMR experiments and MD simulations of mono- and disaccharides.

<table>
<thead>
<tr>
<th>Disaccharides</th>
<th>#</th>
<th>Residue</th>
<th>gt (NMR)</th>
<th>gt (MD)</th>
<th>gg (NMR)</th>
<th>gg (MD)</th>
<th>tg (NMR)</th>
<th>tg (MD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha)-L-Fucp-(1(\rightarrow))3-(\beta)-D-Glcp-OMe</td>
<td>21</td>
<td>Glc</td>
<td>56</td>
<td>39</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha)-L-Fucp-(1(\rightarrow))3-(\alpha)-D-Galp-OMe</td>
<td>22</td>
<td>Gal</td>
<td>51</td>
<td>5</td>
<td>44</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha)-D-Glcp-(1(\rightarrow))4-(\alpha)-D-Galp-OMe</td>
<td>23</td>
<td>Glc</td>
<td>32</td>
<td>60</td>
<td>57</td>
<td>36</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Gal</td>
<td>59</td>
<td>41</td>
<td>4</td>
<td>7</td>
<td>37</td>
<td>52</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

**Monosaccharides**

| | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| \(\alpha\)-D-Galp-OMe | 24 | Gal | 51 | 3 | 5 | 29 | 44 | 4 |
| \(\alpha\)-D-Glcp-OMe\textsuperscript{ab} | 25 | Glc | 42 | 42 | 49 | 55 | 9 | 3 |
| \(\beta\)-D-Glcp-OMe\textsuperscript{c} | 26 | Glc | 55 | 65 | 36 | 32 | 9 | 3 |

\textsuperscript{a} NMR data from Kapla et al.\textsuperscript{120}; \textsuperscript{b} MD data from Plazinski et al.\textsuperscript{121}; \textsuperscript{c} NMR and MD data from Ruda et al.\textsuperscript{122}

**Figure 4.8:** Change of \(^1\)H NMR chemical shifts as a function of temperature for the anomeric protons in 23 (a) and for Glc-H5 and Gal-H4 (b). The colors correspond to the SNFG color of the monosaccharide.\textsuperscript{9}

From the MD simulation of 23 several inter-residue hydrogen bonds could be detected. Some of these could be attributed to the two states \( \psi^+ \) and \( \psi^- \) of the bimodal distribution of the \( \psi \) torsion angle (Table 4.5). One hydrogen bond was found to primarily stabilize each of the two states, respectively. The hydrogen bond O5' \( \cdots \) HO6 is present during ca. 50% of the trajectory and is only found in the \( \psi^+ \) sub state. On the other hand, the non-conventional hydrogen bonding interaction O3 \( \cdots \) H5', which appeared in ca. 25% of the trajectory, can solely be connected to the \( \psi^- \) substate (Figure 9 in manuscript). The latter interaction can be an explanation for the observed down-field shift
of approximately 0.5 ppm of H5', when compared to the 1H NMR chemical shift of the corresponding methyl glycoside. The difference can indicate spacial proximity to an oxygen atom. This type of non-conventional hydrogen bonding interaction was estimated to stabilize such elements by $\sim 2$ kcal mol$^{-1}$.

Table 4.5: Hydrogen bonding (%) and lifetime (ps) in $\alpha$-D-Glc-$p$-(1→4)-$\alpha$-D-Gal-$p$-OMe 23 calculated from MD simulations.

<table>
<thead>
<tr>
<th>Acceptor</th>
<th>Hydrogen</th>
<th>Percentage$^a$</th>
<th>Lifetime$^{b,c}$</th>
<th>Conformation$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>O5'</td>
<td>HO3</td>
<td>48.8</td>
<td>22.6 (0.8)</td>
<td>$\psi^+$</td>
</tr>
<tr>
<td>O2'</td>
<td>HO6</td>
<td>14.1</td>
<td>20.4 (2.4)</td>
<td>$\psi^+$</td>
</tr>
<tr>
<td>O6'</td>
<td>HO3</td>
<td>8.1</td>
<td>5.4 (0.1)</td>
<td>$\psi^+$</td>
</tr>
<tr>
<td>O6</td>
<td>HO2'</td>
<td>5.1</td>
<td>36.3 (16.0)</td>
<td>$\psi^+$</td>
</tr>
<tr>
<td>O6</td>
<td>HO2'</td>
<td>2.4</td>
<td>70.1 (26.1)</td>
<td>$\psi^+/gg$</td>
</tr>
<tr>
<td>O6</td>
<td>HO2'</td>
<td>2.6</td>
<td>7.3 (0.6)</td>
<td>$\psi^+/tg$</td>
</tr>
<tr>
<td>O3</td>
<td>HO6'</td>
<td>4.2</td>
<td>6.0 (0.4)</td>
<td>$\psi^-$</td>
</tr>
<tr>
<td>O3</td>
<td>H5'</td>
<td>23.4</td>
<td>14.0 (0.4)</td>
<td>$\psi^-$</td>
</tr>
</tbody>
</table>

$^a$ Occurrence of hydrogen bond during 500 ns of MD simulation.  
$^b$ Curve fitting of each hydrogen bond to a monoexponential decay showed $R^2 > 0.99$.  
$^c$ Standard deviations calculated from the five 100-ns trajectories are given in parenthesis.  
$^d$ The conformational region for the glycosidic torsion angle $\psi$ has been sub-divided into $\psi^- < 8^\circ$ and $\psi^+ > 8^\circ$; gauche-gauche (gg) and trans-gauche (tg) refer to the rotamers of the hydroxymethyl group in the galactose residue.

