Remote Participation in Glycosylations Revealed by Cryogenic Infrared Spectroscopy

Nachweis des entfernten Nachbargruppeneffekts in Glykosylierungen durch kalte Infrarotspektroskopie

Master Thesis

submitted on 14$^{\text{th}}$ August 2019

Department of Chemistry

Faculty of Mathematics and Natural Sciences

Humboldt-Universität zu Berlin

accomplished at the Fritz Haber Institute

Department of Molecular Physics

Name:	Kim Greis
First Examiner:	Prof. Dr. Dietrich Volmer
Second Examiner:	Prof. Dr. Kevin Pagel
Supervisor:	M. Sc. Eike Mucha

Abstract

Controlling the stereochemistry during glycosynthesis is usually achieved by using elaborate protecting group techniques. For example, 1,2-*trans*-glycosidic bonds are formed by the introduction of participating protecting groups at the neighboring C2-position. This reaction is believed to proceed *via* a glycosyl cation, where the protecting group interacts with the positive charge at the anomeric carbon, effectively shielding the *cis*-side from nucleophilic attack, leading to the concept of neighboring group participation. The formation of 1,2-*cis*-glycosidic bonds remains, however, a major hurdle to date. Empirical studies have shown that their formation can be aided by the introduction of participating protecting groups at the remote C4- or C6-positions, leading to the phenomenon of remote participation. Due to the short-lived nature of glycosyl cations, they cannot be studied readily in the condensed phase. The gas-phase environment inside a mass spectrometer, however, can be used to isolate them. In a recent publication, cryogenic infrared spectroscopy was used to unravel the structure of glycosyl cations. Evidence for the existence of neighboring group participation was found in various monosaccharide building blocks, as the acetyl protecting group at the C2-position forms a covalent bond with the anomeric carbon. While these results enable a better understanding of neighboring group participation, the structural motifs underlying remote participation remain poorly understood. Their study would help optimizing reaction conditions aiding the formation of 1,2-*cis*-glycosidic bonds. Here, a combination of first-principles theory and cryogenic infrared spectroscopy in the low-temperature environment of superfluid helium droplets (0.4 K) is used to decipher the structure of galactose building blocks exhibiting remote participation. Galactose building blocks carrying an acetyl protecting group at the C4-position form α-selective dioxolenium-type structures, exhibiting a covalent bond between the anomeric carbon and the oxygen atom of the carbonyl group of the acetyl protecting group. Contrary, galactose carrying an acetyl group solely at the C6-position does not exhibit remote participation and forms non-selective oxocarbenium-type structures. Furthermore, a novel type of interaction between benzyl protecting groups and the anomeric carbon, leading to α-selective oxonium-type structures, is observed. The recorded data can be used by algorithms based on artificial intelligence to predict the best reaction conditions for performing glycosylation reactions with a defined stereochemical outcome.

Contents

List of Abbreviations

AGA	Automated Glycan Assembly
AO	Atomic Orbital
ATR	Attenuated Total Reflectance
CBS	Complete Basis Set
CC	Coupled Cluster
CID	Collision-Induced Dissociation
DCM	Dichloromethane
DFT	Density Functional Theory
DLPNO	Domain based Local Pair Natural Orbital
DNA	Deoxyribonucleic Acid
Fafoom	Flexible Algorithm For Optimization Of Molecules
FF	Force Field
FHI-FEL	Fritz Haber Institute Free-Electron Laser
GA	Genetic Algorithm
GGA	Generalized Gradient Approximation
HF	Hartree-Fock
HO	Harmonic Oscillator
HPLC	High-Performance Liquid Chromatography
IM	Ion Mobility
Imidate	Trichloroacetimidate
IMR	Ion-Molecule Reaction
IR	Infrared
IRMPD	Infrared Multiphoton Dissociation
IVR	Intramolecular Vibrational Redistribution
LDA	Local Density Approximation
LG	Leaving Group
MMFF94	Merck Molecular Force Field 94

MO	Molecular Orbital
MP	*Møller-Plesset*
MPIKG	Max Planck Institute of Colloids and Interfaces
MS	Mass Spectrometry or Mass Spectrometer
nESI	Nano Electrospray Ionization
NMR	Nuclear Magnetic Resonance
NP-HPLC	Normal-Phase High-Performance Liquid Chromatography
PES	Potential Energy Surface
PG	Protecting Group
RI	Resolution of Identity
SCF	Self Consistent Field
Thiolate	Ethanethiolate
ToF	Time-of-Flight
UFF	Universal Force Field
UV	Ultraviolet
WF	Wave Function
XC	Exchange-Correlation

1 Introduction

1.1 Motivation

Carbohydrates are presumably the most diverse class of biomolecules. In nature, carbohydrates exhibit essential biological functions such as providing energy, forming bio-scaffolds or recognizing cellular processes.[1] They are the most abundant class of biomolecules, and among them cellulose is the most abundant biopolymer.[2] The structural diversity of carbohydrates, especially in chain length, leads to many applications in different fields. On the one hand, the biopolymer cellulose is used in material science as a *green* adsorbent for various pollutants such as heavy metal ions, dyes or pharmaceuticals.[3–6] On the other hand, therapeutics based on carbohydrates (mainly oligosaccharides) are currently emerging.[7–9]

The field of glycomics[10,11], analogous to proteomics[12] or genomics[13], is still in its infancy and even though progress has been made in the past years, no reliable high-throughput technique to analyze glycans has been established so far. This lag in development is mainly due to the structural complexity of glycans, contrary to linearly linked peptides or oligonucleotides. Nevertheless, high-resolution data of glycans has been recorded using NMR spectroscopy and X-ray crystallography.[14,15] Besides high sample consumption, accurate structural analysis *via* NMR is often time-consuming both in terms of recording and evaluating data. Furthermore, inter- and intramolecular interactions as well as solvent effects can disturb the analysis of glycans. These drawbacks combined with the complexity of carbohydrates have prevented a growth of glycomics at the same speed as genomics or proteomics.

However, in the last few years, some major analytical advancements have been made in the field of glycomics. Pagel *et al.* have shown that the gas-phase analysis of carbohydrates using mass-spectrometry based techniques is a valuable approach to identify and differentiate glycan isomers differing in composition, connectivity or configuration. On the one hand, ion mobility-mass spectrometry (IM-MS) can be used to separate molecules according to their mass, charge, size and shape,[16–18] while, on the other hand, cryogenic vibrational spectroscopy in helium nanodroplets yields vibrational fingerprints exhibiting well-resolved absorption features that are diagnostic even for minor structural variations.[19,20] Complex carbohydrates can be structurally analyzed by combining both techniques.[21–27]

Recently, it has been demonstrated by Pagel *et al.* that cryogenic vibrational spectroscopy can also be used for the structural analysis of glycosyl cations, the key intermediate in glycosylations.[28] Even though the first glycosylation reaction was performed already in 1893 by Emil Fischer,[29] its mechanism remains poorly understood until now.[30] Whether the mechanism proceeds *via* a dissociative S_N1- or a concerted S_N2-type mechanism depends

on many factors, such as the donor, the activator, the solvent or the temperature, which have been empirically analyzed.[31–33] Due to the lack of structural data on the glycosyl cation, the S_N1-mechanism is particularly poorly understood. It is empirically known that 1,2-*trans*-glycosides can selectively be synthesized through the introduction of participating protecting groups (PGs) at the neighboring C2-position, which is a commonly used method in chemical synthesis. Participating PGs at the C2-position effectively shield the *cis*-side of the anomeric carbon from nucleophilic attack, leading to 1,2-*trans*-glycosides.

Empirical data shows that the selective formation of 1,2-*cis*-glycosides can be induced through participating PGs at the C4- and C6-positions.[34] It is believed that participating PGs at these positions interact remotely with the positive charge at the anomeric carbon, leading to shielding of the *trans*-side.[35] However, while neighboring participation is acknowledged, the concept of remote participation is not fully understood yet. It has been reported that fully benzylated galactose donors yield a high α-selectivity, even though benzyl is formally a non-participating PG.[34] A type of remotely participating benzyl ethers has also been proposed in mannuronic acid lactones, leading selectively to 1,2-*cis*-mannosides.[36] The fact that some PGs act differently than presumed shows that the fundamental chemistry behind glycosynthesis remains elusive. Studying the structure of the glycosyl cation is the key to understanding S_N1-type reactions in glycosynthesis.[37,38]

However, due to the short-lived nature of this intermediate, its comprehensive study has not been possible until recently. Using cryogenic vibrational spectroscopy, it has been demonstrated that neighboring participating acetyl PGs interact with the positive charge of the anomeric carbon in glycosyl cations.[28] The gas-phase environment of a mass spectrometer, where nucleophiles and solvent molecules are explicitly absent, presents perfect conditions to analyze these reactive intermediates. Furthermore, the permittivity of the vacuum, in which these ions are studied, is close to the permittivity of apolar solvents, commonly used in glycosynthesis.[39]

Glycosyl cations were recently studied by Boltje *et al.* using infrared multiphoton dissociation (IRMPD), who were able to get the first experimental evidence for neighboring participation.[40] Furthermore, remote participation was observed in bicyclic 6,3-uronic acid lactones.[41] Even though the data is in agreement with predictions, the presence of multiple conformers, due to the elevated internal energy of the probed ions, leads to line broadening and congested IR spectra and eventually impedes their reliable analysis.[42] Another approach is chosen by Blériot *et al.*, who show that glycosyl cations can be generated and stabilized in the condensed phase using super acids (HF/SbF_5) and subsequently be studied with NMR spectroscopy.[43–45] However, it is unclear if the data can be utilized to anticipate the stereochemical outcome of a glycosylation reaction, as the analyzed glycosyl cations usually carry four positive charges. Besides, the interaction of the super acidic solvent on the conformation of the glycosyl cation is not understood yet.

In this work, a consistent set of fully protected galactose building blocks bearing different combinations of acetyl and benzyl PGs is analyzed *via* cryogenic vibrational spectroscopy and computational methods. The analyte ions are introduced into a mass spectrometer using a nESI source. The ions are mass-selected before being accumulated in a hexapole ion trap, where they are embedded in superfluid helium nanodroplets (0.4 K) and transported to the detection region. There, the embedded ions are probed with the Fritz Haber Institute free-electron laser (FHI-FEL).[46] Absorption of resonant IR photons causes evaporation of the helium matrix, with subsequent release of the previously embedded analyte ion. The bare unsolvated ions are then detected by a time-of-flight (ToF) detector. Plotting the ion signal of the released analyte ions versus the wavenumber of the laser yields an IR spectrum that can be compared to its computed counterpart to identify the conformation of the glycosyl cation. This approach has been successfully used in the past for probing various substances, including small molecules, proteins, carbohydrates and glycosyl cations.[19,20,28,47,48]

The aim of this work is to find experimental evidence of remote participation by probing a consistent set of glycosyl cations generated by galactose donors carrying benzyl and acetyl protecting groups in four different combinations with cryogenic vibrational spectroscopy and quantum chemical theory. The impact of participating protecting groups at the remote C4- and C6-positions on the conformation of the glycosyl cation, including bond distances and ring puckers, is unravelled and compared to empirical data from glycosynthesis. The recorded data can be used to predict the stereochemical outcome of glycosylation reactions.

1.2   Outline of the Thesis

This work presents a combination of experimental and theoretical results on glycosyl cations. The general structure of carbohydrates, the anomeric effect and ring puckers are presented in **Chapter 2 − Fundamentals**. Furthermore, the role of the glycosyl cation in glycosynthesis and infrared spectroscopy including modern action spectroscopy methods are discussed. The theoretical basics of the quantum chemical methods applied to explore the conformational space of the glycosyl cations are explained. In **Chapter 3 − Experimental Section**, the investigated structures, the experimental setup and the free-electron laser are presented.

The multistep approach used to study the glycosyl cations theoretically is described in **Chapter 4 − Molecular Modelling of Glycosyl Cations**. A systematic study on cation–π-interactions and improvements of the applied genetic algorithm software as well as the results of the genetic algorithm including final structures are presented. The final experimental results are shown in **Chapter 5 − Infrared Spectroscopy of Glycosyl Cations**. Experimental and theoretical data are thoroughly analyzed and compared to

the stereochemical outcome of glycosylation reactions. The influence of the leaving group and its configuration on the structure of the glycosyl cation are discussed.

2 Fundamentals

2.1 Carbohydrates

Carbohydrates, synonymously known as glycans, are the class of biomolecules that possesses the largest structural diversity. They are constituted of monosaccharides, which are polyhydroxylated aldehydes or ketones. Monosaccharides are grouped according to their number of carbon atoms into families such as tetroses, pentoses, hexoses or heptoses. The large structural diversity of glycans is due to the large number of stereocenters in each monosaccharide. The prominent isomeric aldohexoses glucose, galactose and mannose, for example, are defined by the configuration of four stereocenters. The focus of this thesis is solely laid on galactose. The structural diversity of monosaccharides is shown schematically in Figure 2.1 using galactose as an example.

Galactose is known as the C4-epimer of glucose, and thus the only difference between these two species is the stereoconfiguration of the C4-carbon atom. Depending on the stereochemistry of the C5-carbon atom, two different enantiomeric forms, L- and D-galactose, exist. Generally, mainly the D-enantiomeric forms of sugars are biologically relevant, whereas L-fucose is a prominent example for a naturally occuring monosaccharide exhibiting an L-configuration. Each aldohexose bears five hydroxyl and one aldehyde group in the short-lived open-chain form. The hydroxyl groups at the C4- and C5-position can attack the sp^2-hybridized carbonyl carbon atom to form a five- (furanose) or a six-membered ring (pyranose). While the reaction is in terms of enthalpy favoured in both cases, the latter is preferred due to less torsional strain. During the reaction, a new stereogenic center is formed at the C1-carbon atom, the anomeric center. In the α-anomer, the anomeric center has the same configuration as the C5-carbon atom, whereas the stereoconfiguration of both centers are different in the β-anomer.

It is observed that in aqueous solution both anomers interconvert into each other *via* the short-lived open-chain form (mutarotation). For glucose, the equilibrium α:β-ratio at room temperature is 39:61 of the pyranose form, whereas only traces of the open-chain and the furanose form are present.[50] β-D-glucose is favored over the α-anomer because of steric interactions. While all substituents adopt the less hindered equatorial orientation in the β-anomer, the C1-substituent adopts the less stable axial orientation in the α-anomer, leading to destabilizing 1,3-*syn*-axial interactions. Subsequently, if solely these effects are taken into account, a higher ratio towards the β-anomer would be expected. The unexpectedly high α-ratio is caused by the *anomeric effect*, which is defined as "the thermodynamic preference for polar groups bonded to the C1-carbon atom to take up an axial position".[51] The anomeric effect is usually explained by either classical physics or quantum mechanics. The classic explanation considers the dipoles involving the endocyclic
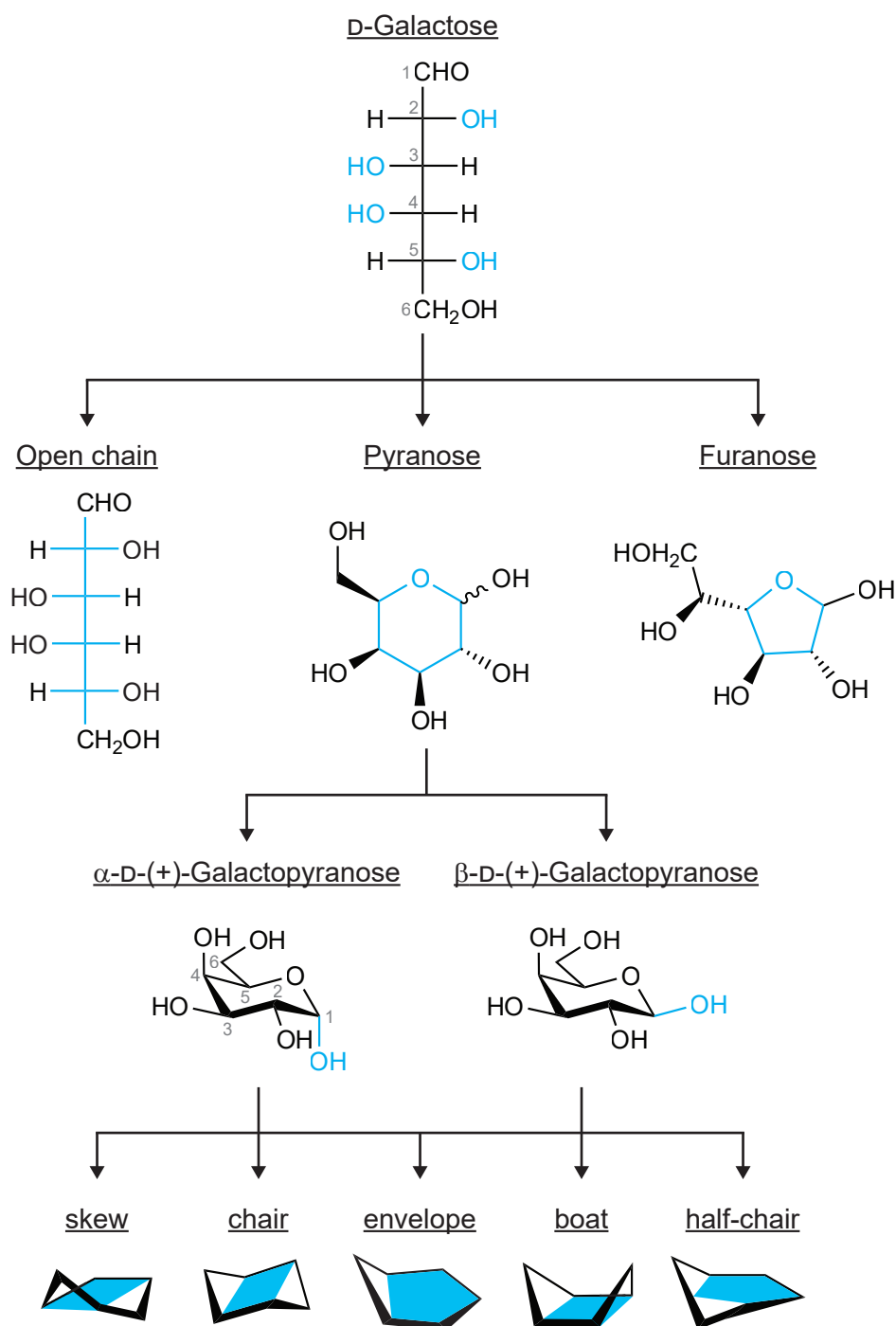
Figure 2.1: Schematic illustration of the structural diversity of monosaccharides using galactose as an example. Pucker representation is adopted from HILL *et. al.*[49]

oxygen and the substituent at the C1-position. The dipoles are partially aligned, if the substituent at the anomeric carbon is orientated equatorially, and thus repelling each other. However, if the substituent is orientated axially, the dipoles are opposing each other, which is energetically favored (Figure 2.2a). The quantum mechanical explanation involves negative hyperconjugation, which is a stabilizing interaction between an unshared electron pair on the endocyclic oxygen and the σ*-orbital of the axial C–O-bond (Figure 2.2b). The effect can also be observed for electronegative heteroatoms that are different from oxygen. Both explanations are questioned because they are not always in line with experimental data.[52]
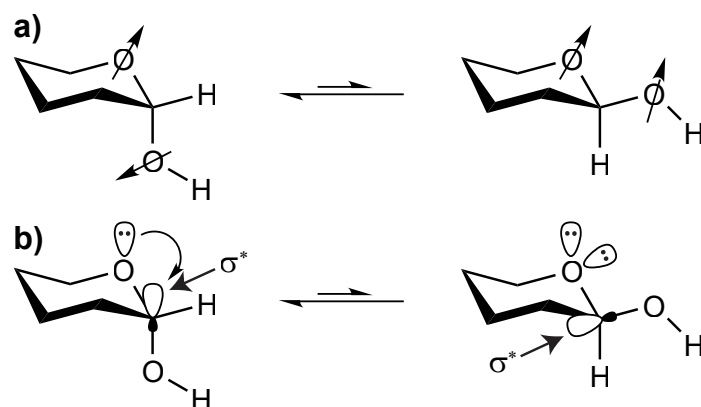

Figure 2.2: Contemporary explanations for the anomeric effect: (a) Opposing dipoles are energetically favored, whereas aligned dipoles are disfavored. (b) Energetically favored electron donation from the lone pair of oxygen in the empty σ*-orbital of the C–O-bond is possible in axially substituted but not in equatorially substituted monosaccharides.