To further address the conformational distribution of the $\psi$ torsion angle in disaccharide 23, different experimentally determined values such as interproton distances and coupling constants were correlated with computational data. The experimentally obtained H1'-H6$_{pro.S}$ distance (Table 4.1) and the transglycosidic $^3J_{C1',H4}$ coupling constant were analyzed as a function of the $\psi$ conformational state (Figure 4.9). Both indicated a preference for the $\psi^-$ conformation with $\sim 70\%$ and $60\%$, respectively. This is in good agreement with the H1'-H4 distance noted above.
Figure 4.9: (a) Calculated interproton distance $H1', H6_{\text{pro-S}}$ in 23 as a function of the conformational state of the torsion angle $\psi$. The experimentally derived distance of 2.84 Å is indicated by a horizontal line and the width of the shaded area indicates the propagated error ($\pm 0.10$ Å) based on two standard deviations of the cross relaxation rate. The rotamer distributions of the $\omega$ torsion angle from the MD simulations and NMR experiment were used to calculate the distance from the former (red circle) and the latter (black square). (b) Calculated heteronuclear $^3J_{C1', H4}$ scalar coupling constant in 23 as a function of the conformational state of the torsion angle $\psi$. The experimentally determined magnitude from J-HMBC experiments of 5.3 Hz is indicated by a horizontal line and one standard deviation by the shaded area.

When the crystal structures of 21$^{101}$ and 22$^{102}$ are compared a seemingly small difference in stereochemistry at C4 of the reducing end sugar leads to a large shift in the $\psi$ torsion angle preference, whereas the conformational space sampled by MD simulation only differs by ca 10°. A closer examination of the crystal structure of 21 reveals a dense hydrogen bonding network due to co-crystallized water molecules. This might lead to the close resemblance of the solution state conformation. A database search of the of the Protein Data Bank (PDB) performed with GFDB$^{127,128}$ revealed the presence of the substructure of 21 and 23 in several deposited crystal structures. The structures were filtered to only contain structures solved from x-ray diffraction and are based on a resolution of 2.0 Å or better. The data set was visually examined and some entries disregarded due to structural inconsistencies. In total, 43 structures corresponding to 21 and 20 corresponding to 23 were found. The average glycosidic conformations found were $\langle \phi \rangle = 51^\circ \ (7)$ and $\langle \psi \rangle = 25^\circ \ (5)$ for 21 as well as $\langle \phi \rangle = 48^\circ \ (4)$ and $\langle \psi \rangle = 13^\circ \ (2)$ for 23 (RMSD in parenthesis). A comparison of this data set with the conformational space sampled in MD simulations (Figure 4.10) showed, that the solid and solution state structures of
21 are highly similar. On the other hand, the crystal structure of 22 is devoid of water and seems to have a vastly different conformational preference compared to the solution state structure. Furthermore, all crystal structures containing the element \(\alpha-D-Glc-p-(1\rightarrow4)-D-Galp\) were situated around the \(\psi^-\) state, which was also found to be the higher populated conformation according to NMR experiments.

![Figure 4.10](image)

**Figure 4.10**: Comparison of the populated conformational space for the glycosidic torsion angles \(\phi\) and \(\psi\) in 21 – 23 (a – c, respectively) color-coded as in Figure 4.7 and crystal structures. Glycosidic torsion angle conformation of crystal structures for compounds 21 and 22 are indicated by a red cross and a database search using the glycan fragment database (GFDB)\textsuperscript{127,128} of the Protein Data Bank for structures containing \(\alpha-L-Fuc-p-(1\rightarrow3)-D-Glc-p\) (a) and \(\alpha-D-Glc-p-(1\rightarrow4)-D-Galp\) (c) are indicated by black circles.

### 4.3 Conformational studies of oligosaccharides related to bacterial polysaccharides by NMR spectroscopy (Paper IV)

#### 4.3.1 Introduction

A large cost is associated with the clinical treatment of infectious diseases, many caused by drug-resistant microorganisms. Vaccination is a promising and cost-effective way to control and prevent infections from antibiotic-resistant pathogenic bacteria.\textsuperscript{129} Very potent carbohydrate-based vaccines have been developed in the past decades.\textsuperscript{130} In order to rationalize the development of carbohydrate-based vaccines, their 3D conformation and dynamics, especially during interactions with antibodies, need to be studied. Conformational re-
striction or rigidification of small molecule ligands is a common approach in medicinal chemistry, in order to achieve enhanced binding affinity and selectivity towards a designated target protein, as well as improve pharmacokinetic properties. Oligosaccharides are known to be inherently flexible and compared to other modalities such as peptides, the impact of conformational rigidification has been studied to a lesser extent. Commonly, carbohydrate ligands display dissociation constants in the milli- to micromolar range and in order for them to be valuable therapeutics modifications of the epitope towards higher affinity need to be introduced. Due to the flexibility of the glycosidic linkage, introduction of a molecular tether and locking the glycan epitope into its bound state could be an attractive way to increase the binding efficiency of carbohydrate ligands via preorientation and thereby mitigating thermodynamic losses during the binding event due to entropic changes. Higher affinity of carbohydrate based ligands could be achieved due to structure editing and conformational restriction. Other benefits of conformational restriction can be the increased cell-wall permeability and higher resistance to enzymatic degradation, as well as increased catalytic activity of enzymes.

In order to interpret the NMR spectroscopic data, models are needed. In most cases these originate from MD simulations. Highly accurate models can only be obtained from well parameterized FFs. Therefore, constant development and improvement of FFs is needed to increase the accuracy of the simulations. Transglycosidic scalar coupling constants are a sensitive measure of the conformation and dynamics of glycosidic linkages and in turn provide a useful reference point for FF development.

### 4.3.2 Results and Discussion

#### Compound Selection

The Gram-negative bacterium *Shigella flexneri* is a major cause of diarrhea in humans, especially in developing countries and poses a great challenge. Of all known *S. flexneri* serotypes all but serotype 6 share the same common backbone structure:

\[
\rightarrow 2\text{-})\alpha-L\text{-Rhap}-(1\rightarrow 2)\text{-})\alpha-L\text{-Rhap}-(1\rightarrow 3)\text{-})\alpha-L\text{-Rhap}-(1\rightarrow 3)\text{-})\beta-D\text{-GlcNAc}-(1\rightarrow
\]

The disaccharides 27-30 (Figure 4.11) correspond to all of the glycosidic linkages found in the backbone of *S. flexneri* O-antigens. Knowledge of the transglycosidic coupling constants enables the comparison to related structures and derivatives thereof.