Carbohydrate rings can adopt various conformations, known as ring puckers, to minimize steric repulsion of the substituents. In total there are 38 canonical puckers in pyranose rings. The conformations can be categorized as chair (C), boat (B), envelope (E), half-chair (H) or skew (S) (Figure 2.1). For each conformation, a reference plane consisting of at least four carbon atoms is defined and the atoms above and below the plane are denoted together with the reference letter to yield the IUPAC nomenclature.[53] A more general definition of ring puckering coordinates was established by CREMER and POPLE, who discovered that the conformation of each pyranose ring can be fully described by three parameters (R, θ, ϕ), including distortions not covered by the canonical puckers.[54,55]

2.2 Glycosyl Cations in Glycosynthesis

Glycosyl cations, also called oxocarbenium ions, are short-lived intermediates in glycosynthesis. They are formed by cleavage of a leaving group at the C1-position of a monosaccharide building block. Glycosylation reactions can be described by two mechanisms, the first one

being unimolecular S_N1-reactions involving a glycosyl cation intermediate and the second one being bimolecular S_N2-reactions that proceed through a concerted associative transition state (Figure 2.3). There is, however, a large gap between these two extrema and most glycosylation and chemical reactions in general are rather situated in a mechanistic continuum in between. While S_N2-reactions are usually highly stereoselective, provoking an inversion of the stereocenter, the stereoinformation is lost in S_N1-reactions. In some glycosylation mechanisms the formation of contact ion pairs has been observed, even though formally dissociative, exhibiting the nature of S_N2-type reactions in terms of stereoselectivity.[31]
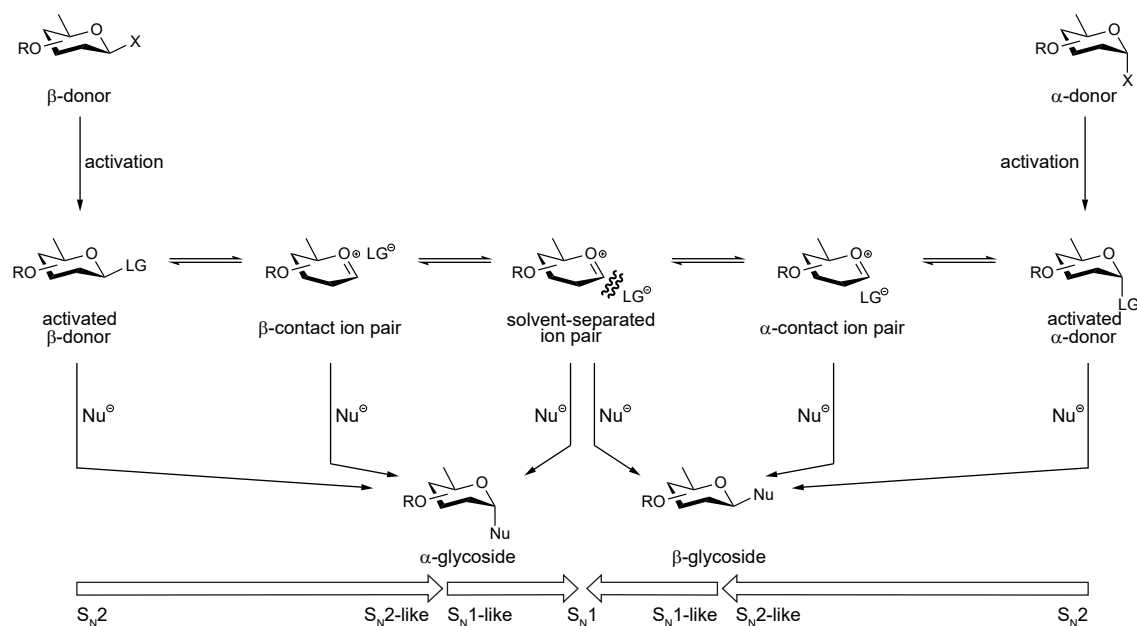

Figure 2.3: Glycosylation reaction with two associative S_N2-mechanisms and one central dissociative S_N1-mechanism. Adapted from ADERO *et al.*[31]

Many parameters are believed to influence the S_N1–S_N2-interface, as for instance the nature of the donor, the acceptor, the leaving group, the concentration, the reaction time, the activator, the solvent or the temperature.[33] While the selectivity of S_N2-type reactions in glycosylations is mainly dependent on the stereoconfiguration of the anomeric carbon, the prediction of the selectivity of S_N1-type reactions is more complicated. Here, the short-lived glycosyl cation is formed and it has been postulated that its conformation, which is mainly influenced by its substituents, is responsible for the stereochemistry of the products.[35]

A commonly used method to introduce stereoselectivity in S_N1-type glycosylation reactions is the use of participating protecting groups, as for example acetyl (OAc) or benzoyl (OBz), at the neighboring C2-position. Thereby, the selective formation of 1,2-*trans*-

glycosidic linkages[1] is achieved. Neighboring participation leads to effective shielding of the *cis*-side from nucleophilic attack. The shielding occurs *via* the formation of dioxolenium-type structures exhibiting a covalent bond between the carbonyl oxygen of the participating protecting group and the anomeric carbon, as recently shown for glucose, galactose and mannose.[28,40]

The formation of 1,2-*cis*-glycosides is generally less understood. Empirical data shows that the introduction of participating protecting groups at the C4- and C6-positions leads to α-selectivity for galactose, which indicates the existence of *remote participation*.[34] Here, the protecting group interacts remotely with the anomeric carbon. The type of interaction is, however, unknown and can range from pure oxocarbenium- to dioxolenium-type structures (exemplarily shown for C4-acetylated galactose in Figure 2.4). Oxocarbenium-type structures exhibiting a long-range interaction between the participating PG and the anomeric carbon are situated between both types of structures. Recently, first experimental evidence of remote participation for a very restricted set of samples of bicyclic glycosyl cations has been published.[41] Despite the dissociative nature of the S_N1-type mechanism, the selective shielding of one side by participating PGs during nucleophilic attack leads to a high stereoselectivity. A further type of remote participation with non-participating benzyl PGs, leading to increased selectivity, has been postulated.[34,36] Here, the oxygen atom of the benzyl ether PG interacts remotely with the anomeric carbon to form oxonium-type structures.

2.3 Infrared Spectroscopy

Infrared spectroscopy, also called vibrational spectroscopy, is a method to probe the vibrational modes of molecules and molecular ions. Each molecule exhibits $3N$ degrees of freedom, with N being the number of atoms in the molecule, while three degrees of freedom are occupied by translation and rotation, respectively, whereas the remaining $3N - 6$ degrees of freedom are occupied by vibration. Linear molecules only exhibit two rotational degrees of freedom and thus possess $3N - 5$ vibrational degrees of freedom. Vibrational modes are visible in an IR spectrum if a change in the dipole moment occurs during the vibration. Modes that are not IR active can be probed with the complementary method of Raman spectroscopy, in which vibrational modes are visible that exhibit a change in polarizability of the electrons. The spectra are usually recorded in the mid-IR range (400–$4000\,\mathrm{cm}^{-1}$) and are diagnostic for each molecule. The spectra can usually be divided into two regions. In the fingerprint region below $1500\,\mathrm{cm}^{-1}$ numerous C–H, C–C and C–O absorption bands define the spectrum, showing a distinct pattern for each

---

[1]The *cis/trans*-nomenclature needs to be distinguished from the α/β-nomenclature. In the former, the carbon atoms involved are explicitly specified, while in the latter the relative stereoconfiguration at the C1- and C5-carbon atoms is considered in hexoses.
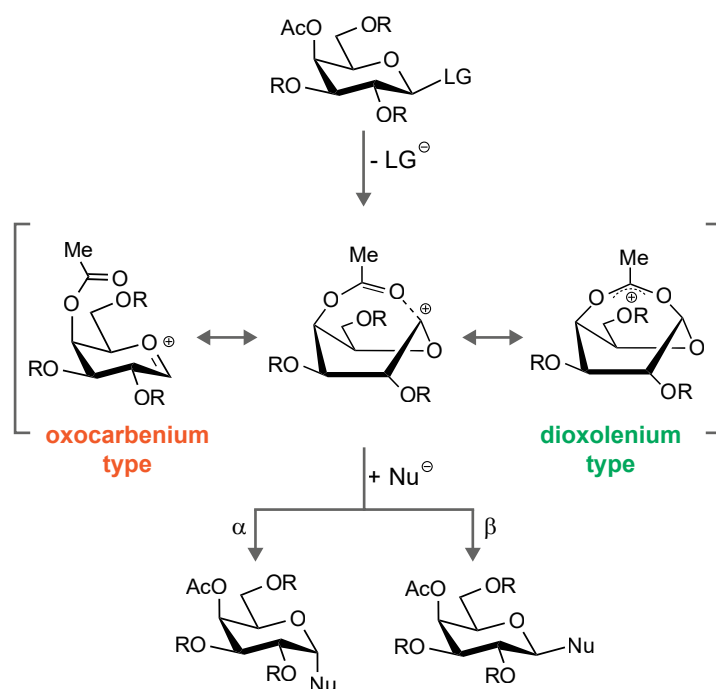
Figure 2.4: Proposed glycosyl cation intermediates with C4-remote participation. The type of interaction can range from oxocarbenium-type ions to hybrid structures and dioxolenium-type ions with a covalent bond between the carbonyl oxygen of the participating protecting group and the anomeric carbon. Adapted from DEMCHENKO *et al.*[35]

molecule that cannot be comprehensively understood and predicted accurately by quantum chemical theory in most cases. Above $1500\,\mathrm{cm}^{-1}$, the distinct vibrations of functional groups, such as the C=O stretching vibration at roughly $1700\,\mathrm{cm}^{-1}$, can be identified. However, it is difficult to correlate the obtained spectra with the structure of the molecule without complementary quantum chemical calculations.

The potential energy $E(r)$ of a chemical bond as a function of the distance r between two atoms can be described with the Morse potential (Equation 1), which is an approximation that correctly describes the anharmonic behaviour of a chemical bond at short and long distances. If two atoms are closer than equilibrium bond distance r_e, the potential energy of the chemical bond increases exponentially, while it only rises to the dissociation energy D_e for increasing distances. The constant factor a is called the *stiffness* of the chemical bond. The Morse potential might describe the chemical bond in an adequate way, but it is computationally demanding, which is why the behaviour of the chemical bond is often approximated with the model of the harmonic oscillator (HO), in which the chemical bond behaves like a classic spring, with k being its force constant (Equation 2). In this model, the potential energy of the chemical bond rises quadratically with the changes in bond distance from r_e. The discrete energetic levels of the vibrational states of the

quantum mechanical HO are equidistant and quantized as $E_\nu = (\nu + \frac{1}{2})\hbar\omega$ $(\nu \in \mathbb{N}_0)$, with ω being the frequency of oscillation. In the Morse potential, the discrete energy levels get closer with rising quantum number ν. The HO is a good approximation in the equilibrium region, but it fails to explain the limiting behaviour of very short and very long distances (Figure 2.5). Thus, the HO does not explain the dissociation of a chemical bond.

$$E(R) = D_e \cdot \left(1 - e^{-a \cdot (r - r_e)}\right)^2 \tag{1}$$
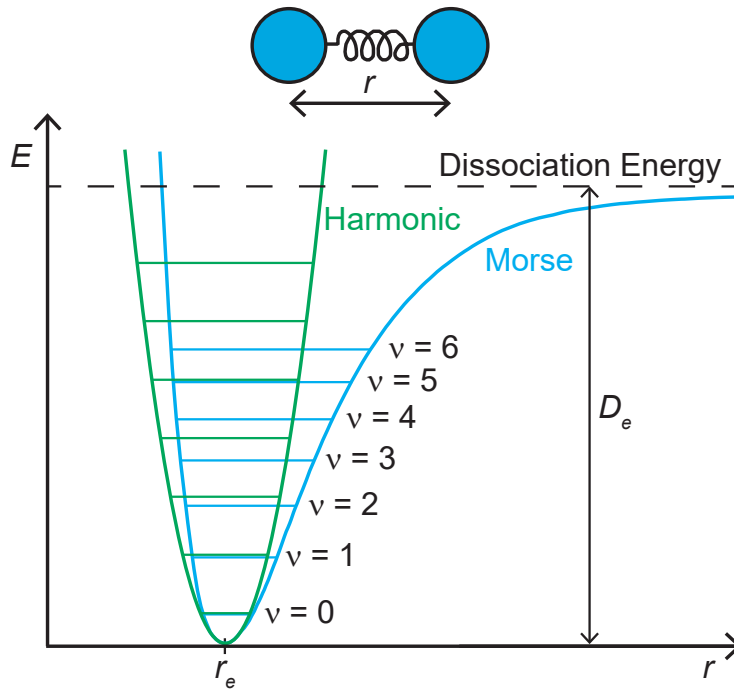
$$E(R) = \frac{1}{2}k \cdot (r - r_e)^2 \tag{2}$$

Figure 2.5: Schematic representation of the harmonic oscillator (green) and the Morse potential (blue). Contrary to the energy levels of the HO, the energy levels of the Morse potential are not equally spaced. The Morse potential accurately describes the dissociative behavior of a diatomic molecule for increasing interatomic distances *r*, with a dissociation energy D_e.

In a classical IR experiment, vibrational modes are detected by absorption spectroscopy. Here, the attenuation of the intensity of IR light I_0 that traverses a sample with the number of density n, the path length L and the wavelength dependent absorption cross section $\sigma(\lambda)$ is measured. The transmitted intensity of IR light is denoted as $I(\lambda)$. Vibrational

excitation of a molecule occurs if a resonant photon is absorbed. Photons able to excite a vibration have the wavelength of the (mid)-IR domain. The absorbance A can be plotted against the wavelength λ or as typically done in IR spectroscopy against the wavenumber $\tilde{\nu}$ to give an IR spectrum. Absorption spectroscopy can be described by the Beer-Lambert law (Equation 3).

$$I(\lambda) = I_0 \cdot e^{-\sigma(\lambda) \cdot L \cdot n} \Leftrightarrow A(\lambda) = \ln \frac{I_0}{I(\lambda)} = \sigma(\lambda) \cdot L \cdot n \tag{3}$$

Absorption spectroscopy is usually carried out either in the condensed or in the solid phase, causing some disadvantages. In the condensed phase, the vibrations of the solvent may overlap with the ones of the sample and interactions between the solvent and the sample may lead to a less accurate spectrum. In the solid phase, the sample can be either incorporated into a KBr pellet or measured with Attenuated Total Reflectance (ATR) IR spectroscopy. This requires however that the solid sample is stable and does not react with its environment. Also, ATR-IR spectra are rather broad. The biggest disadvantage is that short-lived intermediates cannot be measured with these techniques. A suitable method for studying these species is transferring them to the gas-phase environment of a mass spectrometer, where non-desired reactants are absent. Typically, IR spectroscopy requires a large number of density of the species of interest, which is not fulfilled in most gas-phase experiments. Thus, the difference between I and I_0 would not be measurable. However, IR spectra can still be recorded in the gas-phase using *action spectroscopy*, which makes use of the effect of IR radiation on the analyte. The fraction of affected molecules $N(\lambda)$ is expressed in Equation 4, where $\Phi(\lambda)$ is the photon fluence experienced by the sample at a distinct wavelength.

$$N(\lambda) = N_0 \cdot \left(1 - e^{-\sigma(\lambda) \cdot \Phi(\lambda)}\right) \tag{4}$$

There are different types of action spectroscopy that are commonly applied, as for example *IR multiphoton dissociation* (IRMPD), messenger (tagging) or vibrational spectroscopy in helium droplets. The foundation of these methods is usually an action event that occurs during the absorption of a single or multiple photons. Such events are fragmentation, dissociation of a tag or the release of the analyte ions from helium droplets. The signal yield of the thereby generated ions can be monitored with ToF detectors and plotted versus the wavenumber to give an IR spectrum.[56]

In IRMPD, fragmentation of the analyte ion occurs due to the absorption of multiple resonant photons. After the absorption of a resonant photon, the vibrational energy quickly dissipates *via intramolecular vibrational redistribution* (IVR). In this process, anharmonic coupling to the vibrational background states leads to depletion of the excited vibrational

mode while increasing the internal energy of the molecular ion. Upon absorption of multiple photons and subsequent IVR events, the molecule is thermally activated and statistical fragmentation occurs (Figure 2.6a). The main disadvantages of IRMPD are due to the IVR mechanism: the elevated internal energy of the molecule results in the formation of multiple conformers, with absorption bands differing only by a few wavenumbers, causing line broadening. The increased internal energy also causes frequency shifts of the absorption bands. These drawbacks often lead to congestion in IR spectra, especially in the case of glycans that exhibit a high conformational flexibility. Nevertheless, IRMPD has been widely used to study peptides, small carbohydrates and metal clusters.[42,57–60]

In messenger spectroscopy, a tag (or messenger) binds weakly to the analyte ion. The absorption of a single resonant photon leads to the dissociation of the tag from the bare analyte ion, whose signal intensity is monitored (Figure 2.6b). This leads to IR spectra with well-defined narrow bands. The messenger usually is a noble gas atom, H_2, D_2 or N_2. The perfect messenger should exhibit a high tagging efficiency, not change the geometry of the tagged ion, be IR transparent and be easily detachable upon the absorption of a photon. The attachment of a tag requires cryogenic temperatures depending on the nature of the tag (e.g. 3 K for tagging He).[61] The slow cooling of the ions that enter the cooled ion trap also lead to conformer selectivity, which makes this method interesting for studying biological molecules. However, recent findings have shown that the energetically most stable conformer might not always be found using this method, especially when there are multiple energetically low-lying conformers.[62] Drawbacks are changes in geometry that may be induced by the tag and low tagging efficiencies.

A method that combines the advantages of both techniques is cryogenic vibrational spectroscopy in helium nanodroplets, which is also used for the experiments presented in this thesis. The ions that are accumulated in an ion trap are embedded in a matrix consisting of a superfluid helium droplet (0.37 K) that is composed of ca. 10^5 atoms. Then, they are transported to the detection region, where the analyte ions are probed with an IR laser. Upon absorption of a resonant photon, IVR occurs in the analyte ion, leading to the dissipation of energy to the vibrational background states. Even though the internal energy of the analyte ion rises by this process, the energy is quickly transferred to the cryogenic helium matrix through relaxation. Thus, the analyte ion reattains the energetical ground state, while the helium matrix slowly evaporates. This prevents heat activation of the analyte ion and thus it is believed that the trapped conformers will neither be activated by the laser nor the geometry be changed by the matrix. Upon subsequent absorption of a certain number of resonant photons the helium droplet releases the bare ion, which is detected afterwards, for example, by a ToF detector (Figure 2.6c). A drawback is that even though the helium droplets are very cold, the ion trap is still at ambient temperature. Depending on the flexibility of the analyte ions, there might be multiple conformers present
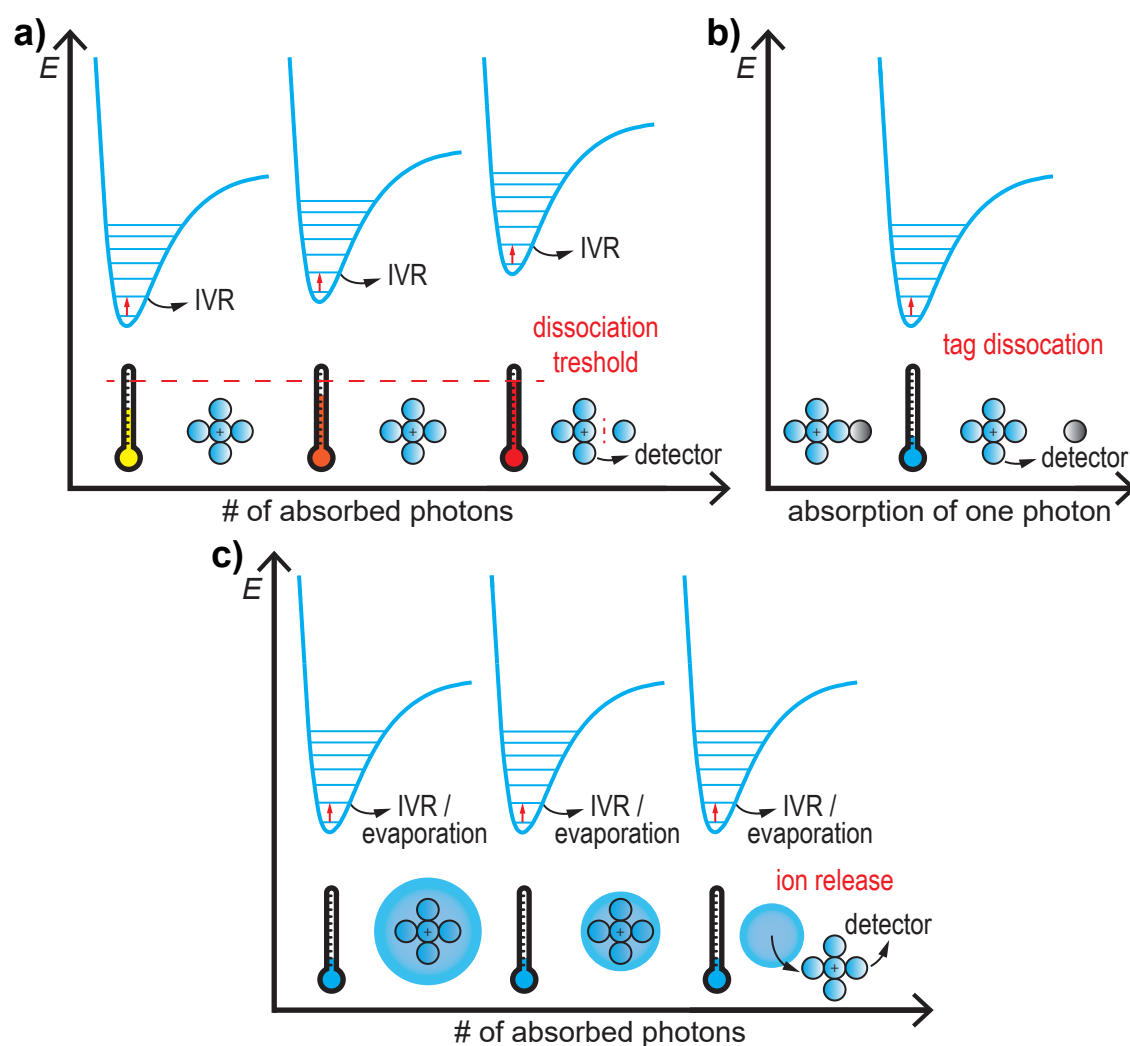
13

Figure 2.6: Schematic comparison between the underlying principles of (a) IRMPD, (b) messenger spectroscopy and (c) vibrational spectroscopy in helium droplets. (a) In IRMPD, the internal energy of the analyte ions rises through absorption of resonant photons and subsequent IVR, leading to fragmentation. (b) In messenger spectroscopy, the tag dissociates from the analyte ion after the absorption of a single photon, thus no thermal activation occurs. (c) In helium droplets, the absorbed energy is dissipated to the helium bath, which subsequently evaporates, to conserve a low temperature.

at room temperature, which will subsequently be kinetically trapped in the helium droplet matrix and then probed, leading to less defined spectra. A way to circumvent that issue is cooling the ion trap with liquid nitrogen to 78 K. This method is a valuable approach to record IR spectra with narrow absorption bands of various biomolecules.

2.4   Molecular Modelling

In the following section, a brief overview of the theoretical methods to determine the low-energy structures of glycosyl cations is given. While force field methods and molecular dynamics are usually a reasonable choice to explore the conformational space of biomolecules, the increase in computing power during the past decades allows to use more costly quantum mechanical methods for exploring the potential energy surface (PES) of a molecule within a reasonable amount of time. The PES is defined as the potential energy of a system as a function of its atomic coordinates. Valuable information such as the molecular geometry, vibrations or transition states can be extracted from the PES.

Quantum Mechanical Methods 1: Wave Function Based Methods

Fundamental understanding of a molecular system arises through studying its electronic nature. Electrons, however, are very light particles and can therefore not be studied by classical mechanics. A reasonable approach to get the wave function (WF) of a system, from which the energy of the system can be extracted, is to solve the time-independent Schrödinger equation (Equation 5).[63] In *ab-initio* methods only the electronic structure is considered, contrary to *(semi-)empirical* methods, where some parameters are fitted to experimental data in order to save computational time. The Born-Oppenheimer approximation allows solving the Schrödinger equation while neglecting coupling of the motion of nuclei and electrons because of the different time scales of their movements due to their huge mass difference. However, the Schrödinger equation can only be solved analytically for one-electron systems such as the H-atom or the He$^+$-cation. Thus, numerous approximations have been introduced to obtain reliable results for systems possessing multiple electrons, some of which are presented in the following sections.

$$H\Psi = E\Psi \tag{5}$$

The Hartree-Fock (HF) theory[64,65] was the first to give results of acceptable accuracy. Here, the motion of each electron is considered to be independent from the dynamics of all other electrons, while the interactions are taken into account in an average fashion. Each electron is described as an orbital and all occupied orbitals form the wave function. Electrons are fermions (particles with spin $\frac{1}{2}$) and are thus antisymmetric (change of sign upon interchanging two electrons). That behavior is taken into account by arranging the orbitals into a Slater determinant. The orbitals that give the lowest energy (HF orbitals) are determined by the variational principle. The HF theory generates the best single-determinant WF. For improving the results systematically, electron correlation needs

to be taken into account, which is achieved through multideterminant methods, being more costly in terms of computational power.

Further *ab-initio* methods that are worth mentioning and routinely used to calculate the correlation energy are the *Møller-Plesset Perturbation Theory*[66] (MP) and the *Coupled Cluster Theory*[67] (CC). Commonly thereof derived methods are MP2 and CCSD(T). In MP2, the second-order energy correction is considered, which means that the correlation between pairs of electrons is described. Also, doubly excited determinants are taken into account, while higher excitations are neglected. Considering higher terms is possible and usually known as MP3 or MP4. However, MP3 often yields worse results than MP2, whereas MP4 is for most systems too expensive. Typically the MP2 correlation energy covers 80-90 % of the correlation energy and is the most economical method including electron correlation.[68]

The idea of CC is to include *all corrections of a given type to infinite order*.[69] This means that each conceivable correlation between a given number of electrons is considered. Moreover, in infinite CC, all excited states would be considered. However, considering all corrections to infinite order is hardly feasible, which is why usually truncated CC is used. Most commonly only singly, doubly and triply excited states are considered, which is known as CCSDT. A good approximation for triply excited states is their approximation *via* perturbation theory and adding this energy as a correction to the CCSD energy. This method is known as CCSD(T).

A severe problem, as already mentioned, is computational cost. Each method scales differently with the number of electrons N of the system. While HF and MP2 scale with a factor of N^{3-4} and N^{4-5}, respectively, CCSD(T) scales with a factor of N^{6-7}. Therefore, approximations have been introduced for MP2 and CCSD(T) that give results of similar accuracy, with a more generous scaling behavior. The *resolution of identity* approximation RI-MP2 decreases the computational cost by one order of magnitude, while maintaining a high accuracy almost equal to MP2.[70] For CCSD(T) the *Domain based Local Pair Natural Orbital* (DLPNO) approximation has been introduced. This method calculates up to 99.9 % of the CCSD(T) correlation energy, maintaining an almost linear scaling behaviour.[71,72]

Quantum Mechanical Methods 2: Density Functional Theory

Commonly used and computationally less costly methods to accurately calculate the energy of a system are density functional theory (DFT) based methods. While the previously presented methods extract the ground state electronic energy from the wave function of the system, the ground state energy in DFT is completely determined by the electron density ρ, as proven by Hohenberg and Kohn.[73] Contrary to the WF that contains $4N$ variables, N being the number of electrons, the electron density only depends on three

spatial coordinates, independent of the number of electrons. However, the functional that connects the electron density and the ground state energy is not known. Later, Kohn and Sham found out that the kinetic energy can be calculated from an auxiliary set of orbitals used for representing the electron density.[74] Thus, the exchange-correlation (XC) energy, which is only a small fraction of the total energy, remains unknown. However, the introduction of orbitals increases the dependence of the electron density from 3 to $3N$ variables. Modelling the XC part remains a major hurdle until now and yielded numerous different DFT functionals that are more or less accurate for a given system.

The first functionals used the so-called *Local Density Approximation* (LDA). Here, the electron density is treated locally as a uniform electron gas. This method is often very inaccurate for molecular calculations and was improved by considering a non-uniform electron gas. This is fulfilled by making the XC energies not only dependent of the electron density, but also of its derivatives. This approximation is called *Generalized Gradient Approximation* (GGA). Pure DFT functionals have a high accuracy for the correlation energy, while they often fail to reliably describe the exchange energy, which is more accurately covered by the HF method. This leads to the nowadays asserted idea of mixing HF and DFT energies. Such DFT functionals are called *hybrid* functionals (e.g. B3LYP). They provide relatively good results for low computational costs. *Double-hybrid* functionals are an even more sophisticated method. Here, the DFT energy is not only mixed with the HF energy, but also with the MP2 energy. The elevated accuracy, however, comes with a price, as hybrid methods require additional HF and/or MP2 calculations and thus adapt their scaling behavior.[75]

DFT methods also fail to describe long-range interactions such as dispersion that are especially in larger systems of great importance. A reasonable approach to include these type of interactions is adding an empirical r^{-6} dependent dispersion term to the DFT calculated energy. The method has been further refined and can be used with a negligible increase in computational cost.[76,77]

Besides the method, each calculation also needs a basis set, which is a set of basis functions that describe approximatively atomic orbitals (AOs). Molecular orbitals (MO) are a linear combination of AOs. Each AO can potentially be described with an infinite number of basis functions, which would then give the most accurate result. However, this is computationally not feasible and thus each orbital is described with a restricted number of basis functions. An efficient approach applied in most basis sets is to model the valence orbitals with an elevated number of functions, while the core orbitals are only modelled with one basis function. Some basis sets also include polarization functions, which allow antisymmetric behaviour of the MOs. The commonly used basis set def2-TZVPP, which is a *triple zeta valence* basis set, models each orbital with three basis functions and two sets of polarization functions.[78] Ideally, the energy obtained with a *triple-zeta* (tz) and

quadruple-zeta (qz) basis set of the same family, for example, can be extrapolated to the complete basis set (CBS) using Equation 6 for the HF and Equation 7 for the correlation energy, which can be added to get the total energy (Equation 8).[68,79,80]

$$E_{HF}^{CBS} = \frac{2 \cdot E_{HF}^{qz} - E_{HF}^{tz}}{\exp\left(\alpha \cdot (\sqrt{4} - \sqrt{3})\right) - 1}, \quad \alpha = 7.88 \tag{6}$$

$$E_{corr}^{CBS} = \frac{E_{corr}^{tz} \cdot 3^{\beta} - E_{corr}^{qz} \cdot 4^{\beta}}{3^{\beta} - 4^{\beta}}, \quad \beta = 2.97 \tag{7}$$

$$E_{tot} = E_{corr} + E_{HF} \tag{8}$$

Genetic Algorithm

Genetic algorithms (GAs) are used in theoretical chemistry to sample chemical structures based on biological concepts in order to find ultimately the lowest-energy conformer of a given molecule. At first, an initial population is created by random variation of torsional angles, *cis/trans*-bonds or even the pucker of pyranose rings. The fitness of the as-generated structures is determined by comparing their respective energies. Low-energy structures have a high fitness and are more likely to serve as parent structures for the next generation. Features from two parent structures are taken to generate a set of child structures, whose energy is determined to calculate their fitness. A suitable portion of the best parent and child structures are used for sampling the next generation. To add some randomness to the algorithm, mutation is allowed to happen in all steps. The genetic algorithm is usually terminated when no change in the best structure occurs for a certain number of generations or after a limited number structures have been sampled.

If the GA uses non-optimized structures, then its philosophy is Darwinian, because the child structures inherit features that the parent structures inherited from their parents and so on. If each generated structure is subjected to a structure optimization, then the philosophy of the GA is rather Lamarckian, because then the features acquired through the structure optimization are inherited to the child structures. This philosophy is for example used in the *flexible algorithm for optimization of molecules* (Fafoom) that is also used in this thesis.[81,82] Every generated structure is sent to an external program for optimization with force fields (FFs) or quantum chemical methods. FFs use empirical parameters to calculate the electronic energy as a function of the nuclear coordinates, while electrons are neglected. Due to the low cost, it is the only valuable method for large biomolecules. Most FFs can only be reliably used for a small number of molecules, for which they have been parametrized and yet deliver always poorer results than quantum chemical methods.

For carbohydrates, the pyranose pucker is often sampled as a chair by most FFs, which especially leads to erroneous results during the consideration of glycosyl cations. Therefore, it is a good approach to sample small molecules, with less than 100 atoms, rather with DFT methods than FFs.

3 Experimental Section

3.1 Investigated Structures

The precursors for the investigated glycosyl cations were prepared by SEEBERGER *et al.* The investigated samples consist of galactose bearing a leaving group (LG) at the C1-position and protecting groups in four distinct combinations. The glycosyl donors consist of fully benzylated, 4-acetylated, 6-acetylated and 4,6-di-acetylated galactose building blocks, while the remaining hydroxyl groups are benzylated. The glycosyl cations are formed by cleavage of trichloroacetimidate or ethanethiolate LGs, abbreviated as imidate and thiolate. The thereby generated glycosyl cations 2,3,4,6-tetra-*O*-benzyl-D-galactopyranoside, 4-*O*-acetyl-2,3,6-tri-*O*-benzyl-D-galactopyranoside, 6-*O*-acetyl-2,3,4-tri-*O*-benzyl-D-galactopyranoside and 4,6-di-*O*-acetyl-2,3-di-*O*-benzyl-D-galactopyranoside are in the following part of the thesis abbreviated as **Bn**, **4Ac**, **6Ac** and **46Ac**. The chemical structures of the investigated samples are shown in Figure 3.1. All samples have been examined with α-imidate LGs, while a limited number of samples were also analyzed with β-imidate and β-thiolate LGs to study the influence of the nature of the leaving group and its configuration on the conformation of the glycosyl cation. The measured samples are summarized in Table 1.
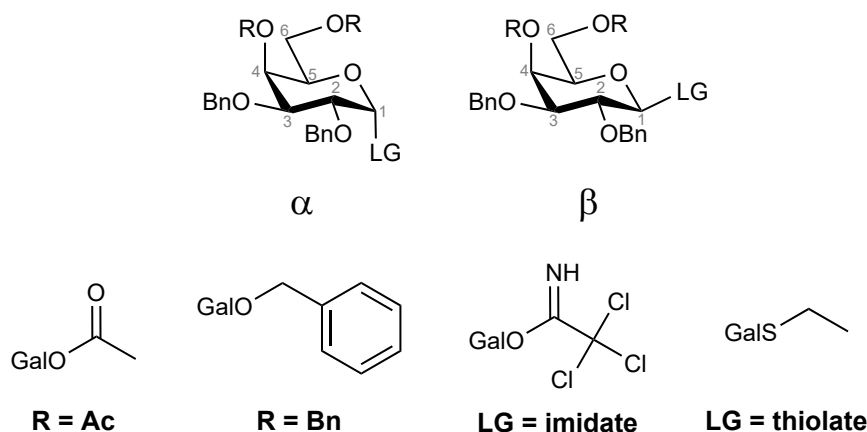
Figure 3.1: Chemical structures of the investigated samples.

3.2 Experimental Setup

The experimental setup used in this thesis has been published previously.[19,22,83,84] It combines mass spectrometry with cryogenic vibrational spectroscopy in helium droplets (Figure 3.2). The setup consists of a modfied Q-ToF Ultima (*Waters Corporation*) mass spectrometer that is combined with a helium droplet source, a hexapole ion trap, a free-electron laser and a time-of-flight detector.

Table 1: Investigated samples are labelled in green, while the non-investigated counterparts are labelled in red.

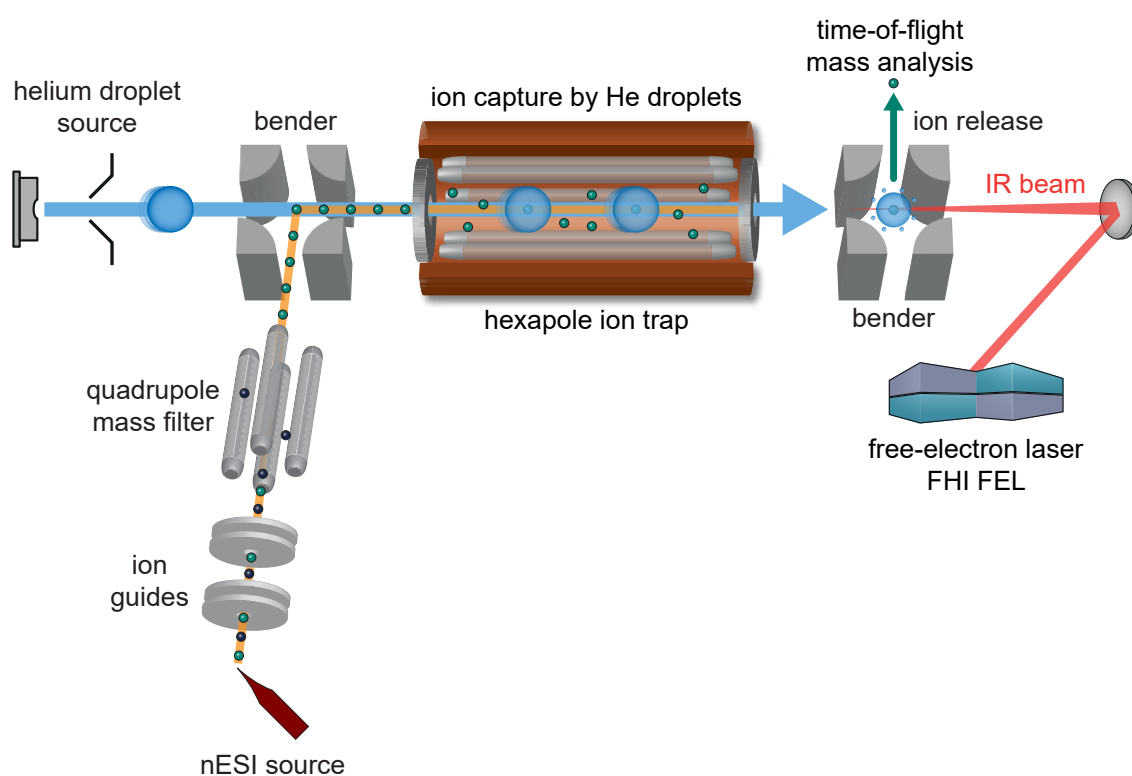| Donor	Imidate		Thiolate	
	α	β	α	β
46Ac	■	■	■	■
4Ac	■	■	■	■
6Ac	■	■	■	■
Bn	■	■	■	■

Figure 3.2: Schematic illustration of the experimental setup. The sample is ionized *via* nESI, the ions are *m/z*-selected in a quadrupole filter and accumulated in a hexapole ion trap. There, the analyte ions are picked up by helium droplets and get transported to the detection region, where they are probed by the FEL.

All samples are dissolved in a mixture of acetonitrile and water (9:1) to yield a 100–200 µM solution. The sample is ionized *via* nano electrospray ionization (nESI) using Pd/Pt coated glass capillaries (Sputter Coater HR 208, *Cressington*), which are pulled to a tip with an inner diameter of 1–2 µm employing a micropipette puller (Model P-1000, *Sutter Instrument*). Charged droplets containing the analyte ions are formed after applying a voltage of 1 kV to the tip of the capillary. Bare analyte ions, that are formed after solvent evaporation and droplet fission, are either deflected towards a skimmer and then guided into the instrument *via* a Z-spray source or directly guided into the instrument *via* a capillary source. The instrument setup allows to change between the Z-spray and the capillary source. The ions are then guided into a quadrupole mass filter by two ring electrode ion guides, where they get mass-to-charge selected. Then the ions pass through a collision cell, which is only used for guiding the ions to a DC quadrupole bender that the ions can either pass directly to enter a reflectron time-of-flight (ToF) detector or get deflected by 90° to fill a linear hexapole ion trap. In the ToF detector, the measured time of flight of the ions is used to derive the *m/z*-ratio. The detector is mainly used for monitoring the signal during optimization of the source conditions and the voltages of the mass spectrometer.

In the hexapole ion trap, the ions are confined radially by a radio-frequency waveform (1.1 MHz, 400 Vpp) and trapped longitudinally by endcap electrodes at the entrance and the exit that are kept 3 V above the DC offset of the trap. Filling the ion trap with helium buffer gas leads to removal of excess kinetic energy and thermalization of the analyte ions. The copper housing of the hexapole ion trap allows cooling with liquid nitrogen to thermalize the ions to 78 K. The temperature in the ion trap can be adjusted from 78 to 400 K by a heater. The buffer gas is removed when the ion density in the trap is sufficiently high.

Superfluid helium droplets are generated by a pulsed Even-Lavie valve. Expansion of helium (^{4}He) from high pressure and low temperature into vacuum leads to the formation of a fast beam of droplets. In this work, the valve is operated at 21 K, leading to droplets with an average size of 10^5 atoms. The valve is cooled by a closed-cycle helium cryostat and operated at a repetition rate of 10 Hz, an opening time of 17 µs and a stagnation pressure of 68 bar. The droplets pass through a skimmer with an opening of 4 mm diameter that is mounted 135 mm away from the nozzle to enter the instrument. While traversing the hexapole ion trap, the helium droplets can pick up analyte ions and cool them rapidly to 0.4 K. The temperature of the droplets is maintained through evaporative cooling. The kinetic energy of the doped helium droplets is high enough to overcome the longitudinal trapping and transport the embedded ions to the detection region. Here, the droplets and the IR laser beam overlap. Subsequent absorption of multiple resonant photons leads to the release of bare ions from the droplets that are then guided to a ToF detector by a DC

quadrupole bender. The ToF signal of the released ions monitored against the wavenumber yields the IR spectrum of the analyte ion.

3.3   Free-Electron Laser

As highlighted in Equation 4 in Section 2.3, the number of molecules affected in a gas-phase action IR spectroscopy experiment rises with the photon fluence. In the experiment described herein, tabletop lasers would conventionally not meet the needed photon fluence, which is why a more powerful photon source is needed: the free-electron laser (FEL). In this thesis, the FHI-FEL is used as a source for high intensity laser radiation (Figure 3.3).[46]
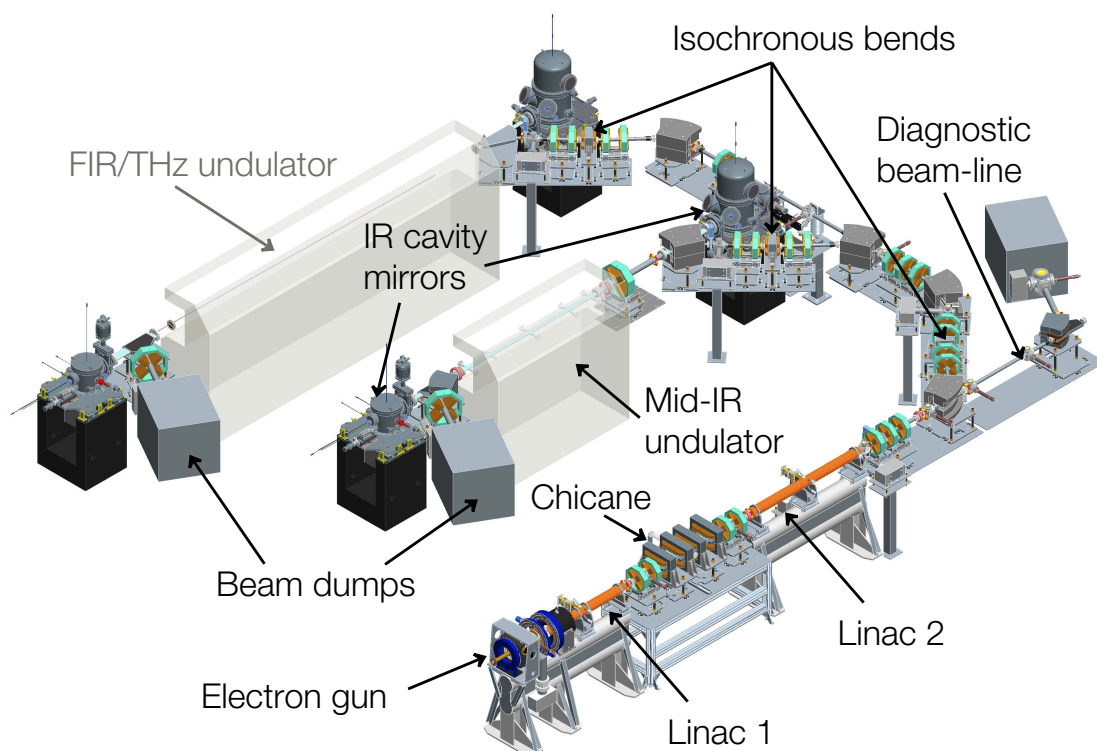

Figure 3.3: Schematic setup of the Fritz Haber Institute free-electron laser (FHI-FEL).

The FHI-FEL consists of an electron gun that releases electrons that are subsequently accelerated to velocities close to the speed of light and energies of up to 50 MeV in two linear accelerators (LINACs). The electron beam is then bent and injected into a resonator that consists of two high-reflectivity mirrors at each end and an undulator. The undulator exhibits a strong magnetic field that is perpendicular to the direction of the electron beam and periodically changes polarity along its length. This causes a periodic deflection of the electron beam, leading to emission of light. Coherent monochromatic

high-intensity radiation is obtained due to the interaction between the radiation and the electron beam, leading to self-amplification. The wavelength can be tuned by changing the gap of the undulator. The FHI-FEL can produce photons in the mid-IR range with a wavelength tunable from 3 to 60 μm. Pulsed radiation with 0.3–0.5% FWHM bandwith of the corresponding wavelength is provided by a micro-pulse repetition rate of 1 GHz, a micro-pulse length of 3–5 ps, a macro-pulse repetition rate of 10 Hz, a macro-pulse length of 10 μs and pulse energies of up to 120 mJ. The kinetic energies of the electrons used for producing high-intensity laser irradiation in this thesis is 36 MeV.

4 Molecular Modelling of Glycosyl Cations

The quantum chemical treatment of glycosyl cations isolated in the gas-phase is presented in this chapter. A multistep strategy is employed to sample the conformational space of glycosyl cations using DFT methods. The most promising structures are reoptimized at a higher level of theory in order to calculate their IR spectra and single point energies. A benchmark study on cation–π-interactions has been carried out using various wave function and DFT methods to find the most reliable theoretical method for calculating single point energies.

4.1 Computational Details

In this section, the computational details of the procedure for generating the final structures of the glycosyl cations are explained. A simplified scheme of the employed multistep procedure is shown in Figure 4.1.

The geometry files of the starting structures are generated using Avogadro[85] or Gaussview and are then subjected to the genetic algorithm Fafoom.[81] All available rotatable bonds and ring puckers have been selected as degrees of freedom in the GA. For each glycosyl cation, ten individual GA runs are carried out. The settings of the GA are shown in Table S1. The GA automatically sends each structure to an external software for structure optimization and uses the acquired properties of the optimized structures for generating new structures. Here, the external software FHI-aims[86] is used at PBE-vdW^{TS}/$light$[87,88] level of theory.

The initial set of structures obtained after GA sampling is clustered into two subsets depending on the interaction of the participating protecting group with the anomeric carbon. If the O–C1-bond distance is smaller than $2\,\text{Å}$, O is covalently bound to the anomeric carbon to form a dioxolenium- or an oxonium-type cation, while it is not or only weakly binding when the O–C1-bond distance is larger than $2\,\text{Å}$ (oxocarbenium-type cation). For acetyl protecting groups, the C=O–C1-distance is measured, while the Bn-O–C1-distance is measured for benzyl protecting groups. Ring puckers are assigned according to CREMER-POPLE coordinates.[54] Bond distances and ring puckers are determined using a self-written *Python* script (Listing 1 in the appendix).