In studies by Bundle and coworkers, the synthesis and binding affinity evaluations of different substituted and tethered trisaccharides related to the
O-antigen from *S. flexneri* serotype Y is described. Based on crystal structures of a monoclonal antibody bound to an oligosaccharide, the conformationally restricted structures 32 - 34 were designed to mimic the bound trisaccharide epitope conformation as closely as possible.

The Gram-negative bacterium *Salmonella enteritidis* is a cause for gastroenteritis in humans. Glycans 35 and 36 represent a part and the full RU of the O-antigen from *S. enteritidis*, respectively.\(^{140}\) Previously conducted MD studies of these structures revealed a significant population of a non-exo-anomeric conformation of the \(\alpha\)-d-Manp-(1\(\rightarrow\)4)-\(\alpha\)-L-Rhap linkage related to the \(\phi\) torsion. Furthermore, a small population of an anti-\(\psi\) conformation of the same linkage was identified for the tetrasaccharide. Contrary to the initial notion, that the conformational flexibility of 36 would be lowered with regard to the higher degree of substitution at the Manp residue, the opposite was observed from MD simulations, when compared to trisaccharide 35.

O-antigen polysaccharides from *E. coli* O35\(^{141}\) and *S. arizona* O62\(^{142}\) share a highly similar RU structures with the only difference being the modification at C6 of the branching \(\alpha\)-d-GalpNAcA residue. Identical reactivity was observed in immunochemical analyses of both LPS with rabbit antiserum specific to *S. arizonae* O62 O-antigen, due to the structural similarity. Trisaccharide 37\(^{143}\) is a model compound related to the branching region of both polysaccharides.

This study aims to investigate the effect of conformational restriction, as well as the influence of vicinal disubstitution, on heteronuclear transglycosidic \(^3J_{CH}\) coupling constants.

**NMR Spectroscopic Analysis**

Transglycosidic \(^3J_{CH}\) coupling constants were attempted to be determined by three different NMR experimental techniques - J-HMBC,\(^{111}\) IPAP sel-HSQMBC\(^{109,110}\) and 1DLR\(^{107,108}\) (Appendix B). Employing different NMR experimental techniques to extract the same information strengthens the confidence in the obtained values. The values from J-HMBC experiments were determined as an average of multiple experiments with different \(\kappa\) scaling factors. Furthermore, the standard deviation was calculated from the array of experiments, yielding a quantified value over the accuracy of the method.

In order to compare the influence of the tethering on the coupling constant, the respective transglycosidic coupling constants for the corresponding disaccharide substructures were available from literature sources\(^{144}\) (27 and 28) or were determined (29). For completeness and potential future conformational NMR spectroscopic studies related to *S. flexneri* glycans, the coupling constants for disaccharide 30 were determined as well, so that now all coupling
Figure 4.11: Schematic structures of compounds 27 - 37.
constants of the individual sub-units describing the *S. flexneri* backbone are experimentally determined.

Trisaccharide 31 is a derivative of the established binding epitope. The terminal L-Rhap unit was replaced with L-ManpA to facilitate the tethering via amide coupling. The determined transglycosidic coupling constants showed very little difference in their magnitude compared to the ones of the isolated disaccharide sub-units. The introduction of a β-alanyl tether in glycan 32 induced a significant change in both coupling constants related to the φ torsion. Especially the α-L-Rhap-(1→3)-β-D-GlcNAc glycosidic linkage experienced a deviation of ∼+1 Hz. The magnitude of the ψ related coupling constants on the other hand changed to a lesser extent. Substitution of the hydroxyl group in position 2 of the central L-Rhap unit by a chloride (33) resulted in no change of the coupling constants related to the φ torsions, whereas the coupling constant related to the ψ torsion of the α-L-ManpA-(1→3)-α-L-Rhap glycosidic linkage was closer to the value of the untethered trisaccharide 31. The obtained coupling constants related to the φ torsion for the 2-deoxy derivative 34 were comparable to those of 32 and 33. Generally, a larger impact was observed for $^3J_{CH}(φ)$ as a result of the tethering (Figure 4.12). Moreover, the $^{13}$C NMR chemical shift of C3 of the GlcpNAc residue was shifted by 3–3.5 ppm upfield upon tethering, emphasizing the sensitivity of $^{13}$C NMR chemical shift displacements to changes in the 3-dimensional structure. $^{145}$

![Figure 4.12](image-url)
For an *exo-syn* conformation heteronuclear transglycosidic coupling constants of $^3J_{CH}(\phi) \sim 4 \text{ Hz}$ and $^3J_{CH}(\psi) \sim 5 \text{ Hz}$ are typically found. The determined coupling constants for compounds 35 and 36 were highly similar and agree well to the mentioned values for the (1→2)- and (1→4)-linkages. For compound 37 the obtained values deviated to some extent (± 0.5 Hz) for the $\phi$ related coupling constant of the (1→2)-linkage and the $\psi$ related coupling constant of the (1→3)-linkage, indicating conformational changes, when compared to a constituting disaccharide.

### 4.4 Conclusion and Outlook

The study of disaccharides, with regard to the characterization of their conformational preferences of the glycosidic linkages, is challenging. With the absence of additional residues, that may restrict the conformational space, only the inherent torsional potentials, hydrogen bonds and charge interactions contribute to the conformational dynamics. The transglycosidic proton-proton distances derived from NOE experiments agreed excellently with the results from MD simulations carried out with explicit solvent for disaccharides 21 and 22, whereas for 23 the deviation between experiment and simulation could be traced to the appearance of a bimodal distribution of the $\psi$ torsion. Furthermore, it was shown, that the less populated conformational state according to MD simulation, should be the major one in accordance to NOE experiments. This is further supported by the analysis of $^3J_{CH}$ coupling constants related to the $\psi$ torsion. The MD simulation of the (1→4)-linked disaccharide 23 revealed a complex inter-residue hydrogen bonding pattern, which is dependent on the hydroxymethyl torsion angle $\omega$ and the $\psi$ torsion angle at the glycosidic linkage. A comparison between solution and solid state conformational preferences showed good agreement between simulation and crystal structure for 21 with a glucopyranosyl residue at the reducing end. On the other hand, with a galactopyranosyl residue at the reducing end (22) the structures deviated with respect to the crystal structure devoid of water. In the case of 23 the conformations of the crystal structures resided around the lesser populated conformational state according to MD, which from NMR experiments was determined to be the main conformational state. The complementary aspects of NMR experiments, simulations and crystal structures for the investigation of the flexibility and conformational preferences of glycosidic linkages are reflected by the results.