The low-energy structures for each subset are reoptimized at PBE0-D3/6-311+G(d,p) level of theory using default convergence criteria in Gaussian 09, RevD.01.[76,77,89–91] Subsequently, a frequency analysis within the harmonic approximation is performed. Each IR spectrum is normalized and scaled by a linear scaling factor of 0.965. The line spectra are convoluted by gaussian functions.
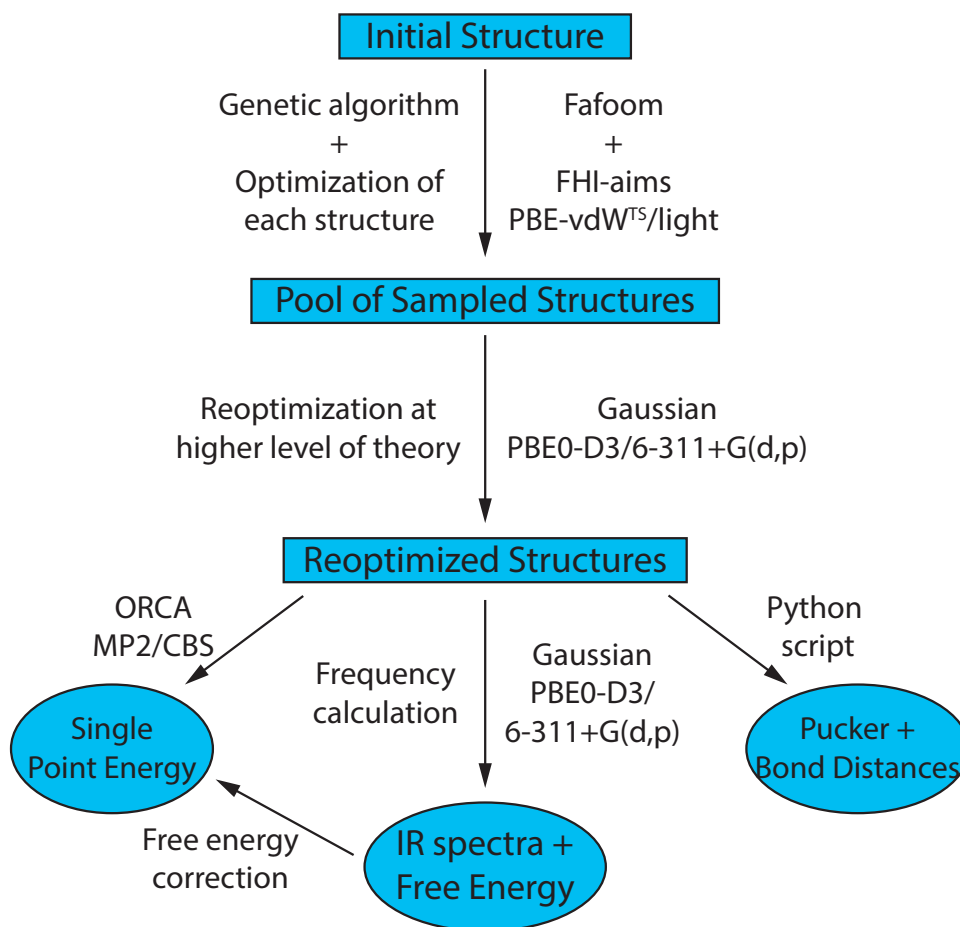
Figure 4.1: Schematic representation of the employed multistep procedure described in this section to compute IR spectra and final energies.

The RI-MP2[70] energies extrapolated to the complete basis set are calculated for each reoptimized structure in ORCA 4.1.[92] Two-point extrapolation using the energies obtained for the basis sets def2-TZVPP and def2-QZVPP that are also used as auxiliary basis sets, is carried out using Equations (6)–(8).[78,93] Grid5 settings and tight SCF convergence are requested. The obtained single point energies are corrected by a free energy contribution at 78 K derived from harmonic vibrational calculations.

4.2   Benchmarking Cation–π-Interactions

Cation–π-interactions are expected to play a major role in glycosyl cations.[94] The positive charge in the glycosyl cations is often stabilized by the π-system of up to two benzyl protecting groups, if no participating group is interacting with the anomeric carbon. To accurately determine the energy of these systems, a benchmark study has been carried out on $NaC_6H_6^+$ and $Na(C_6H_6)_2^+$ ions. In these systems, an atom with a positive charge

(sodium) is coordinated by either one or two π-systems (benzene). These systems are reasonably similar to those studied in the further sections and small enough to serve as a model to study cation–π-interactions in detail.

The ions $NaC_6H_6^+$ and $Na(C_6H_6)_2^+$ have been optimized at PBE0-D3/def2-TZVPP level of theory. Then, a series of single point energy calculations using DLPNO-CCSD(T)[71], RI-MP2 and various DFT functionals have been carried out, with variable distances between the sodium ion and the benzene ring(s). The utilized DFT functionals are the double hybrid functionals DSD-PBEP86[95] and PWPB95[96], the hybrid functionals ωB97X,[97] M06, M062X,[98] X3LYP,[99] B3LYP[100–102] and PBE0[89] as well as the GGA functionals SCANfunc[103] and PBE.[87] Each DFT calculation has been performed using the D3-dispersion correction, except for X3LYP, for which the dispersion correction has not been yet implemented into the ORCA software. The def2-TZVPP basis set has been used for all calculations, except for RI-MP2 and DLPNO-CCSD(T), where def2-TZVPP and def2-QZVPP have been used to extrapolate the energy to CBS. The MP2-part for the calculations using double hybrid DFT functionals has been calculated using the RI-approximation. In total, 787 calculations have been carried out for this study in ORCA 4.1 using Grid5 settings and tight SCF convergence criteria.

The energy is plotted as a function of the $Na^+-C_6H_6$ distance (Figure 4.2). For $NaC_6H_6^+$ (Figure 4.2a), the $Na^+-C_6H_6$ distance of the optimized structure is 2.38 Å. The distance has been decreased to 1.78 Å and increased to 2.98 Å respectively in 0.1 Å intervals. From 2.98 Å to 4.98 Å, the distance has been increased in 0.5 Å intervals. For $Na(C_6H_6)_2^+$, the $Na^+-C_6H_6$ distance of the optimized structure is 2.45 Å. In Figure 4.2b, the $Na^+-C_6H_6$ distance remains constant for one benzene ring, while in Figure 4.2c the $Na^+-C_6H_6$ distance between Na^+ and both benzene rings has been varied. The distance has been decreased to 1.75 Å and increased to 3.05 Å respectively in 0.1 Å intervals. From 3.05 Å to 5.05 Å, the distance has been increased in 0.5 Å intervals. All energies have been referenced to the energies of completely dissociated C_6H_6 and Na^+ or $NaC_6H_6^+$, respectively.

Generally, the lowest energies can be found using DLPNO-CCSD(T)/CBS and RI-MP2/CBS. CCSD(T) is the "gold standard" in computational chemistry and thus the method to go.[104] As the RI-MP2 energies are almost equal to the DLPNO-CCSD(T) energies, it can be estimated that RI-MP2 can calculate cation–π-interactions sufficiently. The RI-MP2 energies are lower than the DLPNO-CCSD(T) energies, because the correlation energy is underestimated in MP2. The energies calculated with DFT methods are generally higher than those calculated with wave function based methods. There is no need to extrapolate the DFT energies to CBS, because they do not depend as much on the basis set as the correlation energy does in wave function methods. For all three systems, the DFT methods are not able to reproduce a similar curve, in terms of both shape and energetics, as the wave function based methods. Furthermore, the difference in energetics seems to
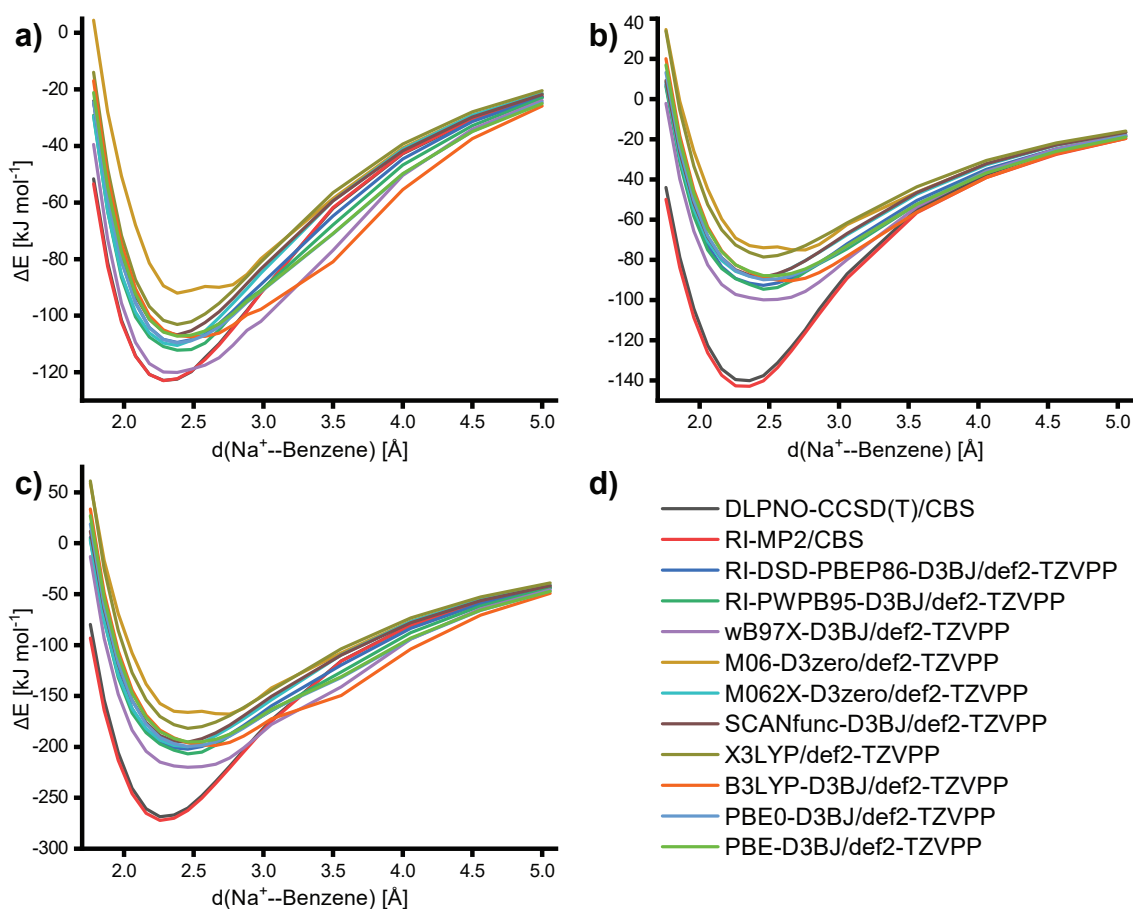
Figure 4.2: Single point energy of $NaC_6H_6^+$ (a) and $Na(C_6H_6)_2^+$ (b,c) plotted as a function of the distance between a sodium ion and benzene. (a) sodium is coordinated by one benzene ring, (b) sodium is sandwiched between two benzene rings, while only one Na^+–Benzene distance is variable, (c) sodium is sandwiched by two benzene rings, while both Na^+–Benzene distances are variable, (d) list of methods used with color code.

increase with increasing number of π-systems interacting with the positive charge. Thus, the best compromise between computational cost and efficiency is RI-MP2/CBS, which is used in the following sections to calculate the energies of the reoptimized structures.

4.3   Modelling the Conformational Space of Glycosyl Cations

Force Fields: A Valuable Approach to Study Glycosyl Cations?

The conformational space of glycosyl cations has been modelled using the evolutionary algorithm Fafoom. The conformational space of molecules is usually sampled using empirical force fields methods because of their low computational cost. However, these methods rely only on empirical parameters and are therefore biased. In the systems studied herein,

there are two main reasons why force fields methods are not a reasonable choice. First, for carbohydrates, force fields tend to optimize the structures into chairs or other puckers, which are often different from those predicted by more accurate quantum chemical methods. Second, the interactions between a protecting group and the anomeric carbon cannot be grasped by force fields.

Due to the low computational cost of force fields, the conformational space of **46Ac** has been sampled using the *Merck Molecular Force Field 94* (MMFF94) and the *Universal Force Field* (UFF), that are implemented in RDKit and which can be used with Fafoom. It is theorized that the acetyl groups in **46Ac** interact with the anomeric center to form dioxolenium-type structures. This would require the formation of a covalent bond between the carbonyl oxygen of one acetyl group and the anomeric carbon. Therefore, the energetics of the sampled structures have been plotted as a function of the C=O4–C1- and the C=O6–C1-distance for the MMFF94 (Figure S1) and the UFF (Figure S2) force field. For both methods, it is observable that only oxocarbenium-type structures are generated, as no covalent bond is formed between either carbonyl oxygen and the anomeric carbon. Only a few structures show weak interactions, where the C=O4–C1-distance is less than $3\,\text{Å}$. Moreover, from all sampled structures (437 for MMFF94 and 450 for UFF), more than $60\,\%$ have optimized in a chair pucker ($^1C_4$ or $^4C_1$). This shows that force fields are not a good approach to study the systems treated in this master thesis. As a consequence, the systems are studied with DFT and WF based methods in the following sections.

Proof of Principle Study: Linking Fafoom to Gaussian

Fafoom has been developed to be used in combination with FHI-aims for DFT optimizations and RDKit for force field optimizations. However, since its development, parsers have been written to use Fafoom with other quantum chemical software packages as ORCA or NWChem. No parser was available for Gaussian, which is unquestionably one of the most often used software in quantum chemistry. The development of such a parser and a proof of principle study of its performance as well as further modifications to Fafoom are described in this section.

Fafoom and all its components are written using the programming language *Python*. Thus, the Gaussian parser needs to be written in the same language. The parser that has been written within the scope of this thesis is represented in Listing 2 in the appendix. The parser initializes the arguments that need to be specified in the Fafoom input file if Gaussian is chosen to optimize the generated structures. Some arguments are required, while the others are optional (default values are used, if they are not specified). In the next step, the parser creates the Gaussian input file (.gjf) from the initial geometry file (.sdf) and the information specified in the Fafoom input file. Then, a subprocess is opened

to submit the Gaussian input file. The output file is opened and checked for convergence. Then the final energy and structure are saved and the former is being converted from Hartree to eV.

To check if the parser works properly, the conformational space of **46Ac** has been sampled using Fafoom in combination with Gaussian 16[105] at PBE-D3/6-31G level of theory. Fine grid settings and loose optimization criteria have been chosen. In total 112 structures have been generated and their energies have been plotted as a function of the C=O4–C1- and the C=O6–C1-distance in Figure 4.3. The plot clearly shows that numerous structures have been generated that exhibit a covalent bond between the C4-acetyl PG and the anomeric carbon. The most stable dioxolenium-type structure is $38\,\mathrm{kJ\,mol^{-1}}$ more stable than the most stable oxocarbenium-type structure. Structures exhibiting a covalent bond between the C6-acetyl PG and the anomeric carbon have not been found. $43\,\%$ of all optimized structures adopt the same pucker as the most stable structure $^1S_5$. Only one structure adopts a chair pucker. The results are in agreement with those obtained with FHI-aims and experimental results, as presented in the following sections. Thus, DFT is contrary to force fields a valuable approach to study glycosyl cations and is less biased from empirical parameters. Besides, the parser to link Fafoom to Gaussian works and is likely to find global minima after thorough sampling. Structures exhibiting a dioxolenium-structure formed by the C6-acetyl group have not been found with Gaussian, even though such structures have been found with FHI-aims and are supposed to be local minima on the PES of **46Ac**. A possible explanation might be the low number of structures that have been sampled.

Besides the implementation of the Gaussian parser into Fafoom, further changes have been included to the genetic algorithm. A feature to mutate the pucker of the pyranose ring has been implemented and Fafoom now recognizes glycosyl cations and takes the planar geometry of the sp^2-hybridized anomeric carbon into account during the structure generation before optimization. The newest version of Fafoom with traceable changes from previous versions can be downloaded from `https://github.com/kimgreis/fafoom-dev/tree/kim-mod`.

Sampling the Potential Energy Surface of Glycosyl Cations using FHI-aims

The conformational space of the donors **46Ac**, **4Ac**, **6Ac** and **Bn** has been sampled using the evolutionary algorithm Fafoom in combination with FHI-aims at PBE-vdW^{TS}/*light* level of theory. The number of generated and unique structures by the GA is shown in Table S2. The relative energy as a function of the C=O4–C1-distance is shown in Figure S3 and Figure S4 for **46Ac** and **4Ac**, while the respective plots for **6Ac** and **Bn** are shown in Figure S5.
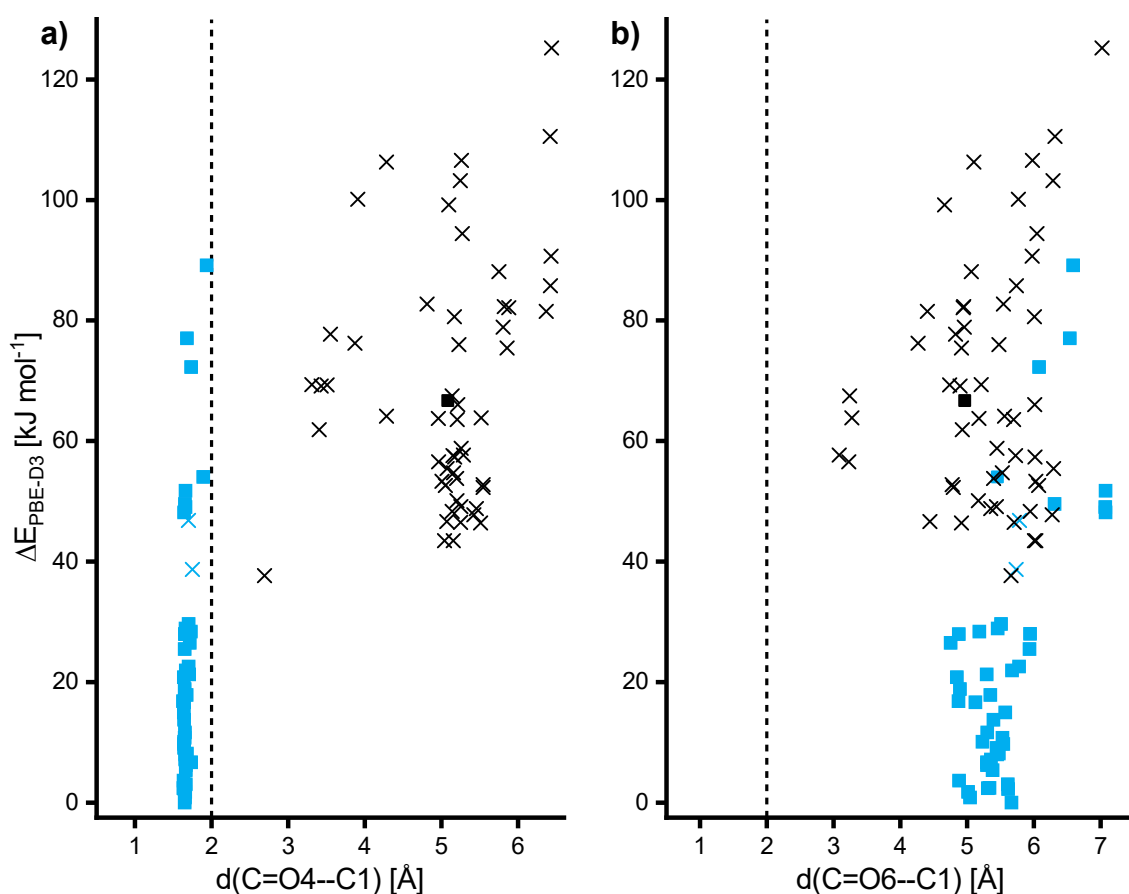
Figure 4.3: Energy hierarchies of various conformers of **46Ac** sampled using Fafoom with Gaussian 16 at PBE-D3/6-31G level of theory. The relative energies are represented as a function of the (a) C=O4–C1- and the (b) C=O6–C1-distance. Oxocarbenium-type structures are represented in black, while C4-acetyl dioxolenium-type structures are represented in blue. The square marks highlights the structures that have the same ring pucker as the energetically most stable structure ($^1S_5$), while the other ring puckers are represented by a cross.

As the number of structures is too high to discuss every single one in detail, only the main trends are presented. Generally, the structures can be divided in two main groups: either the acetyl protecting group is covalently bound to the anomeric carbon (dioxolenium-type structure) or it is not bound (oxocarbenium-type structure). These main groups can be divided into further subgroups. For **46Ac**, both acetyl PGs can bind to the anomeric carbon and thus two distinct dioxolenium-type structures are possible. For **4Ac**, **6Ac** and **Bn**, theory predicts that the non-participating benzyl PGs at the C4- and C6-position can interact with the anomeric carbon to form oxonium-type structures. However, these structures are not distinguished from the oxocarbenium-type structures in

the energy hierarchy plots for **4Ac** and **6Ac**. Only in the case of **Bn**, where dioxolenium-type structures are absent, the O6–C1-distance is used to distinguish oxonium- from oxocarbenium-type structures. Here, the interaction of the C6-benzyl PG leads to a major stabilization.

4.4   DFT- and MP2-Modelling

A set of low-energy structures of each main group is reoptimized followed by a frequency analysis at PBE0-D3/6-311+G(d,p) level of theory. Then, the RI-MP2/CBS single point energies of each optimized structure are calculated and a harmonic free energy correction at 78 K is added. The energy hierarchies of the newly generated structures are shown in Figures S6-S8. The relative energy, the pucker, the O4–C1- and the O6–C1-distances of the lowest-energy structure of each subgroup are summarized in Table 2.

Table 2: Low-energy conformers of **46Ac**, **4Ac**, **6Ac** and **Bn** reoptimized at PBE0-D3/6-311+G(d,p) level of theory. Puckers and bond distances d (in Å) are derived from reoptimized structures. The energy is recomputed at RI-MP2/CBS level of theory, while the harmonic free energy correction at 78 K is derived from a frequency calculation at PBE0-D3/6-311+G(d,p). $\Delta E_{MP2} + \Delta F_{PBE0}$ (in kJ mol^{-1}) is abbreviated as ΔE. Puckers labelled with an asterisk (*) are distorted. The glycosyl cation **K** labelled with a dagger (†) has been generated by *chemical intuition* instead of the GA.

Glycosyl cation		Conformation	Pucker	d(O4–C1)	d(O6–C1)	ΔE
	A	dioxolenium (C4)	$^1S_5$	1.5	5.3	0
46Ac	**B**	dioxolenium (C6)	$^1C_4$	5.2	1.5	20
	C	oxocarbenium	^3E	5.1	3.6	59
	D	dioxolenium (C4)	$^1S_5$	1.5	4.6	0
4Ac	**D'**	dioxolenium (C4)	$^1S_5$	1.5	4.6	3
	E	oxonium (C6)	$^1C_4^*$	4.6	1.5	16
	F	oxocarbenium	^4E	5.2	4.7	50
	G	dioxolenium (C6)	$^1C_4$	4.1	1.5	0
6Ac	**H**	oxonium (C4)	$^{1,4}B$	1.5	4.0	16
	I	oxocarbenium	$^5H_4^*$	4.1	2.8	20
	J	oxonium (C6)	$^1C_4$	4.0	1.5	0
Bn	**K**†	oxonium (C4)	$^{1,4}B$	1.5	4.4	51
	L	oxocarbenium	^4E	3.1	4.7	19

Table 2 gives insight into some facts that have already been foreshadowed by the GA data. While the energetics have changed substantially, the bond distances and puckers have only been slightly altered, due to the already optimized nature of the structures

after the GA. Five main motifs are distinguishable: C4- and C6-acetyl dioxolenium-, C4- and C6-benzyl oxonium- and oxocarbenium-type structures. Generally, dioxolenium-type structures exhibit the lowest energy, with C4-acetyl dioxolenium-type structures being more stable than C6-acetyl dioxolenium-type structures. Oxonium-type structures are more stable than oxocarbenium-type structures. For **4Ac**, a major energetic stabilization of $34\,\mathrm{kJ\,mol^{-1}}$ (relative to the lowest-energy oxocarbenium-type structure) can be observed upon interaction of the C6-benzyl PG with the anomeric center. For **6Ac**, a similar interaction between the C4-benzyl PG and the anomeric carbon only leads to an energetic stabilization of $4\,\mathrm{kJ\,mol^{-1}}$. For **Bn**, only C6-benzyl oxonium-type structures have been found by the GA, which are $19\,\mathrm{kJ\,mol^{-1}}$ more stable than oxocarbenium-type structures. Because C4-benzyl oxonium-type structures have not been found by the GA, but can potentially be formed by this glycosyl cation, such a structure has been generated by *chemical intuition*. Then, it has undergone the same multistep procedure as the other structures. However, the as-generated structure is less stable than the related C6-benzyl oxonium-type structure by $51\,\mathrm{kJ\,mol^{-1}}$. Even though this structure has been generated using the previously sampled geometry of the lowest-energy structure of **6Ac** exhibiting C4-benzyl participation, it is difficult to predict whether this structure is a low-energy structure of **Bn** for this type of interaction or if it presents just a high-energy local minimum on the PES of this structure. Structures generated without using a GA (or similar methods) are unlikely to represent low-energy conformers and can only be regarded as a rough guideline. The facts that the GA has not generated these type of structures and that a guessed structure is highly elevated in energy are certainly good hints that their formation is rather disfavored.

Regarding the ring puckers, some trends can clearly be observed for the low-energy structures. C4-acetyl dioxolenium-type structures prefer $^{1}S_{5}$ skew puckers, while C6-acetyl dioxolenium- and C6-benzyl oxonium-type structures prefer $^{1}C_{4}$ chair puckers. Oxocarbenium-type structures exhibit ^{4}E or ^{3}E envelope puckers, except for **6Ac**, where the closest matching pucker is $^{5}H_{4}$.

4.5   Theoretical Infrared Spectra

The harmonic frequencies of the reoptimized structures have been computed at PBE0-D3/6-311+G(d,p) level of theory. The calculation of the frequencies leads directly to an IR spectrum. The theoretical IR spectrum of the lowest-energy conformer of **46Ac** (conformer **A**) is shown in Figure 4.4, highlighting the general features and regions in the IR spectra of the probed samples. The spectra are shown in the experimentally accessible range of 980 to $1800\,\mathrm{cm^{-1}}$.
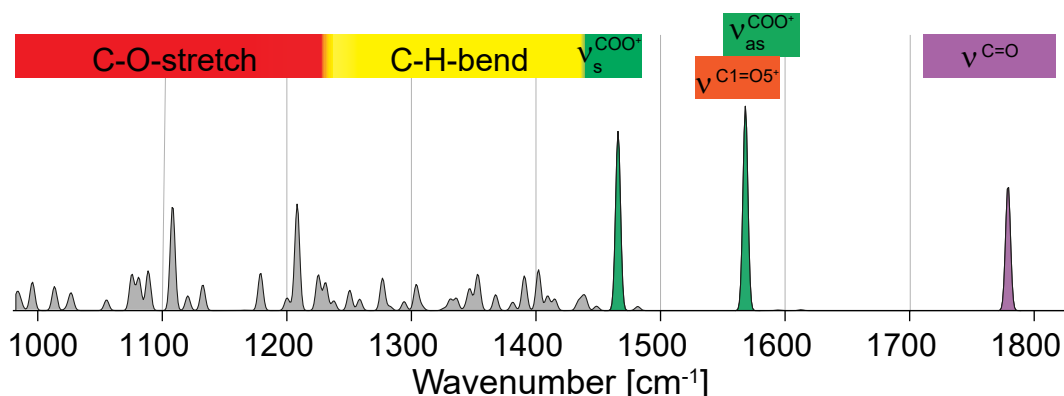
Figure 4.4: DFT calculated IR spectrum of conformer **A** (**46Ac**) at PBE0-D3/6-311+G(d,p) level of theory, serving as an example. The spectrum can be divided into three major parts: C–O-stretching vibrations (980–1240 cm^{-1}), C–H-bending motions (1240–1440 cm^{-1}), COO$^+$- and C=O-stretching modes (1440–1810 cm^{-1}). The latter can be subdivided into symmetric ν_s and asymmetric ν_{as} dioxolenium (COO$^+$)-stretching modes, oxocarbenium (C1=O5$^+$)-stretching vibrations and classical C=O-stretching motions. Oxocarbenium-stretching vibrations are not visible in the presented spectrum.

The spectra can be divided into three major regions of different importance. From 980 to 1240 cm^{-1}, C–O-stretch bands dominate the spectrum. The high transition dipole moment of these vibrations leads to absorptions with a large intensity. The second region from 1240 to 1440 cm^{-1} mainly consists of weak absorptions caused by C–H$_X$-bending motions. The weak intensity of the bands can be explained by the small change in transition dipole moment during the C–H$_X$-bending motions. Both regions are hard to model theoretically with DFT methods within the harmonic approximation. Except some distinct C–O-stretch vibrations with large intensity, the features in these regions are mainly ignored during the discussion.

The third region between 1440 and 1810 cm^{-1} consists of COO$^+$- and C=O-stretch absorption bands. These absorption bands are characteristic and appear at distinct positions in the spectrum depending on the conformation. The dioxolenium-motif (COO$^+$) is represented by two distinct bands, the symmetric and the antisymmetric COO$^+$-stretch, appearing at roughly 1440–1480 and 1550–1610 cm^{-1}, respectively. The oxocarbenium-stretch (C1=O5$^+$) appears in a similar region, from 1530 to 1590 cm^{-1}. It is, however, unlikely that both signals overlap, since the formation of the dioxolenium-motif usually disables the oxocarbenium-stretch. Besides, the dioxolenium-stretches are nonetheless characteristic, since they exhibit two distinct absorption bands. The third feature in this region is the classical C=O-stretch, which is usually located between 1700 and 1810 cm^{-1}.

The absorption bands that are located in the third region are often diagnostic for a distinct conformer and are used in the next section to establish matches between

experimental and theoretical spectra to elucidate the gas-phase structure of the analyzed glycosyl cations. Further theoretical spectra are shown and directly compared to the experimental spectra in the following section.

5 Infrared Spectroscopy of Glycosyl Cations

5.1 Experimental Details

The precursors of the investigated samples have been synthesized by Alonso Pardo and Sooyeon Moon of the Seeberger group at the Max Planck Institute of Colloids and Interfaces (MPIKG). Sooyeon Moon has carried out the glycosylation reactions using flow chemistry and determined the α:β-ratios of the products using an automated microreactor/HPLC platform.

The infrared spectra of the glycosyl cations have been recorded using the experimental setup and the samples described in Chapter 3. The probed glycosyl cations are generated through in-source fragmentation of the thiolate and imidate precursors. Glycosyl cations can be formed *via* collision-induced dissociation (CID).[40,41] However, in the used experimental setup, the collision cell is not activated for performing CID. Furthermore, the collision cell is situated behind the quadrupole mass filter, rendering a mass-selection after fragmentation impossible. This may lead to signals of multiple ions and a decrease of total signal in the IR spectrum. Thus, the glycosyl cations were generated by CID *via* in-source activation.[28]

An exemplary mass spectrum of the α-trichloracetimidate precursor of **46Ac** is shown in Figure 5.1. Besides the glycosyl cation at m/z 427, the sodium adduct of the precursor can be observed at m/z 610. The glycosyl cations are mass-to-charge selected using a quadrupole mass filter.

The spectra have been recorded in the range of 980 to 1800 cm^{-1} in steps of 2 cm^{-1}. The intensity of the signals scales nonlinearly with the absorption cross-section and the laser energy, as a result of the multiple photon process. As a first-order approximation, a linear correction has been performed by dividing the signal by the laser energy. Thus, the intensities of the absorption bands should just serve as a rough approximation, while the peak positions are not expected to be influenced by the multiphoton process of the experiment. It is presumed that the ions are probed in their vibrational ground state. However, trapping of multiple conformers is not necessarily prevented using the described setup. Due to the large number of possible puckers and rotatable bonds in the investigated systems, many conformers are likely to coexist. Elevated temperatures lead to an increased number of distinct conformers, due to thermal population of higher energy conformers. The conformation of the trapped ions is not expected to change after being embedded in the helium droplet, as they are rapidly cooled to 0.4 K. A change in conformation is usually connected to an activation barrier (transition state) that cannot be overcome at ultracold temperatures. To reduce the number of conformers in the ion trap as much as possible and prevent congestion of the signals in the spectra, the ion trap has been cooled to 78 K using

liquid nitrogen. In summary, the recorded spectra are likely to represent a cross-section of all the conformers of an analyte ion in the hexapole ion trap at a given temperature.

For each investigated glycosyl cation, at least two IR spectra have been recorded that are subsequently corrected regarding the laser energy. The final spectrum is composed of an average of two individual spectra showing a reproducible IR signature.
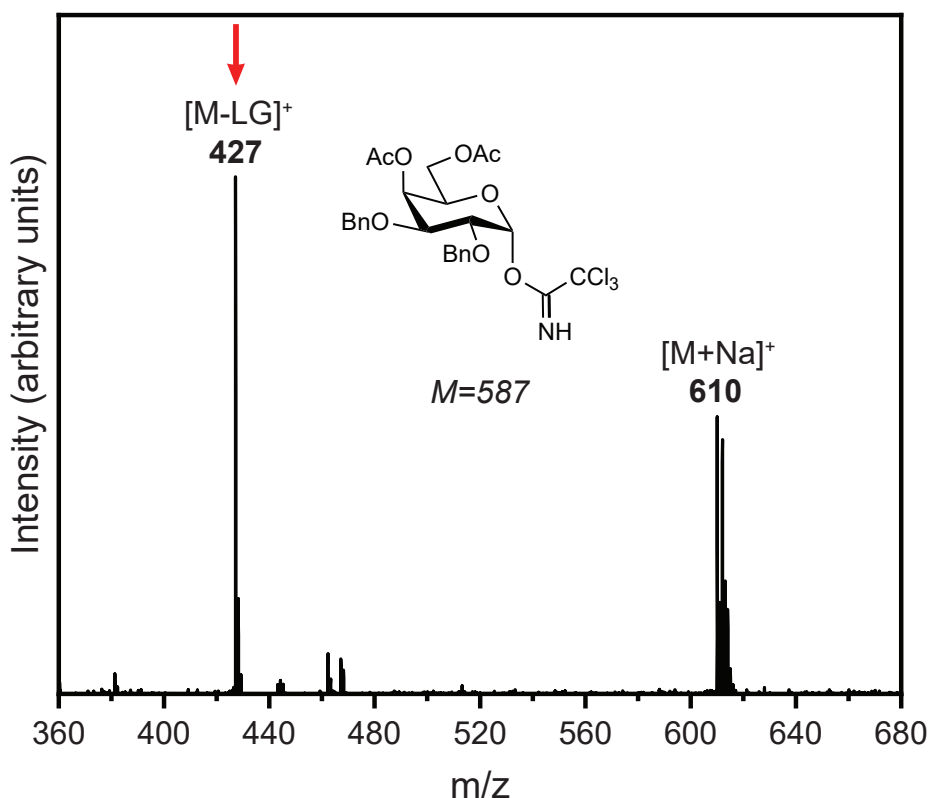

Figure 5.1: Exemplary mass spectrum of 4,6-di-O-acetyl-2,3-di-O-benzyl-α-D-galactopy-ranosyl trichloacetimidate, precursor of **46Ac**. Glycosyl cations ($m/z = 427$) are generated using in-source fragmentation of the respective precursor ions. The signal of the sodiated precursor shows the isotope pattern of Cl_3, which is clearly visible in the spectrum, while the signal of the glycosyl cation mainly consists of one isotope, besides some minor signals due to ^{13}C.

5.2   Stereoselectivity in Glycosynthesis

The α:β-ratios during the glycosynthesis of trisaccharides using the β-thiolate precursors of **46Ac**, **4Ac**, **6Ac** and **Bn** were determined by HAHM *et al.* The trisaccharides were synthesised by automated glycan assembly (AGA) at -20 °C.[106,107] Each building block was coupled to the same disaccharide, which is linked to a solid phase and subsequently

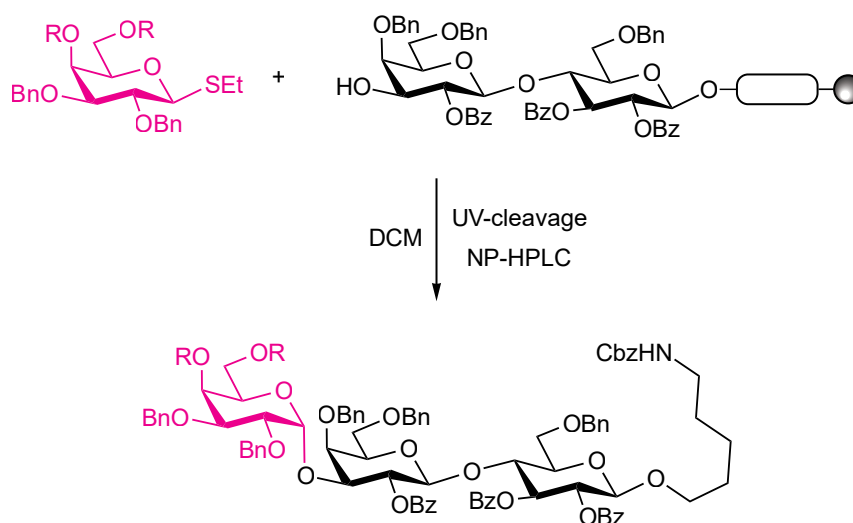released by UV-cleavage (Figure 5.2). The ratios were determined with NP-HPLC and are presented in Table 3.[34]

Figure 5.2: Reaction of a galactosidic donor with a disaccharide linked to a solid phase. The product is separated from the solid phase *via* UV-cleavage and the anomeric ratios are determined by NP-HPLC. Adapted from HAHM *et al.*[34]

Table 3: Impact of remote participation on the stereochemical outcome of glycosidic bond formation in galactose in automated synthesis of trisaccharides. Depending on the protecting groups (Bn = benzyl, Ac = acetyl) different anomeric ratios are obtained. Participating acetyl protecting groups at the C4– and C6-positions exclusively lead to the α-product, which implies that remote participation occurs.[34]

C4-O**R**	C6-O**R**	α:β
Bn	Bn	13.8:1
Bn	Ac	6.4:1
Ac	Bn	26.5:1
Ac	Ac	α only

In all building blocks the formation of α-products is favored. While the usage of **46Ac** and **4Ac** precursors leads almost solely to the α-product, the **Bn** precursor exhibits an unprecedented high α-selectivity, although only bearing non-participating benzyl protecting groups. The lowest α-selectivity is detected for the **6Ac** precursor. It is conjectured that C6-acetylation may even have adverse effects on α-selectivity in glycosylations.[108,109] While this empirical study shows that C4-acetylation seems to play an important role for

providing high α-selectivity, the effects leading to the lower but yet unquestionably high selectivity of the **6Ac** and **Bn** precursors remain concealed.

Reproducibility of ratios is often a problem in batch chemical synthesis, which can, however, be improved with flow chemistry. Flow chemistry provides unique control over the key reaction parameters and offers further benefits such as better mixing and a more efficient heat transfer.[110] Flow chemistry setups have been successfully used in glycosylation reactions.[32] Thus, glycosylation reactions using the α-imidate donors of **46Ac**, **4Ac**, **6Ac** and **Bn** have been carried out by our collaborators of the Seeberger group. The donors have been reacted with 2-propanol in DCM at five different temperatures between -50 and 30 °C. The percentage of α-product is plotted as a function of the temperature for the four donors (Figure 5.3).

The plot reveals that the donors divide into two categories that either yield mainly α-glycosides (**46Ac** and **4Ac**) or little to moderate α-glycosides (**6Ac** and **Bn**). Furthermore, it is observable that the α-ratio rises for each building block with rising temperatures. The consistent difference in stereoselectivity between both groups for all measured temperatures suggests that different reaction intermediates are responsible for the outcome of the reactions. It is known that these type of reactions proceed *via* a S_N2-type mechanism at low temperatures, while the S_N1–S_N2-interface shifts towards S_N1-type reactions with rising temperatures. S_N1-type mechanisms are favored with rising temperatures due to the gain in entropy caused by the dissociative nature of the mechanism.[31] For α-imidate donors, the formation of β-glycosides would thus be preferred at low temperatures, whereas the stereoselectivity at higher temperatures would be determined by the conformation of the glycosyl cation intermediate. If no shielding of the anomeric carbon occurs, the formation of a racemate is expected, while the shielding of the β-side due to remote participation would lead to a high α-selectivity.

For **46Ac** and **4Ac**, low to moderate selectivity towards the α-product is observed at low temperatures, whereas high selectivity is observed at higher temperatures. The moderate α-selectivity can be explained by a mixture between S_N1- and S_N2-type mechanisms at low temperatures that are both selective for the respective anomer, thus leading to moderate selectivity in general. The preference towards the S_N1-side at higher temperatures then leads to α-selectivity. Thus, remote participation seems to play a role in **46Ac** and **4Ac**.

For **6Ac** and **Bn**, high selectivity towards the β-product is observed at low temperatures, while low to moderate selectivity towards the α-product is observed at higher temperatures. Again, a mixture between both mechanisms is presumably seen at low temperatures. However, the S_N1-side of the reaction seems to be not or less selective, as observable for high temperatures, where the formation of a racemate can be observed. It is not certain whether remote participation occurs in **6Ac** and **Bn** or if it acts differently in these building blocks in the condensed phase.
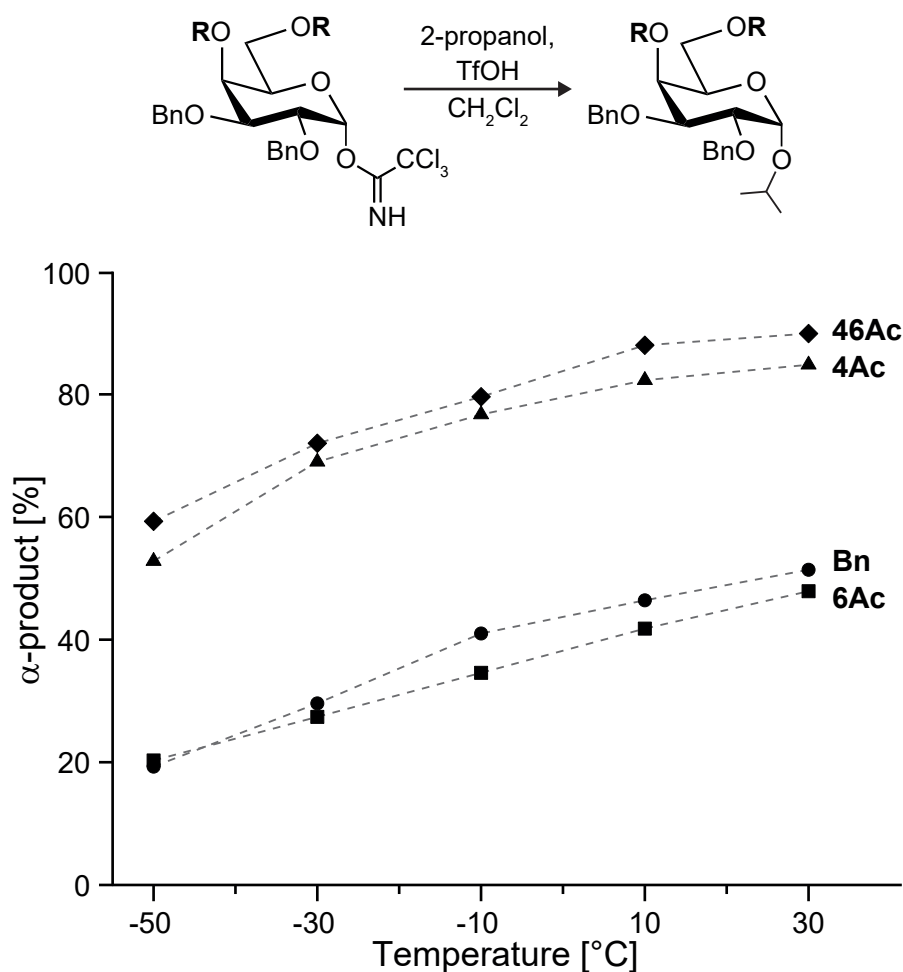
Figure 5.3: Reactions of galactose building blocks (α-imidates) to assess the impact of different protecting group combinations on the stereochemical outcome of the glycosylation reaction. The reactions have been carried out using flow chemistry at various temperatures with 2-propanol as a nucleophile. The relative yield of α-glycosides increases with increasing temperature, indicating that increasing temperatures lead to shift of the mechanistic continuum towards S_N1-type reactions. Building blocks carrying an acetyl group at the C4-position (**46Ac** and **4Ac**) consistently show a higher α-selectivity than the building blocks carrying either only an acetyl group at the C6-position or no acetyl groups (**6Ac** and **Bn**).

Translating the gained knowledge to the data published by HAHM *et al.* explains the observed stereoselectivity rather well. The reactions have been carried out at -20 °C using β-thiolate donors. Presumably a mixture between S_N1- and S_N2-type reactions is observed at this temperature. However, presuming that remote participation takes place, both mechanisms would lead to α-glycosides, explaining the generally high α-selectivity observed

for all building blocks. The decreased selectivity for **6Ac** and **Bn** can be explained by the formation of non or less α-selective glycosyl cations in the S_N1-mechanism. However, the stereochemical outcome of glycosylation reactions with a S_N2-mechanism is not solely dependent on the stereoconfiguration of the leaving group. It is presumed that the strength of the nucleophile plays a big role in glycosynthesis. HAHM *et al.* utilized a disaccharide, which is a rather weak nucleophile, while isopropanol, used by our collaborators, is a strong nucleophile.

5.3   Establishing Links between Experiment and Theory

In the following section, the experimental IR spectra of each glycosyl cation are presented in combination with the computed vibrational spectra of the low-energy conformers of each subgroup. The direct comparison of experimental and theoretical spectra allows to establish structural matches to elucidate the gas-phase structure of each glycosyl cation and draw conclusions on the stereochemistry observed in glycosynthesis. Glycosyl cations are the key intermediates in the dissociative S_N1-mechanism. Their investigation allows to elucidate the S_N1-side of the S_N1–S_N2-continuum.

46Ac

The **46Ac** building block exhibits the highest selectivity for α-glycosidic bond formation. The infrared spectrum of the corresponding glycosyl cation (Figure 5.4) displays six well-resolved absorption bands. The signals at 1467 and 1564 cm^{-1} are clearly situated in the regions of the symmetric and the antisymmetric dioxolenium-stretches and the absorption bands at 1217 and 1771 cm^{-1} can be associated with the C–O- and the C=O-stretch of the unbound acetyl PG, indicating that dioxolenium-type structures are adopted by the **46Ac** glycosyl cation. The calculated spectrum of the low-energy structure **A**, that adopts a $^{1}S_5$ ring pucker and possesses a C4-acetyl dioxolenium-motif, matches the experimental spectrum. The characteristic bands associated with the dioxolenium-motif are predicted at 1465 and 1568 cm^{-1} and the bands predicted at 1207 and 1779 cm^{-1} of the C–O- and the C=O-stretch of the non-interacting C6-acetyl group, respectively, are in agreement with the experiment.