The conformation and dynamics of oligo- and polysaccharides are of uttermost importance, especially during carbohydrate-protein interactions, with the possibility of the presence of multiple conformations. Hence, heteronuclear transglycosidic coupling constants were acquired for a number oligosaccha-
rides related to polysaccharides from *S. flexneri*, *S. enteritidis*, *E. coli* O35 and *S. arizonae* O62. The influence of structural tethering and vicinal disubstitution were discussed based on the obtained $^{3}J_{CH}$ coupling constants. The conformational restriction of glycans is compared to other classes, such as peptides, less explored. To systematically study the influence of tethering on the conformation and NMR observables like long-range scalar couplings, model compounds 38-40 (Figure 4.13) with varying tether lengths could be employed.

![Schematic structures of model compounds 38-40](image)

**Figure 4.13:** Schematic structures of model compounds 38 - 40 to study the influence of conformational tethering on transglycosidic coupling constants.

To gain a deeper understanding of the conformational preferences and dynamics of tethered and vicinally disubstituted carbohydrates MD simulations could be used to interpret the NMR data. The obtained $^{3}J_{CH}$ coupling constants can furthermore aid in the developments of FFs.
5. On the Influence of Solvent on the Stereoselectivity of Glycosylation Reactions (Paper V)

5.1 Introduction

The coupling reaction of a glycosyl donor and a glycosyl acceptor is known as a glycosylation reaction. Stereocontrol is the most challenging task involved. Different factors can influence the outcome, such as the nature of donor and acceptor, protecting groups, activator/promoter system, solvent, temperature and water content among others. Several of these were investigated experimentally. The reaction mechanism is best described as a continuum with the two extremes, namely $S_{N1}$ and $S_{N2}$ (Figure 5.1).

![Mechanism continuum to describe the stereochemical outcome of glycosylation reactions](image)

**Figure 5.1:** Mechanism continuum to describe the stereochemical outcome of glycosylation reactions (CIP = contact ion pair, SSIP = solvent separated ion pair). Adapted and reproduced with permission from reference 17.

This study aims to investigate the role and influence of the solvent in glycosylation reactions of thioglucopyranosides with non-participating protecting groups in position 2.
5.2 Results and Discussion

The glucosyl donors 41 - 44 (Figure 5.2) employed in this study carry a non-participating protecting group in position 2. The steric and electronic properties were altered by introduction of methyl-, propyl-, benzyl- and 2-hydroxyethyl-substituents and were synthesized from partly protected thiopyranoside 47\textsuperscript{154} (Scheme 5.1) or in the case of 42 according to literature procedures.\textsuperscript{155} The donors were activated with NIS/AgOTf in the presence of external acceptor EtOH (42 - 44) or in the absence thereof (41). The glycosylation reactions were carried out under identical conditions only employing different solvents. The reactions were high-yielding (>80%) and complete conversion was observed within 2 hours. The ratio of α to β product was determined via integration of suitable resonances of \textsuperscript{1}H NMR spectra of the purified mixtures (Table 5.1).

![Figure 5.2: Thioglucopyranosides used during the course of the study.](image)

![Scheme 5.1: Synthesis of thioglycosides 41, 43 and 44 from partly protected thioglycoside 47.](image)

The selectivity of the glycosylation reactions of donors 41-44 could be tuned from higher α-selectivity in solvents such as Et\textsubscript{2}O, 1,4-dioxane or DMF to high β-selectivity using solvents like acetone or MeCN. The observed selectivities were correlated with different quantitative physico-chemical solvent parameters. Among the dielectric constant \(\varepsilon\), Dimroth-Reichard ET scale,\textsuperscript{157} Z-value\textsuperscript{158} and Kamlet-Taft solvent parameters,\textsuperscript{156} the Kamlet-Taft solvent polarizability parameter \(\pi^*\) showed a seemingly linear correlation. The selectivities of the different donors in relationship to the corresponding \(\pi^*\)-value of the
Table 5.1: Summarized results of glycosylation reactions of donors 41 – 44 in different solvents in terms of α-selectivity. The ratio of α- and β-products was determined by integration of 1H NMR spectra.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>π*&lt;sup&gt;a&lt;/sup&gt;</th>
<th>41</th>
<th>42</th>
<th>43</th>
<th>44</th>
</tr>
</thead>
<tbody>
<tr>
<td>Et&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>0.27</td>
<td>80</td>
<td>65</td>
<td>64</td>
<td>63</td>
</tr>
<tr>
<td>1-Chlorobutane</td>
<td>0.39</td>
<td>–</td>
<td>54</td>
<td>54</td>
<td>53</td>
</tr>
<tr>
<td>Toluene</td>
<td>0.54</td>
<td>60</td>
<td>57</td>
<td>52</td>
<td>55</td>
</tr>
<tr>
<td>Dioxane</td>
<td>0.55</td>
<td>70</td>
<td>68</td>
<td>66</td>
<td>63</td>
</tr>
<tr>
<td>Acetone</td>
<td>0.71</td>
<td>68</td>
<td>45</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>MeCN</td>
<td>0.75</td>
<td>42</td>
<td>27</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>DCM</td>
<td>0.82</td>
<td>59</td>
<td>42</td>
<td>40</td>
<td>36</td>
</tr>
</tbody>
</table>

<sup>a</sup> π* values were obtained from reference 156.