In the alternative structure **B**, adopting a $^{1}C_4$ pucker, the dioxolenium-motif originates from a covalent bond between the C6-acetyl protecting group and the anomeric carbon. Its free energy is, however, predicted to be 20 kJ mol^{-1} higher than the energy of structure **A**. Furthermore, the theoretical and experimental vibrational spectra agree less well overall. An oxocarbenium-type structure **C** with two unbound acetyl groups can be eliminated from consideration with a large free energy difference of 59 kJ mol^{-1}. Its spectrum exhibits
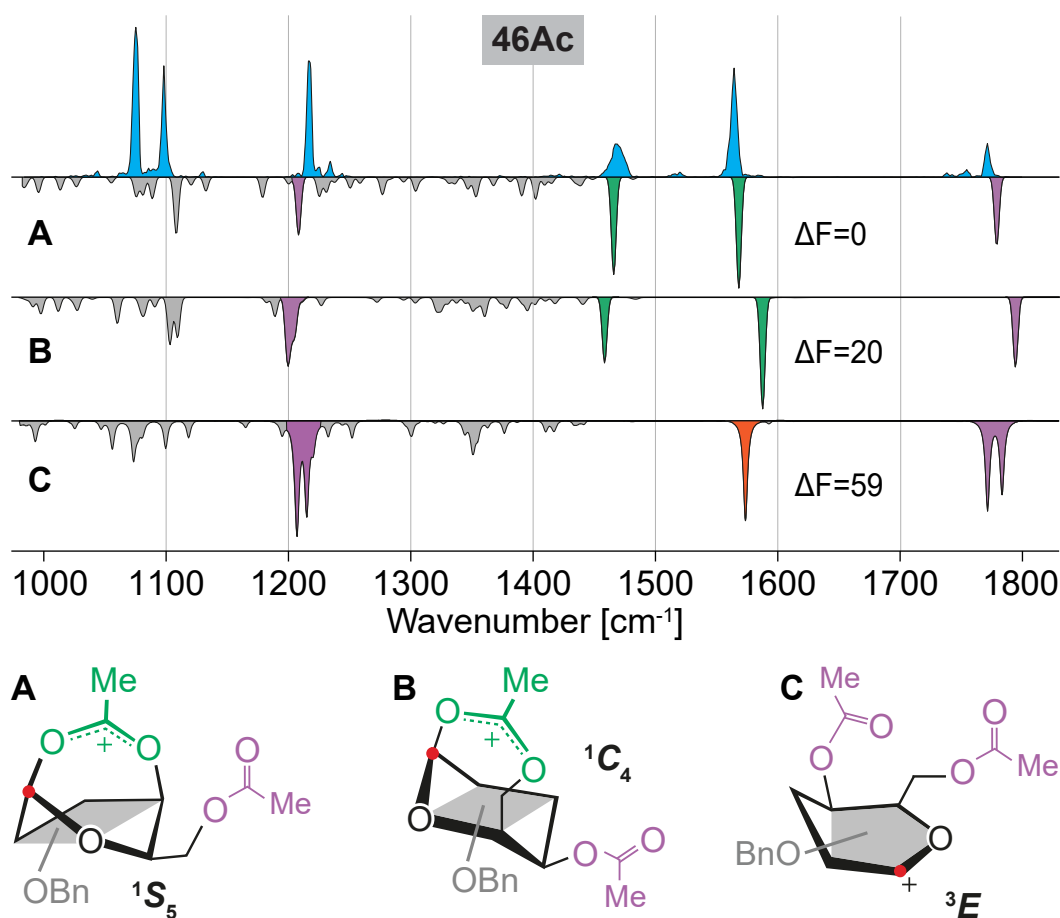
Figure 5.4: Infrared spectra of **46Ac** glycosyl cations. The blue trace represents the experimental IR spectrum, while the grey inverted traces are calculated IR spectra corresponding to the low-energy conformers shown below at the bottom in a simplified representation. The free energies are given in kJ mol^{-1}. The highlighted absorption bands in the theoretical spectra indicate vibrations from free acetyl groups (purple), participating acetyl groups in dioxolenium-type structures (green) and from the C1=O5-stretch in oxocarbenium-type structures (orange). For this building block, the majority of ions adopts dioxolenium-type structures that exhibit a covalent bond between the C4-acetyl group and the anomeric carbon (highlighted with a red dot).

a strong oxocarbenium-stretch and double peaks for the C–O- and the C=O-stretches, resulting from the second unbound acetyl group.

In summary, the IR signature of the **46Ac** galactosyl donor provides direct evidence of C4-acetyl participation through a dioxolenium-type intermediate that effectively shields the β-side from nucleophilic attacks leading to α-selectivity. Participation of the C6-acetyl protecting group is energetically unfavored.

4Ac

The comparable stereoselectivity of **4Ac** suggests that it adopts a similar conformation as **46Ac**. The IR spectrum of **4Ac** displays indeed characteristic bands between 1450 and 1600 cm^{-1}, similar to those observed for **46Ac**, being diagnostic for dioxolenium-type structures (Figure 5.5). Only a small signal can be observed in the C=O-stretch region. The calculated vibrations at 1465 and 1568 cm^{-1} of the dioxolenium-type structure **D** featuring a $^1S_5$ ring pucker match well with the experimental spectrum. The presence of multiple signals in this region indicates that structurally very similar dioxolenium-type intermediates coexist. Structure **D'**, only differing in the spatial position of the C6-benzyl group from structure **D**, exhibits a very similar IR spectrum. Its free energy is 3 kJ mol^{-1} relative to the energy of **D**. The small change in structure leads to a small shift of the absorption bands of the dioxolenium-motif.

The additional low intensity absorption bands at 1220 and 1780 cm^{-1} indicate that a small fraction of glycosyl cations adopts other structures. The oxonium-type structure **E**, yielded by the conformational search, surprisingly exhibits an interaction between the oxygen atom of the C6-benzyl protecting group and the anomeric carbon, even though this protecting group is considered to be *non-participating*. This structure, which adopts a distorted $^1C_4$ ring pucker, possesses a free energy of 16 kJ mol^{-1} above the energy of structure **D** and the vibrations of the free acetyl group at 1227 and 1763 cm^{-1} agree with the experimentally resolved bands.

The sampled low-energy oxocarbenium-type structure **F** exhibits a much higher free energy of 50 kJ mol^{-1} relative to structure **D** and the predicted vibrations agree less with the experiment and can thus be ruled out from consideration. The spectroscopic data provides evidence that the **4Ac** glycosyl cation mainly adopts an α-selective dioxolenium-type structure, similar to that observed for **46Ac**. The influence of the formation of a small fraction of oxonium-type structure **E** on the stereochemistry remains elusive.

6Ac

Contrary to **46Ac** and **4Ac**, the glycosyl donors **6Ac** and **Bn** consistently show low to medium abundance of α-products, suggesting that the underlying intermediates are structurally different. The IR spectrum of the glycosyl cation **6Ac** displays various absorption bands in the region between 1300 and 1500 cm^{-1}, being not diagnostic for the predicted features, while the bands at 1229 and 1756 cm^{-1} are diagnostic for C–O- and C=O-stretches of unbound acetyl groups, respectively (Figure 5.6). A minor signal is observed at 1564 cm^{-1} in the oxocarbenium-stretch region. The presence of multiple high-intensity bands above 1200 cm^{-1} indicates the coexistence of multiple structures in the ion trap. The dioxolenium-type structure **G** with a $^1C_4$ ring pucker exhibits the lowest
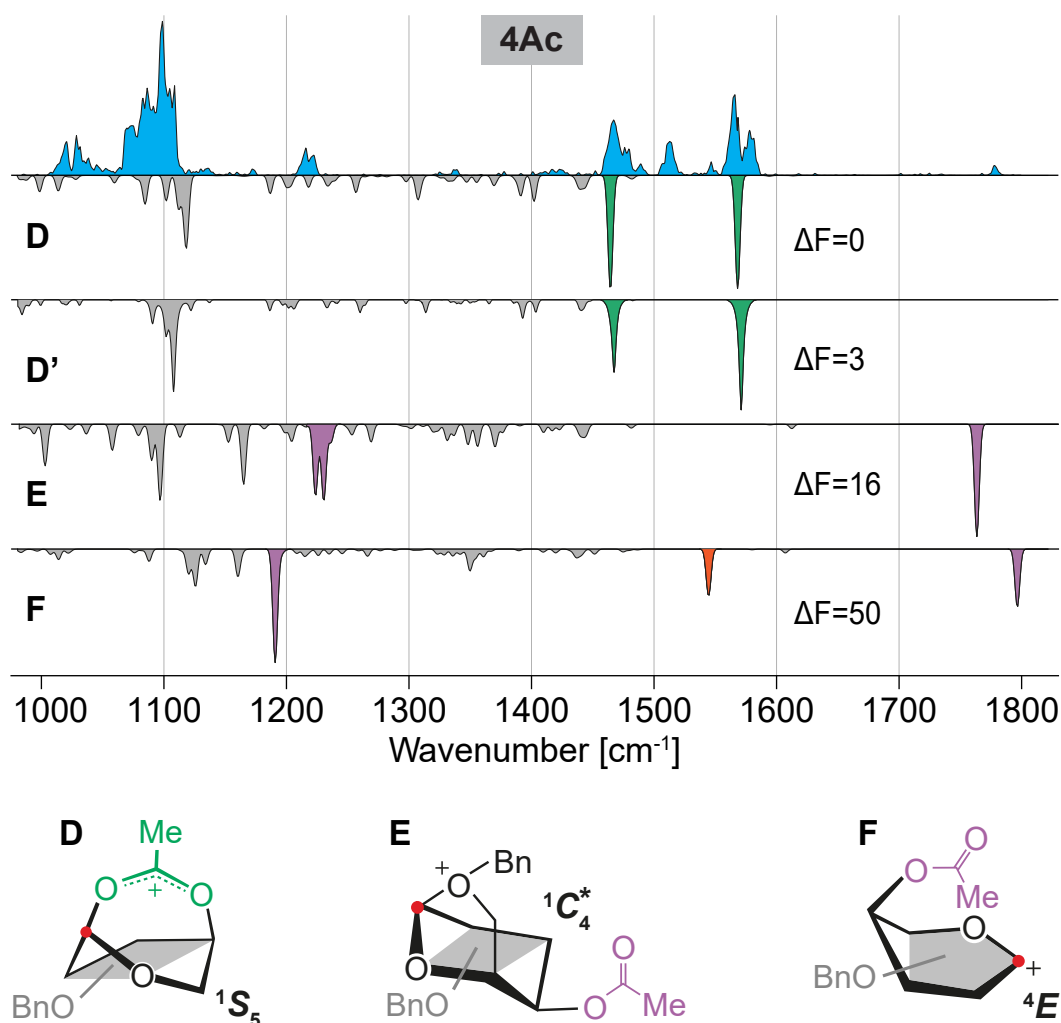
Figure 5.5: Infrared spectra of **4Ac** glycosyl cations. The blue trace represents the experimental IR spectrum, while the grey inverted traces are calculated IR spectra corresponding to the low-energy conformers shown below at the bottom in a simplified representation. The free energies are given in kJ mol^{-1}. The highlighted absorption bands in the theoretical spectra indicate vibrations from free acetyl groups (purple), participating acetyl groups in dioxolenium-type structures (green) and from the C1=O5-stretch in oxocarbenium-type structures (orange). Multiple low-energy dioxolenium-type conformers as **D** and **D'** may explain the occurence of multiple absorption bands in the region of the symmetric and antisymmetric dioxolenium-stretch. For this building block, the majority of ions adopts dioxolenium-type structures that exhibit a covalent bond between the C4-acetyl group and the anomeric carbon. Structure **E** shows a distorted $^1C_4$ ring pucker, indicated by an asterisk (*).

free energy and is predicted to have two characteristic modes at 1457 and 1593 cm^{-1}, which match poorly with the experimental spectrum. Again, structures exhibiting *non-classical*

remote participation can be observed. Oxonium-type structures by C4-benzyl participation can be witnessed in conformer **H**, which adopts a $^{1,4}B$ ring pucker. The predicted vibrations of the free acetyl group at 1220 and 1762 cm^{-1} match with two prominent bands in the experimental spectrum. Its free energy is 16 kJ mol^{-1} higher than the relative energy of structure **G**.
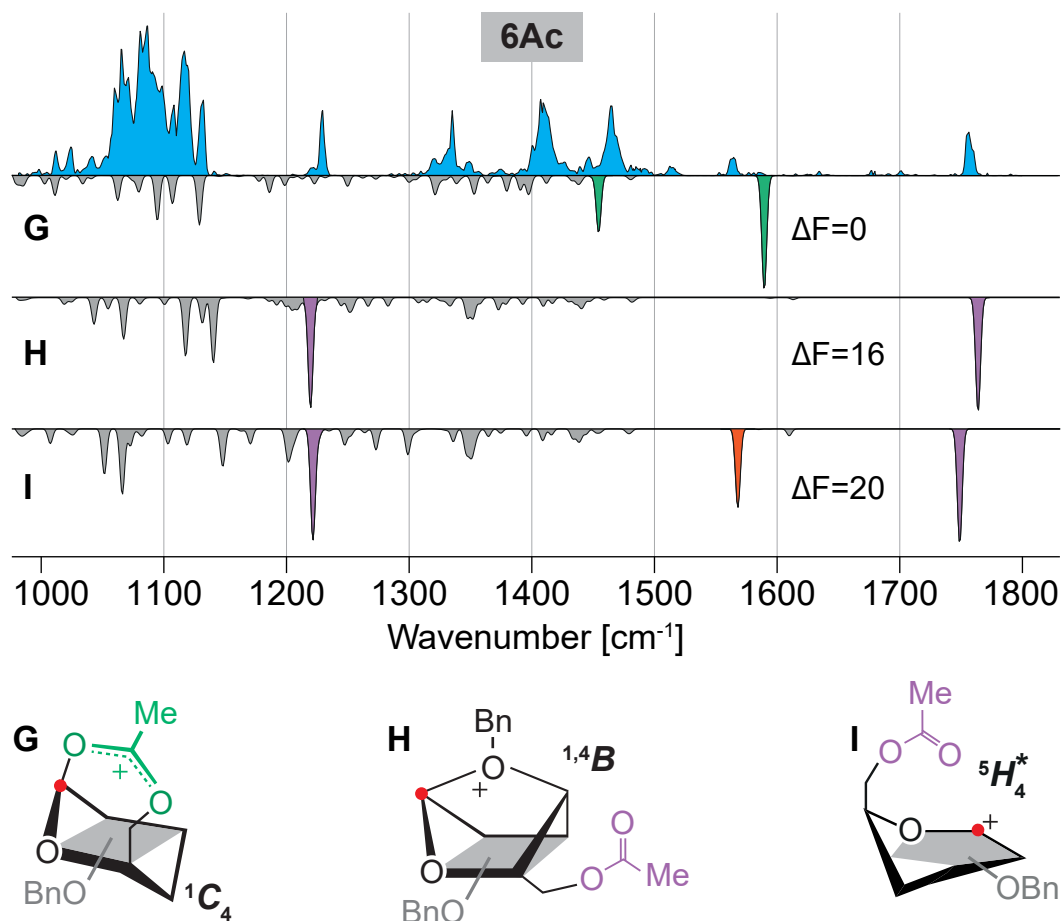

Figure 5.6: Infrared spectra of **6Ac** glycosyl cations. The blue trace represents the experimental IR spectrum, while the grey inverted traces are calculated IR spectra corresponding to the low-energy conformers shown below at the bottom in a simplified representation. The free energies are given in kJ mol^{-1}. The highlighted absorption bands in the theoretical spectra indicate vibrations from free acetyl groups (purple), participating acetyl groups in dioxolenium-type structures (green) and from the C1=O5-stretch in oxocarbenium-type structures (orange). Interestingly, *non-classical remote participation* of benzyl groups at the C4-position leads to oxonium-type intermediates. In this building block, the majority of ions either adopt oxonium- or oxocarbenium-type glycosyl cations. Structure **I** shows a distorted $^5H_4$ ring pucker, indicated by an asterisk (*).

The oxocarbenium-type structure **I** is with a free energy of $20\,\mathrm{kJ\,mol^{-1}}$ very similar to structure **H**. It adopts a heavily distorted ring pucker, which resembles the closest to $^5H_4$. The predicted bands at 1222 and $1749\,\mathrm{cm^{-1}}$ of the free acetyl group and the vibration of the oxocarbenium-stretch at $1567\,\mathrm{cm^{-1}}$ of structure **I** match the experimental spectrum well. The oxocarbenium-type structure is stabilized by a long-range interaction $(2.8\,\text{Å})$ between the anomeric carbon and the carbonyl oxygen of the C6-acetyl protecting group. Furthermore, the oxocarbenium-type structure **I** is also stabilized by cation–π-interactions between the anomeric carbon and the C2-benzyl protecting group. Even though structures **H** and **I** are energetically less favored than structure **G**, the comparison between theory and experiment shows clearly that either both **H** and **I** or one of these glycosyl cations exist as the dominant structures, while structure **G** is not formed. These findings are in line with the low to medium abundance of α-products in the test reactions, suggesting that the oxocarbenium-type structure **I** with reduced α-selectivity, compared to dioxolenium-type structures, is the main intermediate in this reaction. The oxonium-type structure **H**, although being energetically favoured over **I**, would exhibit a high α-selectivity, not observed in the test reactions and is, thus, not the main intermediate. The origin of the high intensity absorption bands between 1300 and $1500\,\mathrm{cm^{-1}}$ remains elusive. Such signals in this region are rather unexpected for the analyzed type of ions, suggesting that intramolecular rearrangement occurs to a substantial degree.

Bn

The fully benzylated building block **Bn** shows a comparably low abundance of α-glycosides as **6Ac** in the test reactions. The experimental IR spectrum of the corresponding glycosyl cation (Figure 5.7) does not show any significant absorption bands above $1200\,\mathrm{cm^{-1}}$. Due to the conjectured non-participating nature of benzyl protecting groups, a large C1=O5-stretch would be expected at around $1560\,\mathrm{cm^{-1}}$. However, as previously shown, benzyl groups at the C4- and C6-positions can interact with the positive charge at the anomeric carbon and form a covalent bond to yield oxonium-type structures, completely deactivating the expected oxocarbenium-stretch. Conformers with a participating C6-benzyl group have been sampled, while C4-participating conformers could only be generated *via* chemical intuition. Spectra of both type of conformers are matching well with the experimental spectrum, showing no major signals above $1200\,\mathrm{cm^{-1}}$. The free energy of conformer **J** $(^1C_4)$ exhibiting C6-benzyl participation is, however, $51\,\mathrm{kJ\,mol^{-1}}$ lower than the free energy of conformer **H** $(^{1,4}B)$, exhibiting C4-benzyl participation.

The spectrum of the sampled oxocarbenium-type structure **L** (^4E) matches overall less with the experimental spectrum and it exhibits a dominant oxocarbenium-stretch at $1561\,\mathrm{cm^{-1}}$ that is absent in the experimental spectrum. Considering that its relative
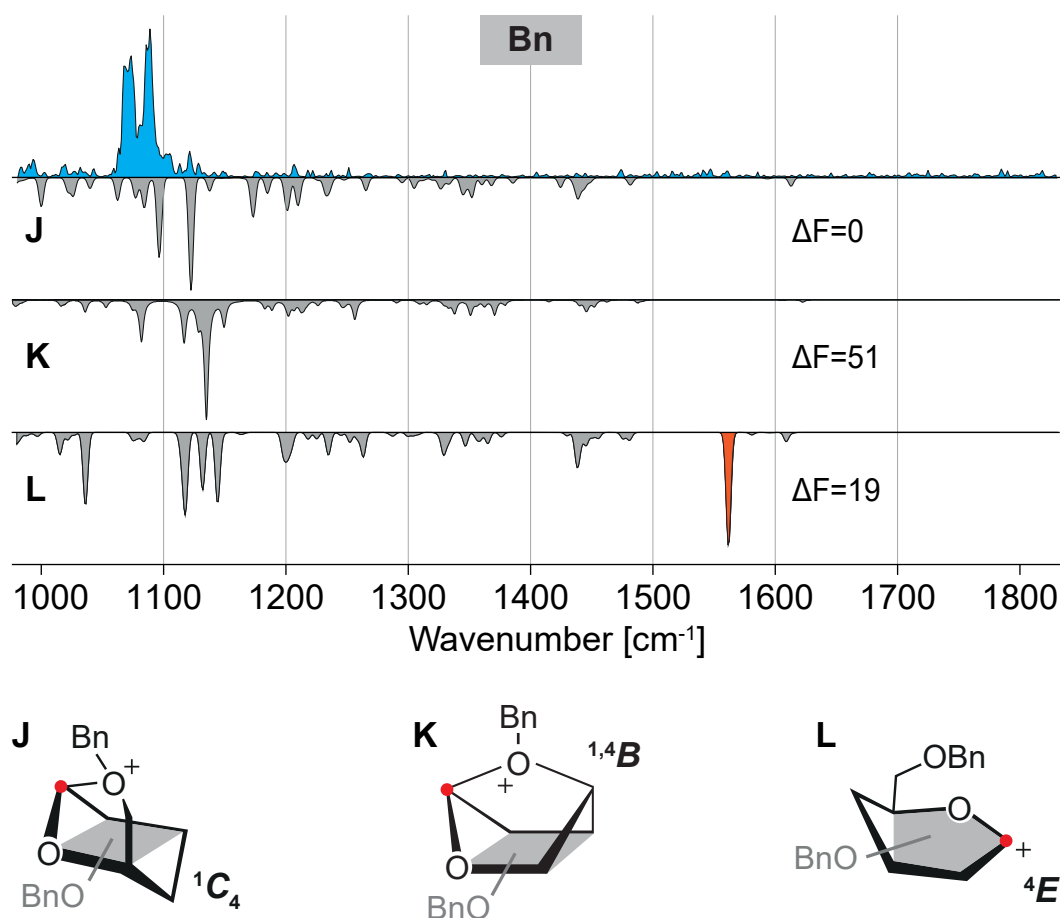
Figure 5.7: Infrared spectra of **Bn** glycosyl cations. The blue trace represents the experimental IR spectrum, while the grey inverted traces are calculated IR spectra corresponding to the low-energy conformers shown below at the bottom in a simplified representation. The free energies are given in $kJ \, mol^{-1}$. The highlighted absorption band indicates vibrations from the C1=O5-stretch in oxocarbenium-type structures (orange). Interestingly, *non-classical remote participation* of benzyl groups at the C6-position leads to oxonium-type intermediates, completely disabling the oxocarbenium-stretch in the experimental spectrum, suggesting that all formed glycosyl cations adopt an oxonium-type conformation.

free energy is $19 \, kJ \, mol^{-1}$ above the energy of conformer **J**, oxocarbenium-type structures are most likely not formed by **Bn**. The dominant structure is therefore both in terms of energetics and spectral match the oxonium-type structure **J**.

The sampled structures suggest that shielding of the β-side *via* oxonium-type structures should lead to an increased α-selectivity. Generally, C6-benzyl participation leads to a large energetic stabilization of $19–34 \, kJ \, mol^{-1}$ relative to oxocarbenium-type structures, whereas C4-benzyl participation does either lead to no or only a minor stabilization of $4 \, kJ \, mol^{-1}$. Experimentally, **Bn** exhibits a slightly higher α-selectivity than **6Ac**, which suggests that

benzyl participation leads to β-shielding, which is less efficient than in building blocks that form C4-acetyl dioxolenium-type structures. However, it has been reported that C6-acetylation on galactosyl donors may even have an adverse effect on α-selectivity.[108] Even though oxonium-type structure **J**, potentially exhibiting a high α-selectivity, is formed in the gas-phase, its formation in the condensed phase may be hindered by interactions with the solvent or the activator and thus no significant increase in α-selectivity can be observed. A more in-depth study on this phenomenon needs to be carried out to exactly determine the influence of *non-classical remote participation* of benzyl protecting groups on the stereochemistry in glycosylation reactions.