Solvent are shown in Figure 5.3. The trendlines indicated by dashed lines proceed almost parallel to one another. The high α-selectivity when employing DMF (magenta diamond in Figure 5.3a) as solvent can be illustrated in analogy to the halide-assisted glycosylation<sup>159</sup> with an equilibrium between anomeric halides, in which the α-anomer is stabilized, leading to an S<sub>N</sub>2-like displacement of the β-halide by the incoming acceptor, thus yielding high levels of α selectivity.

Computational investigations based on density functional theory (DFT) were undertaken to better understand the influence of the solvent in glycosylation reactions of thioglucosides carrying a non-participating protecting group a position 2. A solvent continuum model<sup>160</sup> was employed in all calculations and a broad range of π*-values was studied including Et<sub>2</sub>O, 1,4-dioxane and acetonitrile. For computational efficiency the benzyl ethers at O3, O4 and O6 were replaced by methyl ethers. Firstly, the possibility of a pure S<sub>N</sub>1 mechanism was evaluated. The calculation of the transition state for the dissociation of the leaving group, such as triflate in this study, is not feasible using a solvent continuum model, since this is not sufficient to stabilize the strong electronegative charge of the LG. The anomeric triflate is reformed, even if the triflate group is placed at a distance of over 3 Å. This can be circumvented by introduction of a positively charged ion in the vicinity of the LG.<sup>161</sup> With the transition state energy inaccessible for the dissociation, the relative free energies of the stationary points along the S<sub>N</sub>1 pathway, such as the naked oxocarbenium ion <b>II</b>, were computed in the different solvents (Scheme 5.2 and Table 5.2). When comparing the relative free energies of <b>II</b> obtained in the different solvents, it is evident, that the reaction is not likely to proceed via a S<sub>N</sub>1 mechanism in the case of Et<sub>2</sub>O and 1,4-dioxane, since the computed energies, 21.3 and 44.1 kcal mol<sup>−1</sup>, respectively, are too high to observe full conversion of the
reaction within the experimental time. On the other hand, in the polar solvent acetonitrile the positive charge of the oxocarbenium ion is readily stabilized, and a lower energy than the β-anomeric triflate $I_{\beta}$ is found. Considering these values, a $S_N1$ mechanism can be ruled out for the reaction in dioxane, is highly unlikely in $Et_2O$ and appears to be feasible in acetonitrile.

**Scheme 5.2**: Schematic representation of $S_N1$-type mechanism for reaction of donor 41.

Reaction mechanisms with higher $S_N2$ character were investigated next. For donor 41 four different pathways can be proposed connecting each of the anomeric triflates with both reaction products $III_\alpha$ and $III_\beta$ (Scheme 5.3).
Table 5.2: Calculated relative free energies of intermediates in kcal mol\(^{-1}\) for intramolecular glycosylation reaction proceeding through a S\(_{N1}\) mechanism in different solvents.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>I(\alpha)</th>
<th>I(\beta)</th>
<th>II</th>
<th>III(\alpha)</th>
<th>III(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Et(_2)O</td>
<td>0.0</td>
<td>3.9</td>
<td>21.3</td>
<td>-7.5</td>
<td>-7.6</td>
</tr>
<tr>
<td>1,4-Dioxane</td>
<td>0.0</td>
<td>5.7</td>
<td>44.1</td>
<td>-5.2</td>
<td>-5.1</td>
</tr>
<tr>
<td>MeCN</td>
<td>0.0</td>
<td>3.7</td>
<td>3.1</td>
<td>-8.7</td>
<td>-9.0</td>
</tr>
</tbody>
</table>

The nucleophile can approach either from the same side as the triflate (A and C) or from the opposite side (B and D). The relative free energies of the respective transition states in different solvents are shown in Table 5.3 (Figure 5.4). The free energy difference between the barriers associated with TS A, TS C and TS D were calculated to be close in energy (\(\Delta\Delta G < 3.0\) kcal mol\(^{-1}\)) for the studied solvents. Therefore, a reliable prediction of the stereochemical outcome is difficult. Analysis of the S–O bond lengths of the triflate groups of the optimized transition state structures revealed rather short S–O bond lengths of 1.47-1.50 Å, which is comparable to an isolated triflate anion, whereas for the covalent triflates I\(\alpha\) and I\(\beta\) \(r_{S-O}\) distances were longer with \(\sim 1.60\) Å. This, combined with rather long C–O distances, especially in TS A and TS C, hinted towards an exploded S\(_{N2}\) mechanism proceeding through contact ion pair species.

Scheme 5.3: Proposed mechanistic pathways related to transition states of higher S\(_{N2}\) character.

Santana et al. described the \(\beta\)-anomeric triflate, though being present at much lower concentrations than the \(\alpha\)-anomeric triflate, as the one being responsible for the experimentally observed high \(\alpha\) selectivity.\(^{162}\) Therefore, the interconversion of anomeric triflates IV\(\alpha\) and IV\(\beta\) (Figure 5.5) in different
Table 5.3: Calculated relative transition state free energies in kcal mol$^{-1}$ for intramolecular glycosylation reaction in different solvents

<table>
<thead>
<tr>
<th>Solvent</th>
<th>TS A</th>
<th>TS B</th>
<th>TS C</th>
<th>TS D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Et$_2$O</td>
<td>12.0</td>
<td>18.0</td>
<td>12.2</td>
<td>10.4</td>
</tr>
<tr>
<td>1,4-Dioxane</td>
<td>15.8</td>
<td>23.4</td>
<td>n.d.</td>
<td>13.7</td>
</tr>
<tr>
<td>MeCN</td>
<td>7.8</td>
<td>12.2</td>
<td>9.1</td>
<td>6.4</td>
</tr>
</tbody>
</table>

n.d. not determined

Figure 5.4: Optimized transition states for TS A – TS D in MeCN.

Solvents was investigated by DFT calculations. The relative free energies for the associated transition state were calculated to be on the order of 12 – 15 kcal mol$^{-1}$ (Table 5.4) and are closely comparable in magnitude to the barriers found for the intramolecular glycosylations.

Table 5.4: Calculated free energies of intermediates and transition states in kcal mol$^{-1}$ for interconversion between alpha and beta triflate of permethylated glucosyl triflate in different solvents.

<table>
<thead>
<tr>
<th>Solvent</th>
<th>IV$\alpha$</th>
<th>IV$\beta$</th>
<th>TS E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Et$_2$O</td>
<td>0.0</td>
<td>2.4</td>
<td>14.9</td>
</tr>
<tr>
<td>1,4-Dioxane</td>
<td>0.0</td>
<td>2.4</td>
<td>12.8</td>
</tr>
<tr>
<td>MeCN</td>
<td>0.0</td>
<td>2.5</td>
<td>12.0</td>
</tr>
</tbody>
</table>

56
5.3 Conclusion and Outlook

During the investigation of the influence of the solvent on the selectivity of model glycosylation reactions of thiopyranosides from the glucos-series carrying a non-participating group at O2 a seemingly linear trend was found in correlation to the Kamlet-Taft parameter $\pi^*$. Generally, lower values were found to induce a higher selectivity for the $\alpha$ configured product, whereas for larger values the opposite was observed. To rationalize these observations, DFT studies of this reaction were undertaken. Firstly, the possibility of a pure $S_N1$-type mechanism was studied. This mechanistic pathway seems only feasible in highly polar solvents, such as MeCN. On the other hand, for solvents of lower polarity, e.g., Et$_2$O and 1,4-dioxane the relative free energy was calculated to be too high to deem this pathway viable on the reaction timeline. Furthermore, transition states corresponding to pathways with higher $S_N2$ character for the intramolecular glycosylation of 41 were optimized. Four different possible pathways were discussed and all except TS B were characterized with highly similar relative free energies spanning a range of 2-3 kcal mol$^{-1}$. These four transition states can be described as exploded $S_N2$ transition states proceeding through CIP species, which could be concluded from the short S-O bond lengths of the leaving group. Based on the small differences in relative energy it is difficult to draw any certain conclusions and reliable prediction of the selectivity for a glycosylation reaction in a given solvent proves challenging. In studies by Hosoya et al.,$^{163,164}$ the authors describe the computational study of the equilibrium of different intermediates of glycosidic triflates, such as covalent intermediates CIPs and SSIPs, displaying a variety of different ring conformations. With the inclusion of several explicit solvent molecules i.e., using a mixed solvent model, it was possible to optimize the geometries of
such structures and the relative free energies thereof, as well as of the transition states connecting them. The reported transition state barriers are comparable in energy to the ones found in this study. Further experimental and computational studies are needed to decipher the complexity of energetically accessible structures, ranging from covalent all the way to solvent separated intermediates, present during glycosylation reactions. This will aid in the rational design of new synthetic glycosylation methodology.
6. Concluding Remarks

The work described in this thesis covers several areas of glycosciences, spanning from characterization based on NMR spectroscopy, to conformational and mechanistic studies of carbohydrates.

The characterization of oligo- and polysaccharides by NMR spectroscopy is a time-consuming and error-prone task. The narrow spectral window, in which resonances from carbohydrates reside, further complicates the analysis due to spectral overlap of resonances. Computer aided approaches, like CASPER, offer the opportunity to guide and automate this task. The application of CASPER for the NMR chemical shift prediction of glycopeptides of different complexity was exemplified. The introduction of commonly employed aglycones for conjugation to proteins and manufacturing of glycans into the database of CASPER was highlighted in the structural elucidation, as well as the NMR chemical shift prediction of carbohydrate structures bearing such an aglycone at the reducing end. A broad variety of discussed oligo- and polysaccharide showed the wide applicability of CASPER. Derivatization of polysaccharide samples can be the only option, if the NMR analysis is hampered by virtue of the low dissolution of the sample. The application of CASPER with regard to permethylated polysaccharides was investigated and a good agreement between experiment and prediction observed. The developments discussed in this thesis with regard to highly-branched structures, glycopeptides and non-natural aglycones broaden the capabilities of CASPER and present a useful tool for the glycoscience community.

Knowledge of the conformations and dynamics of oligo- and polysaccharides is instrumental in deciphering the rich information content displayed by glycans. NMR spectroscopy in combination with MD simulations can be employed to study these. The study of three disaccharides, with either subtle differences in stereochemistry leading to vastly different solid state structures or previously observed discrepancies between experiment and simulation in a related structure, was presented. The influence of small changes in stereochemistry close to the glycosidic linkage on the solution state conformations was small. A bimodal distribution of the ψ torsion was observed for an α(1→4)-linked disaccharide. The population of these two states was evaluated by distance determination based on 1H,1H-NOESY experiments and the measurement of 3JCH transglycosidic coupling constants. In a different study the influence of conformational tethering and vicinal disubstitution on 3JCH trans-

glycosidic coupling constants was investigated for oligosaccharides, which are related to bacterial polysaccharides. Conformational restriction via introduction of a molecular tether had a strong effect on the heteronuclear coupling constants indicating a conformational change. The obtained values deviated by up to 1 Hz, when compared to a non-restricted disaccharide. Additional studies are needed to systematically understand the conformational behavior dynamics of tethered carbohydrate structures. Molecular models are essential in understanding these and development of FFs necessary to accurately describe these modifications.

Synthetic carbohydrate chemistry focuses on the stereoselective formation of C−O bonds. A number of factors can influence the selectivity of glycosylation reactions. The influence of the solvent on the stereoselectivity was systematically explored in the glycosylation reactions of model compounds. A seemingly linear relationship of the selectivity as a function of the Kamlet-Taft solvent parameter $\pi^*$ was observed. Solvents with smaller values of $\pi^*$ yielded a higher selectivity for the $\alpha$ configured product and vice versa. The obtained experimental results were attempted to be corroborated by DFT calculations employing a solvent continuum model. Possible reaction pathways were identified, but the results of the calculations were inconclusive due to small differences in transition state barriers. Further investigations on the role of the solvent are necessary. The inclusion of explicit solvent molecules in the calculations seems inevitable, in order to study the entire range of possible reactive intermediates. A deeper knowledge of how different parameters influence the stereoselectivity of glycosylation reactions will be beneficial for the development of new synthetic methodology.
Appendix A - Contribution List

I. Developed the database and software. Evaluated the data. Participated in writing of the manuscript.

II. Planned and performed part of the NMR experiments. Developed the database and software. Evaluated the data. Participated in writing of the manuscript.