5.4 The Influence of the Leaving Group on the Conformation of the Glycosyl Cation

In a further step, the influence of the leaving group as well as its configuration on the conformation of the generated glycosyl cations has been assessed. SCHINDLER *et al.* and GRAY *et al.* have previously shown that B- and C-fragments (according to the DOMON and COSTELLO nomenclature[111], shown in Figure 5.8) of both anomers of disaccharides and oligosaccharide standards generated *via* CID exhibit a distinct IR signature.[112,113] This phenomenon is known as *anomeric memory.* The cryogenic vibrational spectra of the glycosyl cations of β-thiolate donors of **46Ac**, **4Ac** and **6Ac** and of the β-imidate donor of **46Ac** have been recorded and compared to the spectra of the α-imidate donors to assess if this phenomenon also occurs in the building blocks examined in this thesis using in-source fragmentation (Figures 5.9 and 5.10).
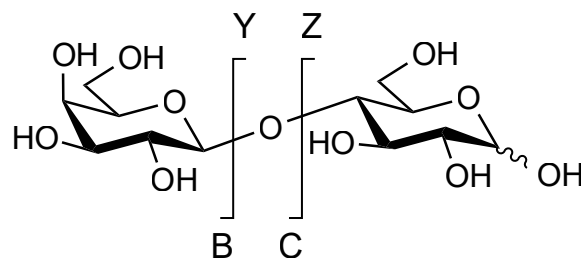

Figure 5.8: Nomenclature of glycan MS fragments, after DOMON and COSTELLO, exemplarily shown for the disaccharide Gal-β-1,4–Glc.[111]

It is clearly visible in Figure 5.9 that the vibrational spectra of the **46Ac** glycosyl cations generated *via* the α- and the β-imidate precursors are almost identical. The formed glycosyl cations are similar to B-fragments of disaccharides. This is clear evidence contradicting the existence of the *anomeric memory* in the analyzed molecules using the described experimental setup. Similarly, as shown in Figures 5.9 and 5.10, the spectra generated by the β-thiolate precursors of **46Ac**, **4Ac** and **6Ac** display the same bands as

the respective α-imidate precursors, demonstrating that the leaving group does not influence the conformation of the generated glycosyl cation. This is in line with empirical data recorded by CHATTERJEE *et al.*[32] The stereochemical outcome of glycosylation reactions of fully benzylated glucose donors with the nucleophile 2-propanol and the activator TfOH have been systematically analyzed using α- and β-imidate as well as β-thiolate donors. In all cases, the observed α:β-ratios were the same, suggesting that these parameters do not influence the stereochemical outcome in glycosylation reactions. Comparably, no influence on the conformation of the glycosyl cation has been observed upon changing these parameters.

However, although displaying the same bands, the vibrational spectra of the β-thiolate precursors show less-resolved bands than the α- or β-imidate precursors and appear to be more noisy. This is mainly due to the low ionization efficiency of thiolate precursors, leading to a less populated ion trap, decreasing the overall signal of the recorded IR spectra. The imidate precursors are ionized more efficiently, leading to an increased signal with less noise.
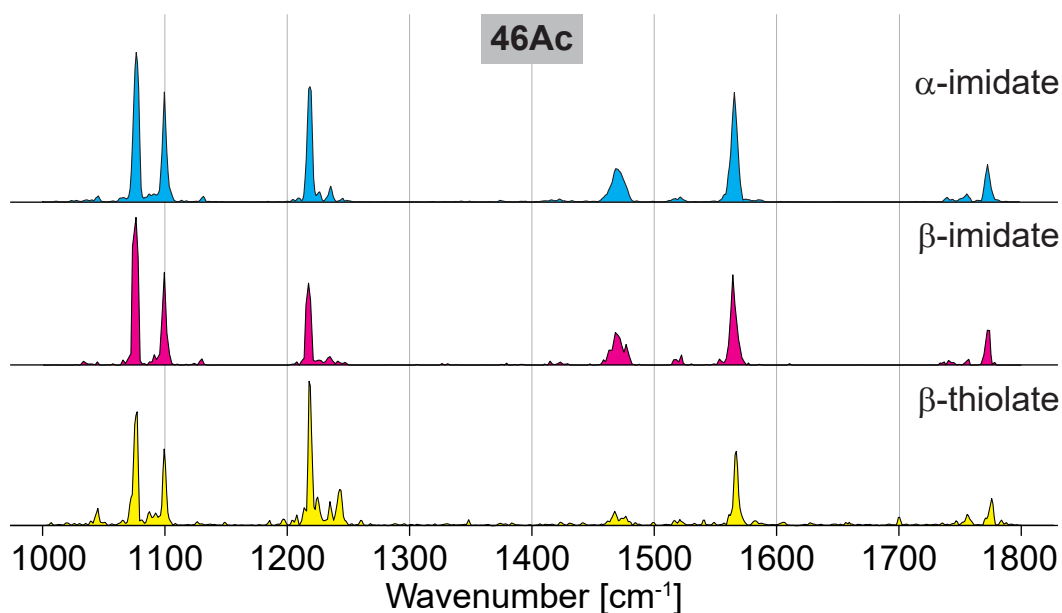

Figure 5.9: Comparison of the experimental spectra of **46Ac** glycosyl cations generated with precursors bearing α-imidate (blue trace), β-imidate (magenta trace) or β-thiolate (yellow trace) leaving groups. The similarity of the traces indicates that the conformation of the probed glycosyl cations is independent from the nature of the leaving group and its configuration.
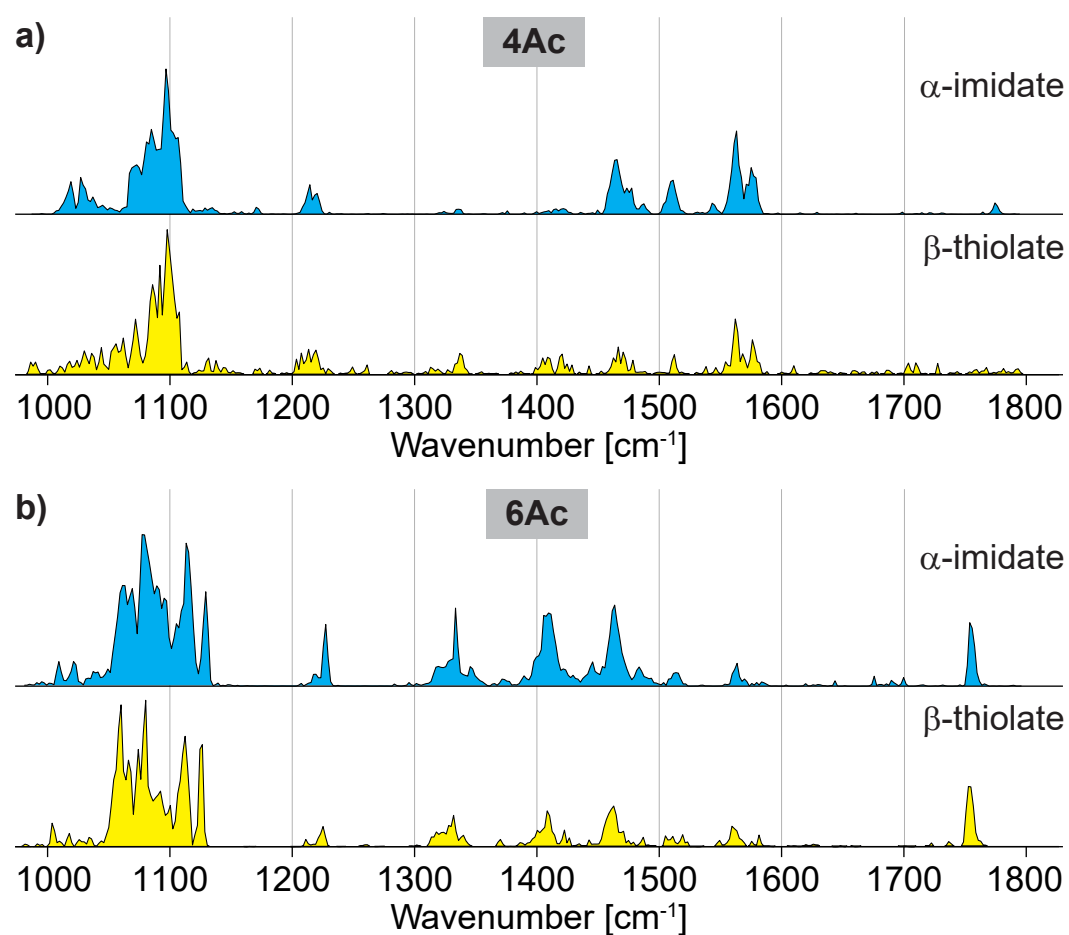
Figure 5.10: Comparison of the experimental spectra of (a) **4Ac** and (b) **6Ac** glycosyl cations generated with precursors bearing α-imidate (blue trace) or β-thiolate (yellow trace) leaving groups. The similarity of both traces indicates that the conformation of the probed glycosyl cations is independent from the nature of the leaving group.

6 Conclusions and Outlook

A combined experimental and theoretical study of galactose building blocks bearing acetyl and benzyl protecting groups in four distinct combinations has systematically been performed in this thesis. Using cryogenic vibrational spectroscopy in helium droplets, the IR signatures of each glycosyl cation have been recorded. In combination with the narrow bandwidth IR free-electron laser, this setup has yielded well-resolved IR spectra with distinct features for each glycosyl cation. In a further step, the conformational space of the glycosyl cations has been sampled using a genetic algorithm, in which each generated structure has been optimized using density functional theory methods. The geometries of low-energy structures featuring distinct conformations have been reoptimized using hybrid DFT methods and the corresponding IR spectra have been calculated. The energies of the reoptimized structures have been recomputed at RI-MP2/CBS level of theory.

The comparison of theoretical and experimental spectra has yielded excellent structural matches. For **46Ac**, **4Ac** and **Bn** the theoretical IR spectra of the lowest-energy structure is in agreement with the experimental spectra, while the best match for **6Ac** is provided by a structure that possesses a free energy that is slightly elevated compared to the lowest-energy structure. The recorded spectra clearly show that remote participation exists in galactose building blocks routinely used in oligosaccharide synthesis. α-selectivity is achieved *via* C4-acetylation, promoting the formation of the energetically preferred dioxolenium-type ions with a covalent bond between the carbonyl oxygen and the anomeric carbon, as shown for building blocks **4Ac** and **46Ac**.

Although such structures are predicted to be energetically more stable than oxocarbenium- or oxonium-type structures in **6Ac**, no experimental proof for the existence of C6-acetyl dioxolenium-type structures has been found, suggesting that C6-acetyl remote participation does not play a major role for this building block. Oxocarbenium- and oxonium-type structures are energetically very close to each other and the spectra suggest that both types of structures coexist as the dominant structures. Oxocarbenium-type structures are stabilized by both long-range interactions between the acetyl protecting group and the anomeric carbon and cation–π-interactions between the anomeric carbon and a benzyl protecting group, whereas oxonium-type structures exhibit non-classical remote participation of benzyl protecting groups, where a covalent bond is formed between the oxygen of the benzyl ether and the anomeric carbon. The latter is particularly surprising, as it is hardly described in literature.

In **Bn**, a building block only carrying non-participating benzyl protecting groups, oxonium-type structures are clearly the dominant structures, shown by the deactivation of the oxocarbenium-stretch by the C6-benzyl group, explaining the increased α-selectivity in **Bn** compared to **6Ac**. These spectroscopic observations, which are in line with empirical

data during glycosylation reactions, provide a structural basis for classical- and non-classical remote participation that is essential during the stereoselective formation of α-glycosidic bonds. This study shows that by removing the influence of solvent and counter ions, the clean room environment of a mass spectrometer enables to study the intrinsic stereoselectivity of glycosyl cations as a basis to design glycosyl donors by tuning the properties of protecting groups. Determining the conformation of the glycosyl cation leads to an understanding of the S_N1-side of the mechanistic continuum in glycosylation reactions. Thus, in this thesis, the first experimental proof for remote participation in building blocks commonly used in glycosynthesis and for non-classical remote participation have been provided. The collected data can be used by algorithms based on artificial intelligence to predict the best reaction conditions for performing glycosylation reactions with a defined stereochemical outcome and will ultimately lead to a major improvement in glycosynthesis.

In the future, further protecting group combinations involving for example pivaloyl (Piv), *tert*-butyloxycarbonyl (Boc) and fluorenylmethyloxycarbonyl (Fmoc) protecting groups will be thoroughly analyzed using the same setup and computational multistep procedure as described in this thesis. Thereby, the influence of these protecting groups on the glycosyl cation will be determined to improve the knowledge about these types of reactions. Furthermore, the newly discovered non-classical remote participation will be investigated by selectively tuning the electron density on the C6-benzyl ether oxygen in **Bn**, by introducing electron donating and electron withdrawing groups in this building block.

Furthermore, the collision cell of the utilized instrument will be activated in the future, to further investigate the anomeric memory. Moreover, either the ion guide region or the collision cell will be modified to allow ion-molecule reactions (IMRs). Through the thereby obtained setup the interactions of glycosyl cations with solvent and nucleophile molecules can be analyzed to further investigate and fully understand the mechanism of glycosylation reactions.

References

[1] A. Varki, *Glycobiology* **1993**, *3*, 97–130.

[2] F. W. Lichtenthaler, *Ullmann's Encyclopedia of Industrial Chemistry, 40 Volume Set*, Wiley-VCH, **2011**.

[3] K. Bethke, S. Palantöken, V. Andrei, M. Roß, V. S. Raghuwanshi, F. Kettemann, K. Greis, T. T. K. Ingber, J. B. Stückrath, S. Valiyaveettil, K. Rademann, *Adv. Funct. Mater.* **2018**, *28*, 1800409.

[4] G. Annadurai, R. Juang, D. Lee, *J. Hazard. Mater.* **2002**, *92*, 263–274.

[5] D. Rana, B. Scheier, R. M. Narbaitz, T. Matsuura, S. Tabe, S. Y. Jasim, K. C. Khulbe, *J. Memb. Sci.* **2012**, *409-410*, 346–354.

[6] D. Setyono, S. Valiyaveettil, *J. Hazard. Mater.* **2016**, *302*, 120–128.

[7] L. Schofield, M. C. Hewitt, K. Evans, M.-A. Siomos, P. H. Seeberger, *Nature* **2002**, *418*, 785–789.

[8] S. I. van Kasteren, S. J. Campbell, S. Serres, D. C. Anthony, N. R. Sibson, B. G. Davis, *Proc. Nat. Acad. Sci.* **2008**, *106*, 18–23.

[9] S. B. Dulaney, X. Huang in *Advances in Carbohydrate Chemistry and Biochemistry*, Elsevier, **2012**, pp. 95–136.

[10] R. F. Service, *Science* **2012**, *338*, 321–323.

[11] J. E. Turnbull, R. A. Field, *Nat. Chem. Biol.* **2007**, *3*, 74–77.

[12] B. Domon, *Science* **2006**, *312*, 212–217.

[13] J. C. Venter *et al.*, *Science* **2001**, *291*, 1304–1351.

[14] S. S. C. Chu, G. A. Jeffrey, *Acta Cryst. B* **1968**, *24*, 830–838.

[15] C. A. Bush, M. Martin-Pastor, A. Imbery, *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 269–293.

[16] J. Hofmann, H. S. Hahm, P. H. Seeberger, K. Pagel, *Nature* **2015**, *526*, 241–244.

[17] J. Hofmann, K. Pagel, *Angew. Chem. Int. Ed.* **2017**, *56*, 8342–8349.

[18] C. Manz, M. Grabarics, F. Hoberg, M. Pugini, A. Stuckmann, W. B. Struwe, K. Pagel, *Analyst* **2019**, DOI 10.1039/c9an00937j.

[19] E. Mucha, A. I. González Flórez, M. Marianski, D. A. Thomas, W. Hoffmann, W. B. Struwe, H. S. Hahm, S. Gewinner, W. Schöllkopf, P. H. Seeberger, G. von Helden, K. Pagel, *Angew. Chem. Int. Ed.* **2017**, *56*, 11248–11251.

[20] E. Mucha, M. Lettow, M. Marianski, D. A. Thomas, W. B. Struwe, D. J. Harvey, G. Meijer, P. H. Seeberger, G. von Helden, K. Pagel, *Angew. Chem. Int. Ed.* **2018**, *57*, 7440–7443.

[21] C. Manz, K. Pagel, *Curr. Opin. Chem. Biol.* **2018**, *42*, 16–24.

[22] E. Mucha, A. Stuckmann, M. Marianski, W. B. Struwe, G. Meijer, K. Pagel, *Chem. Sci.* **2019**, *10*, 1272–1284.

[23] C. Masellis, N. Khanal, M. Z. Kamrath, D. E. Clemmer, T. R. Rizzo, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2217–2222.

[24] N. Khanal, C. Masellis, M. Z. Kamrath, D. E. Clemmer, T. R. Rizzo, *Analyst* **2018**, *143*, 1846–1852.

[25] M. Z. Kamrath, T. R. Rizzo, *Acc. Chem. Res.* **2018**, *51*, 1487–1495.

[26] S. Warnke, A. B. Faleh, R. P. Pellegrinelli, N. Yalovenko, T. R. Rizzo, *Faraday Discuss.* **2019**, *217*, 114–125.

[27] A. B. Faleh, S. Warnke, T. R. Rizzo, *Anal. Chem.* **2019**, *91*, 4876–4882.

[28] E. Mucha, M. Marianski, F.-F. Xu, D. A. Thomas, G. Meijer, G. von Helden, P. H. Seeberger, K. Pagel, *Nat. Commun.* **2018**, *9*, 4174.

[29] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1893**, *26*, 2400–2412.

[30] D. Crich, *Acc. Chem. Res.* **2010**, *43*, 1144–1153.

[31] P. O. Adero, H. Amarasekara, P. Wen, L. Bohé, D. Crich, *Chem. Rev.* **2018**, *118*, 8242–8284.

[32] S. Chatterjee, S. Moon, F. Hentschel, K. Gilmore, P. H. Seeberger, *J. Am. Chem. Soc.* **2018**, *140*, 11942–11953.

[33] S. van der Vorm, T. Hansen, J. M. A. van Hengst, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Chem. Soc. Rev.* **2019**, DOI `10.1039/c8cs00369f`.

[34] H. S. Hahm, M. Hurevich, P. H. Seeberger, *Nat. Commun.* **2016**, *7*, 12482.

[35] A. V. Demchenko, E. Rousson, G.-J. Boons, *Tetrahedron Lett.* **1999**, *40*, 6523–6526.

[36] H. Elferink, R. A. Mensink, P. B. White, T. J. Boltje, *Angew. Chem. Int. Ed.* **2016**, *55*, 11217–11220.

[37] T. Hosoya, T. Takano, P. Kosma, T. Rosenau, *J. Org. Chem.* **2014**, *79*, 7889–7894.

[38] T. G. Frihed, M. Bols, C. M. Pedersen, *Chem. Rev.* **2015**, *115*, 4963–5013.

[39] S. Warnke, J. Seo, J. Boschmans, F. Sobott, J. H. Scrivens, C. Bleiholder, M. T. Bowers, S. Gewinner, W. Schöllkopf, K. Pagel, G. von Helden, *J. Am. Chem. Soc.* **2015**, *137*, 4236–4242.

[40] H. Elferink, M. E. Severijnen, J. Martens, R. A. Mensink, G. Berden, J. Oomens, F. P. J. T. Rutjes, A. M. Rijs, T. J. Boltje, *J. Am. Chem. Soc.* **2018**, *140*, 6034–6038.

[41] H. Elferink, R. A. Mensink, W. W. A. Castelijns, O. Jansen, J. P. J. Bruekers, J. Martens, J. Oomens, A. M. Rijs, T. J. Boltje, *Angew. Chem. Int. Ed.* **2019**, *58*, 8746–8751.

[42] N. C. Polfer, J. J. Valle, D. T. Moore, J. Oomens, J. R. Eyler, B. Bendiak, *Anal. Chem.* **2006**, *78*, 670–679.

[43] A. Martin, A. Arda, J. Désiré, A. Martin-Mingot, N. Probst, P. Sinaÿ, J. Jiménez-Barbero, S. Thibaudeau, Y. Blériot, *Nat. Chem.* **2015**, *8*, 186–191.

[44] T. Hansen, L. Lebedel, W. A. Remmerswaal, S. van der Vorm, D. P. A. Wander, M. Somers, H. S. Overkleeft, D. V. Filippov, J. Désiré, A. Mingot, Y. Blériot, G. A. van der Marel, S. Thibaudeau, J. D. C. Codée, *ACS Cent. Sci.* **2019**, *5*, 781–788.

[45] L. Lebedel, A. Ardá, A. Martin, J. Désiré, A. Mingot, M. Aufiero, N. A. font, R. Gilmour, J. Jiménez-Barbero, Y. Blériot, S. Thibaudeau, *Angew. Chem. Int. Ed.* **2019**, DOI 10.1002/anie.201907001.

[46] W. Schöllkopf, S. Gewinner, H. Junkes, A. Paarmann, G. von Helden, H. Bluem, A. M. M. Todd in Advances in X-ray Free-Electron Lasers Instrumentation III, (Ed.: S. G. Biedron), SPIE, **2015**.

[47] D. A. Thomas, E. Mucha, M. Lettow, G. Meijer, M. Rossi, G. von Helden, *J. Am. Chem. Soc.* **2019**, *141*, 5815–5823.

[48] A. I. González Flórez, E. Mucha, D.-S. Ahn, S. Gewinner, W. Schöllkopf, K. Pagel, G. von Helden, *Angew. Chem. Int. Ed.* **2016**, *55*, 3295–3299.

[49] A. D. Hill, P. J. Reilly, *J. Chem. Inf. Model.* **2007**, *47*, 1031–1035.

[50] S. R. Maple, A. Allerhand, *J. Am. Chem. Soc.* **1987**, *109*, 3168–3169.

[51] *IUPAC Compendium of Chemical Terminology*, (Eds.: M. Nič, J. Jirát, B. Košata, A. Jenkins, A. McNaught), IUPAC, **2009**.

[52] Y. Huang, A.-G. Zhong, Q. Yang, S. Liu, *J. Chem. Phys.* **2011**, *134*, 084103.

[53] A. D. McNaught, *Carbohydr. Res.* **1997**, *297*, 1–92.

[54] D. Cremer, J. A. Pople, *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

[55] H. B. Mayes, L. J. Broadbelt, G. T. Beckham, *J. Am. Chem. Soc.* **2014**, *136*, 1008–1022.

[56] A. M. Rijs, J. Oomens in *Topics in Current Chemistry*, Springer International Publishing, **2014**, pp. 1–42.

[57] J. Seo, W. Hoffmann, S. Warnke, X. Huang, S. Gewinner, W. Schöllkopf, M. T. Bowers, G. von Helden, K. Pagel, *Nat. Chem.* **2016**, *9*, 39–44.