III. Planned and performed the NMR experiments and analysis thereof. Carried out and analyzed the MD simulations. Drafted text and participated in writing of the manuscript.

IV. Planned and performed the NMR experiments and analysis thereof.

V. Performed parts of the synthetic experiments. Performed and evaluated the computational part of the study. Wrote the supporting information and the initial manuscript.
Appendix B - Experimental Data

Paper IV

NMR experiments were carried out on a 700 MHz Bruker AVANCE III equipped with a 5 mm TCI Z-gradient cryoprobe. Samples of oligosaccharides 29 – 37 were prepared in 3 or 5 mm outer diameter NMR tubes as solutions in D$_2$O with concentrations between 22-78 mM. $^1$H and $^{13}$C NMR chemical shifts were referenced externally or to internal 3-trimethylsilyl-(2,2,3,3-$^2$H$_4$)-propionate (TSP, $\delta_H$ 0.00) and externally to a 10% solution of 1,4-dioxane in D$_2$O ($\delta_C$), respectively. The experimental data were processed and evaluated using TOPSPIN 4.0.2 and in-house MATLAB scripts. J-HMBC, IPAP-selHSQMBC$^{109,110}$ and 1DLR$^{107,108}$ experiments were used for the determination of transglycosidic heteronuclear coupling constants at temperatures between 295 - 310 K. IPAP-selHSQMBC experiments were performed and evaluated as described by Saurí et al.$^{110}$

J-HMBC experiments employed a three-fold $J$-filter ($J$ = 120, 145, 170 Hz) for the suppression of $^1J_{CH}$. An acquisition time of 2.5 s in the direct dimension, with either 128, 256 or 512 increments and 8 or 16 scans per $t_1$ increment was used. The spectral widths were set to 10 and 60 ppm in the $F_2$ and $F_1$ dimensions, respectively. Long-range couplings appear in the $F_1$ dimension scaled by $\kappa = 17$-$35$ relative to the NMR chemical shift of the heteronucleus. The 2D data were zero-filled to 4k and 2k data points in $F_2$ and $F_1$, respectively. The peak-separation was determined from automatic peak-picking routines as implemented in TOPSPIN and $^3J_{CH}$ coupling constants extracted via division of the peak-separation (in Hz) by the scaling factor $\kappa$.

For the suppression of $^1J_{CH}$, 1DLR experiments used a delay which was set to either 145 or 169 Hz, depending on the selectively irradiated $^{13}$C resonance. The delay for the evolution of long-range couplings was set to 6 Hz. The spectral width was set to 12 ppm with an acquisition time of 4 s and an inter scan delay of 1 s. The FIDs were zero-filled once and multiplied with an exponential window function using a line-broadening factor of 0.25 – 0.60 Hz. The $^3J_{CH}$ transglycosidic coupling constants appear as anti-phase doublets and were extracted by the $J$ doubling methodology$^{165}$ implemented by an in-house MATLAB script.
### Table B: Transglycosidic NMR $^3J_{CH}$ coupling constants (Hz) of oligosaccharides in D$_2$O determined by J-HMBC, IPAP sel-HSQMBC and 1DLR experiments.

<table>
<thead>
<tr>
<th>Oligosaccharide</th>
<th>#</th>
<th>$J(\phi_2)$</th>
<th>$J(\psi_2)$</th>
<th>$J(\phi_3)$</th>
<th>$J(\psi_3)$</th>
<th>$J(\phi_4)$</th>
<th>$J(\psi_4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-L-Rha-(1→2)-$\alpha$-L-Rha-OMe</td>
<td>27</td>
<td>4.3$^{a,d}$</td>
<td>4.7$^{a,d}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-L-Rha-(1→3)-$\alpha$-L-Rha-OMe</td>
<td>28</td>
<td></td>
<td></td>
<td>4.1$^{a,d}$</td>
<td></td>
<td>5.1$^{a,d}$</td>
<td></td>
</tr>
<tr>
<td>$\alpha$-L-Rha-(1→3)$\beta$-D-GlcNAc-OMe</td>
<td>29</td>
<td>4.2 (0.2)$^a$</td>
<td>4.7 (0.2)$^a$</td>
<td>5.2 (0.2)$^a$</td>
<td>5.4 (0.2)$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$-D-GlcNAc-(1→2)-$\alpha$-L-Rha-OMe</td>
<td>30</td>
<td>4.1 (0.1)$^{a,e}$</td>
<td>4.5 (0.2)$^{a,f}$</td>
<td>4.8 (0.1)$^{a,e}$</td>
<td>5.1 (0.2)$^{a,f}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-L-ManANMe-(1→3)-$\alpha$-L-Rha-(1→3)$\beta$-D-GlcNAc-OMe</td>
<td>31</td>
<td>4.1 (0.1)$^{b,e}$</td>
<td>4.5 (0.2)$^{b,f}$</td>
<td>4.8 (0.1)$^{b,e}$</td>
<td>5.1 (0.2)$^{b,f}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-L-ManANMe(CH$_2$)-[1→3]-$\alpha$-L-Rha-(1→3)+$\beta$-D-GlcN(C=O)CH$_2$-OMe</td>
<td>32</td>
<td>3.5 (0.1)$^{b,e}$</td>
<td>5.4 (0.2)$^{b,f}$</td>
<td>5.4 (0.1)$^{b,e}$</td>
<td>5.0$^{b,f}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-L-ManANMe(CH$_2$)-[1→3]-2-chloro-2-deoxy-$\alpha$-L-Rha-(1→3)$\beta$-D-GlcN(C=O)CH$_2$-OMe</td>
<td>33</td>
<td>3.5 (0.1)$^{b,e}$</td>
<td>5.5 (0.2)$^{b,f}$</td>
<td>4.6 (0.2)$^{b,e}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-L-ManANMe(CH$_2$)-[1→3]-2-deoxy-$\alpha$-L-Rha-(1→3)$\beta$-D-GlcN(C=O)CH$_2$-OMe</td>
<td>34</td>
<td>3.6 (0.2)$^{b,e}$</td>
<td>5.3 (0.1)$^{b,f}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-D-Gal-(1→2)$\alpha$-D-Man-(1→4)$\alpha$-L-Rha-OMe</td>
<td>35</td>
<td>4.2 (0.2)$^b$</td>
<td>4.4 (0.2)$^b$</td>
<td>4.4 (0.2)$^b$</td>
<td>4.5$^b$</td>
<td>5.6 (0.2)$^b$</td>
<td>5.2$^b$</td>
</tr>
<tr>
<td>$\alpha$-D-Gal-(1→2)$\alpha$-Tyr-(1→3)$\alpha$-D-Man-(1→4)$\alpha$-L-Rha-OMe</td>
<td>36</td>
<td>3.9 (0.2)$^b$</td>
<td>4.5 (0.3)$^b$</td>
<td>4.5 (0.2)$^b$</td>
<td>4.5 (0.2)$^b$</td>
<td>5.4 (0.2)$^b$</td>
<td>5.2$^b$</td>
</tr>
<tr>
<td>$\beta$-D-GlcNAc-(1→3)$\alpha$-D-GalNAc-(1→2)$\alpha$-L-Rha-OMe</td>
<td>37</td>
<td>3.5$^b$</td>
<td>4.7 (0.1)$^b$</td>
<td>4.1 (0.1)$^b$</td>
<td>4.1 (0.1)$^b$</td>
<td>5.8 (0.2)$^a$</td>
<td>5.5$^b$</td>
</tr>
</tbody>
</table>

$^a$ Scalar couplings from J-HMBC experiments and are averaged from 4-12 experiments with standard deviations given in parentheses. $^b$ Scalar couplings from IPAP sel-HSQMBC experiments. $^c$ Scalar couplings from 1DLR experiments. $^d$ $^3J_{CH}$ data from Hardy et al. $^{144}$; $^e$ terminal (1→3)-linkage, $^f$ reducing (1→3)-linkage.
Appendix C - Experimental Data

**Table C**: $^1$H and $^{13}$C NMR chemical shifts of $\beta$-D-Glc-p-O(CH$_2$)$_5$NH$_2$ in D$_2$O at 25 °C referenced to external TSP ($\delta_H$ 0.00) and external dioxane in D$_2$O ($\delta_C$ 67.40). The $^1$H NMR chemical shifts were refined via NMR spin simulation using Cosmic Truth.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$-D-Glc-p-O(CH$_2$)$_5$NH$_2$</td>
<td>4.469 (8.02)$^a$</td>
<td>3.270</td>
<td>3.494</td>
<td>3.386</td>
<td>3.462</td>
<td>3.727, 3.928</td>
</tr>
<tr>
<td></td>
<td>102.98</td>
<td>73.94</td>
<td>76.63</td>
<td>70.47</td>
<td>76.74</td>
<td>61.56</td>
</tr>
<tr>
<td>$\beta$-D-Glc-p-O(CH$_2$)$_5$NH$_2$</td>
<td>3.701, 3.949</td>
<td>1.681, 1.687</td>
<td>1.469, 1.474</td>
<td>1.700, 1.715</td>
<td>3.013, 3.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70.86</td>
<td>28.97</td>
<td>22.89</td>
<td>27.29</td>
<td>40.17</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ J$_{H1,H2}$ in parenthesis.
Acknowledgements

I wish to express my sincere gratitude to:

Prof. Göran Widmalm, for your guidance, your support, sharing your knowledge and all the discussions during these years.

Prof. Berit Olfsson and Prof. Belén Martín-Matuté for the support, the interest in this thesis and for the valuable feedback.

Dr. Martina Delbianco for accepting to be the opponent for this dissertation, as well as Dr. Ulrika Westerlind, Prof. Corinne Sandström, Prof. Niklas Hedin and Prof. Thomas Norberg for being part of the committee.

The past members of the GW group: Dr. Axel Furevi, Dr. Alessandro Ruda, Zaheer Timol, Dr. Geeta Karki, Dr. Mateusz Pallach, Dr. Federico Riu, as well as the students Honorine Lardeux, Domenica Spanu, Quentin Raveyre and Clement Maynand, I had the pleasure to work with.

The TA staff at the department and beyond: Karin Larsson, Kristina Romare, Martin Roxengren, Jonas Stähle, Matthew Bennett, Ola Andersson, Sigrid Mattson, Lousie Lehto, Gabriella Ågren and Petra Godin. Thank you for your help.

The great people at the department, who I had the pleasure of meeting throughout the years. I will not list all of you, since I would surely forget about someone.

Meine Freunde aus der Heimat und darüber hinaus: Elena, Ben, Chris, Luis Mi, Markus, Annika, Ines, Alex, Julia und Arianne. Danke für all die schönen Momente, die ich mit euch erleben durfte und für die, die noch kommen werden.

Zuletzt gebührt mein tiefster Dank meiner Familie. Vor allem meinen Eltern Ilona und Andreas, denn ohne eure Unterstützung durch all diese Jahre hätte ich es nicht so weit geschafft.
Bibliography

53. Roslund, M. U.; Säwén, E.; Landström, J.; Rönnols, J.; Jonsson, K. H.; Lundborg, M.; Svensson, M. V.; Widmalm, G. Complete 1H and 13C NMR chemical shift assignments


86. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 1993, 3, 97–130.


144. Hardy, B. J.; Bystricky, S.; Kovac, P.; Widmalm, G. Conformational analysis and molecular dynamics simulation of α-(1 → 2) and α-(1 → 3) linked rhamnose oligosaccharides: Reconciliation with optical rotation and NMR experiments. *Biopolymers* 1997, 41, 83–96.