[58] W. Hoffmann, K. Folmert, J. Moschner, X. Huang, H. von Berlepsch, B. Koksch, M. T. Bowers, G. von Helden, K. Pagel, *J. Am. Chem. Soc.* **2017**, *140*, 244–249.

[59] M. Marianski, J. Seo, E. Mucha, D. A. Thomas, S. Jung, R. Schlögl, G. Meijer, A. Trunschke, G. von Helden, *J. Phys. Chem. C* **2018**, *123*, 7845–7853.

[60] N. C. Polfer, B. Paizs, L. C. Snoek, I. Compagnon, S. Suhai, G. Meijer, G. von Helden, J. Oomens, *J. Am. Chem. Soc.* **2005**, *127*, 8571–8579.

[61] J. Roithová, A. Gray, E. Andris, J. Jašík, D. Gerlich, *Acc. Chem. Res.* **2016**, *49*, 223–230.

[62] C. P. Harrilal, A. F. DeBlase, J. L. Fischer, J. T. Lawler, S. A. McLuckey, T. S. Zwier, *J. Phys. Chem. A* **2018**, *122*, 2096–2107.

[63] E. Schrödinger, *Ann. Phys.* **1926**, *385*, 437–490.

[64] D. R. Hartree, *Math. Proc. Camb. Phil. Soc.* **1928**, *24*, 89–110.

[65] V. Fock, *Z. Phys.* **1930**, *61*, 126–148.

[66] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618–622.

[67] G. D. Purvis, R. J. Bartlett, *J. Chem. Phys.* **1982**, *76*, 1910–1918.

[68] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons Inc, **2017**, 660 pp.

[69] R. J. Bartlett, M. Musiał, *Rev. Mod. Phys.* **2007**, *79*, 291–352.

[70] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143–152.

[71] D. G. Liakos, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 4054–4063.

[72] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2016**, *144*, 024109.

[73] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.

[74] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133–A1138.

[75] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.

[76] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[77] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[78] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

[79] D. G. Truhlar, *Chem. Phys. Lett.* **1998**, *294*, 45–48.

[80] A. J. C. Varandas, *J. Phys. Chem. A* **2010**, *114*, 8505–8516.

[81] A. Supady, V. Blum, C. Baldauf, *J. Chem. Inf. Model.* **2015**, *55*, 2338–2348.

[82] M. Marianski, A. Supady, T. Ingram, M. Schneider, C. Baldauf, *J. Chem. Theory Comput.* **2016**, *12*, 6157–6168.

[83] F. Filsinger, D.-S. Ahn, G. Meijer, G. von Helden, *Phys. Chem. Chem. Phys.* **2012**, *14*, 13370.

[84] F. Bierau, P. Kupser, G. Meijer, G. von Helden, *Phys. Rev. Lett.* **2010**, *105*, 133402.

[85] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, *J. Cheminformatics* **2012**, *4*, 17.

[86] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.

[87] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[88] A. Tkatchenko, M. Scheffler, *Phys. Rev. Lett.* **2009**, *102*, 073005.

[89] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[90] A. D. McLean, G. S. Chandler, *J. Chem. Phys.* **1980**, *72*, 5639–5648.

[91] M. J. Frisch *et al.*, *Gaussian 09 Revision D.01*.

[92] F. Neese, *WIREs Comput. Mol. Sci.* **2017**, *8*, e1327.

[93] A. Hellweg, C. Hättig, S. Höfener, W. Klopper, *Theor. Chem. Acc.* **2007**, *117*, 587–597.

[94] L. Montalvillo-Jiménez, A. G. Santana, F. Corzana, G. Jiménez-Osés, J. Jiménez-Barbero, A. M. Gomez, J. L. L. Asensio, *J. Am. Chem. Soc.* **2019**, DOI `10.1021/jacs.9b03285`.

[95] S. Kozuch, J. M. L. Martin, *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104.

[96] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2010**, *7*, 291–309.

[97] J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

[98] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2007**, *120*, 215–241.

[99] X. Xu, W. A. Goddard, *Proc. Nat. Acad. Sci.* **2004**, *101*, 2673–2677.

[100] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[101] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[102] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[103] J. Sun, A. Ruzsinszky, J. Perdew, *Phys. Rev. Lett.* **2015**, *115*, 036402.

[104] L. Goerigk, S. Grimme, *WIREs Comput. Mol. Sci.* **2014**, *4*, 576–600.

[105] M. J. Frisch *et al.*, *Gaussian 16 Revision A.03*.

[106] O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **2001**, *291*, 1523–1527.

[107] P. H. Seeberger, *Acc. Chem. Res.* **2015**, *48*, 1450–1463.

[108] Z. Li, L. Zhu, J. Kalikanda, *Tetrahedron Lett.* **2011**, *52*, 5629–5632.

[109] J. Kalikanda, Z. Li, *J. Org. Chem.* **2011**, *76*, 5207–5218.

[110] M. B. Plutschack, B. Pieber, K. Gilmore, P. H. Seeberger, *Chem. Rev.* **2017**, *117*, 11796–11893.

[111] B. Domon, C. E. Costello, *Glycoconjugate J.* **1988**, *5*, 397–409.

[112] B. Schindler, L. Barnes, G. Renois, C. Gray, S. Chambert, S. Fort, S. Flitsch, C. Loison, A.-R. Allouche, I. Compagnon, *Nat. Commun.* **2017**, *8*, 973.

[113] C. J. Gray, B. Schindler, L. G. Migas, M. Pičmanová, A. R. Allouche, A. P. Green, S. Mandal, M. S. Motawia, R. Sánchez-Pérez, N. Bjarnholt, B. L. Møller, A. M. Rijs, P. E. Barran, I. Compagnon, C. E. Eyers, S. L. Flitsch, *Anal. Chem.* **2017**, *89*, 4540–4549.

A   Appendix

Figures

Figure S1: Energy hierarchies of various conformers of **46Ac** sampled using Fafoom with MMFF94. The relative energies are represented as a function of the (a) C=O4–C1- and the (b) C=O6–C1-distance. All optimized structures are oxocarbenium-type structures (black). The square marks highlights ring puckers $^{1}C_{4}$ and $^{4}C_{1}$, while all the other puckers are represented by crosses.
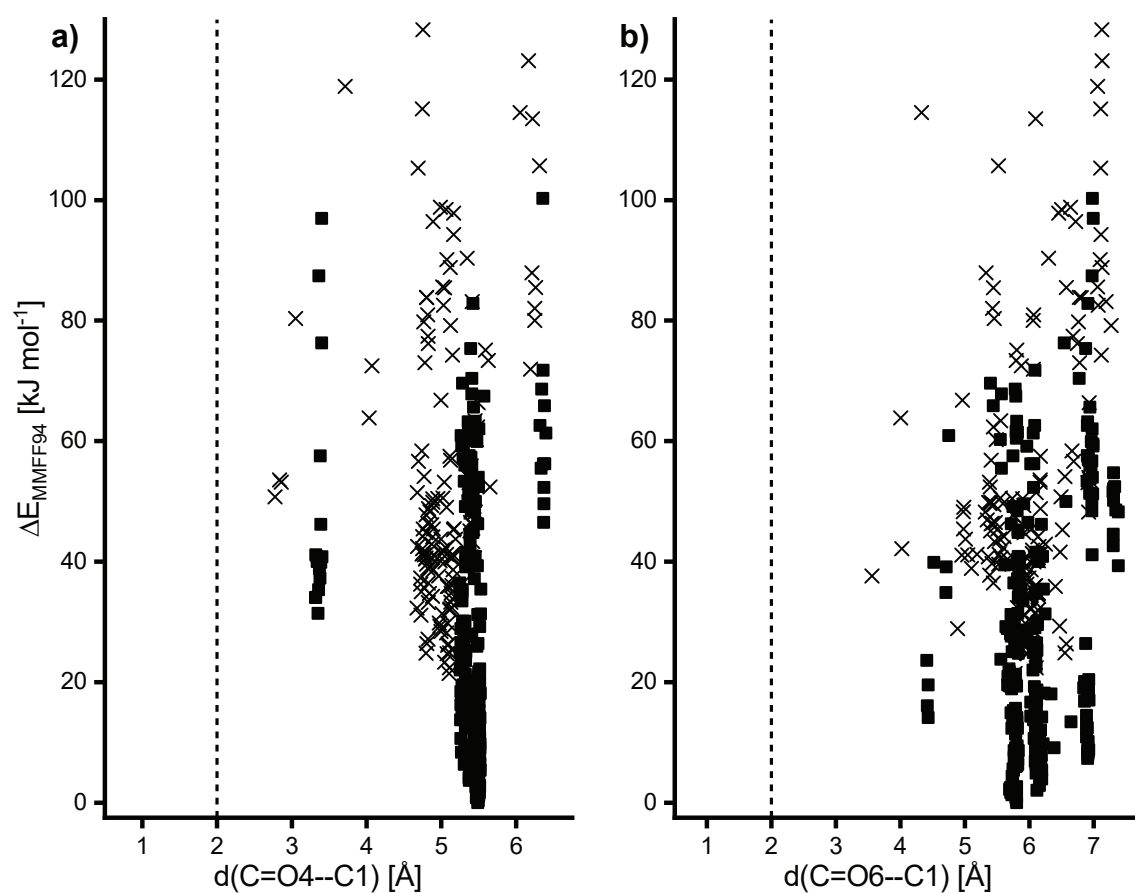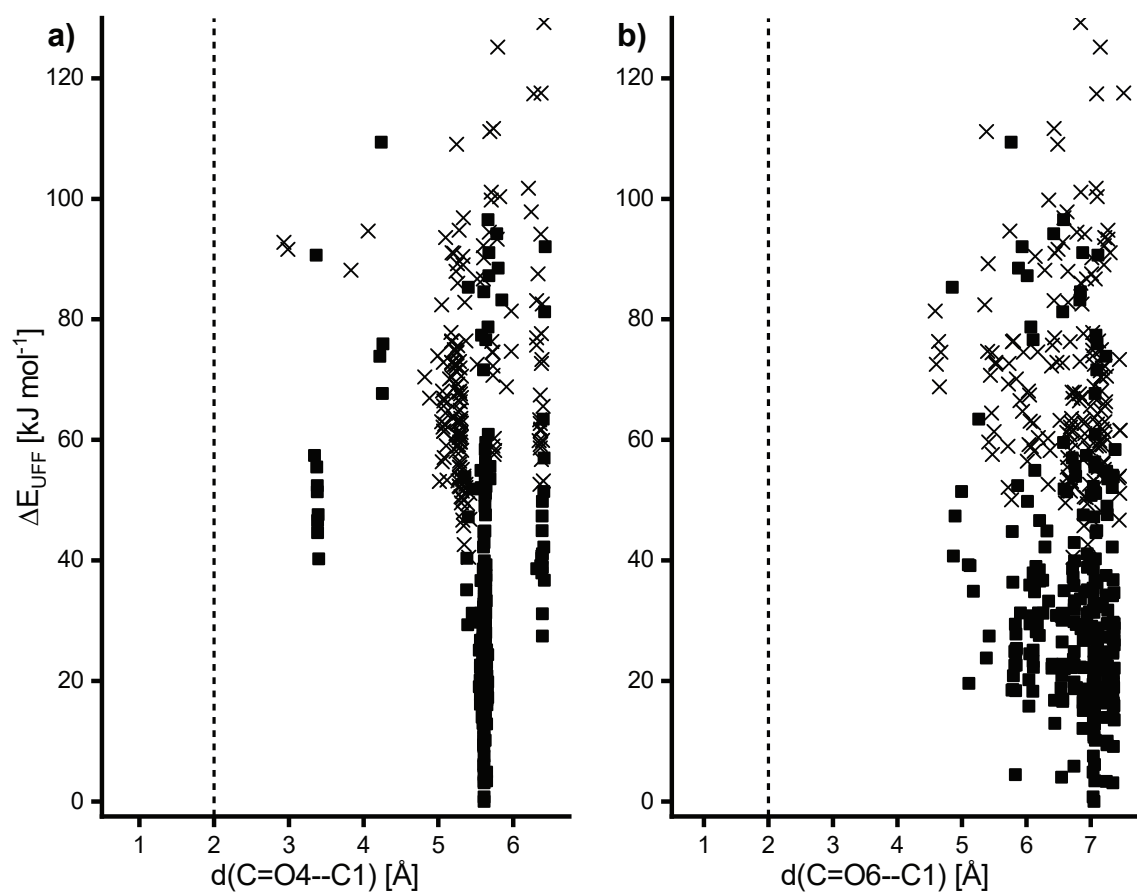
Figure S2: Energy hierarchies of various conformers of **46Ac** sampled using Fafoom with UFF. The relative energies are represented as a function of the (a) C=O4–C1- and the (b) C=O6–C1-distance. All optimized structures are oxocarbenium-type structures (black). The square marks highlights ring puckers $^{1}C_{4}$ and $^{4}C_{1}$, while all the other puckers are represented by crosses.

Figure S3: Energy hierarchies of various conformers of **46Ac** sampled using Fafoom with FHI-aims at PBE-vdW^{TS}/*light* level of theory. The relative energies are represented as a function of the (a) C=O4–C1- and the (b) C=O6–C1-distance. Oxocarbenium-type structures are represented in black, while C4-acetyl dioxolenium-type structures are represented in blue and C6-acetyl dioxolenium-type structures in red. The square marks highlight the structures that have the same ring pucker as the energetically most stable structure ($^1S_5$), while the other ring puckers are represented by a cross.
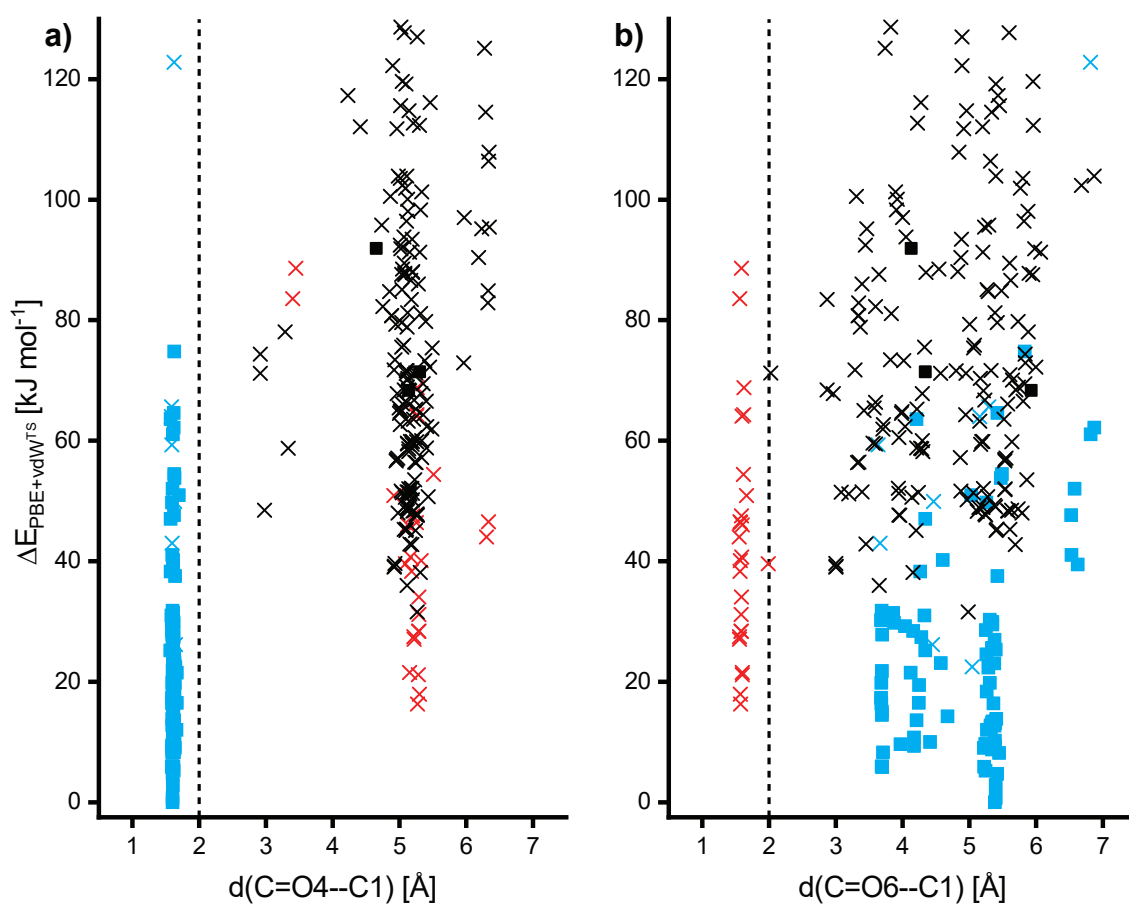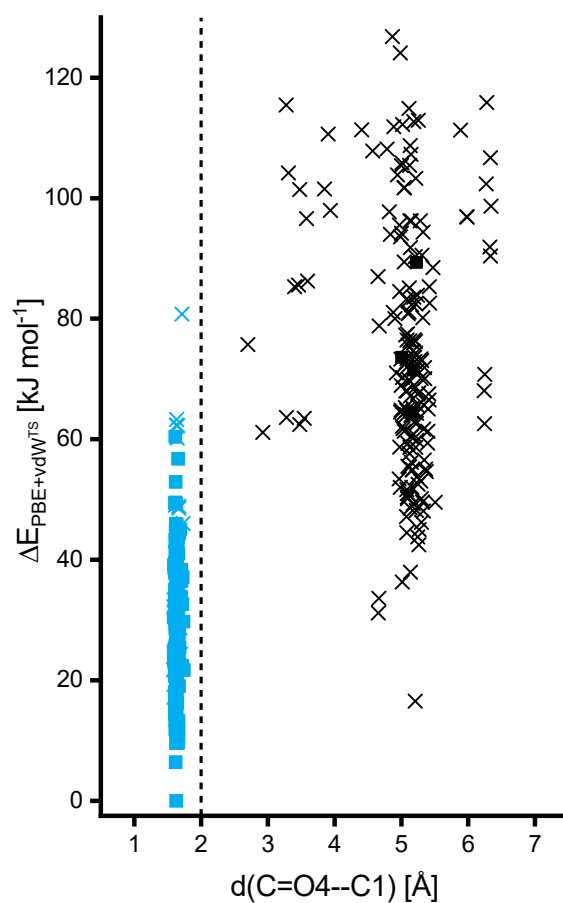
Figure S4: Energy hierarchies of various conformers of **4Ac** sampled using Fafoom with FHI-aims at PBE-vdW^{TS}/*light* level of theory. The relative energies are represented as a function of the C=O4–C1-distance. Oxocarbenium-type structures are represented in black, while C4-acetyl dioxolenium-type structures are represented in blue. The square marks highlight the structures that have the same ring pucker as the energetically most stable structure ($^{1}S_{5}$), while the other ring puckers are represented by a cross.
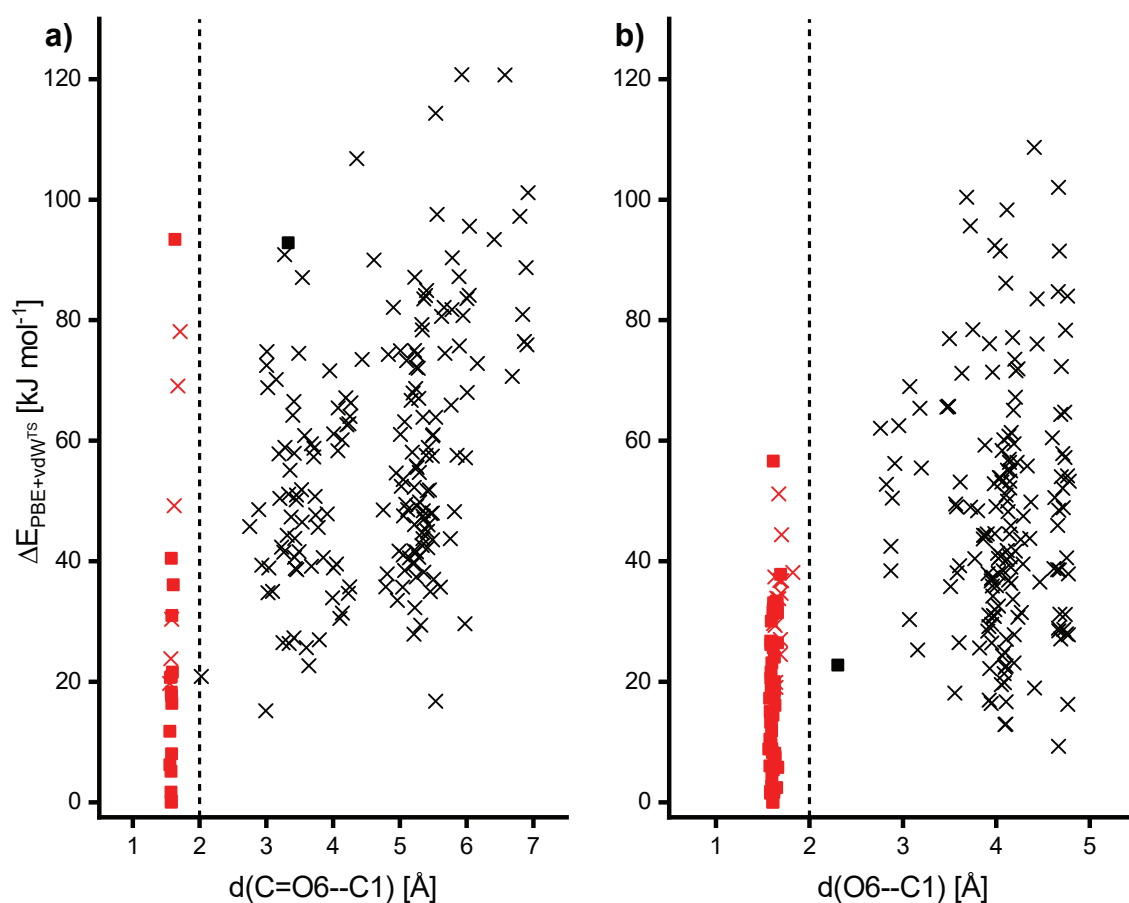
Figure S5: Energy hierarchies of various conformers of (a) **6Ac** and (b) **Bn** sampled using Fafoom with FHI-aims at PBE-vdW^{TS}/*light* level of theory. The relative energies are represented as a function of the (a) C=O6–C1- and the (b) O6–C1-distance. Oxocarbenium-type structures are represented in black, while C6-acetyl dioxolenium- and C6-benzyl oxonium-type structures are represented in red. The square marks highlight the structures that have the same ring pucker as the energetically most stable structure ($^1C_4$), while the other ring puckers are represented by a cross.

Figure S6: Energy hierarchies of reoptimized (PBE0-D3/6-311+G(d,p)) **46Ac** galactosyl cations at RI-MP2/CBS level of theory with a harmonic correction at 78 K. The energies are plotted as a function of the (a) C=O4–C1- and the (b) C=O6–C1-distance. Square marks highlight structures bearing the same ring pucker as the lowest-energy structure, while other puckers are pictured as crosses. C4-acetyl dioxolenium-type structures are shown in blue, while C6-acetyl dioxolenium-type structures are shown in red and oxocarbenium-type structures are represented in black.
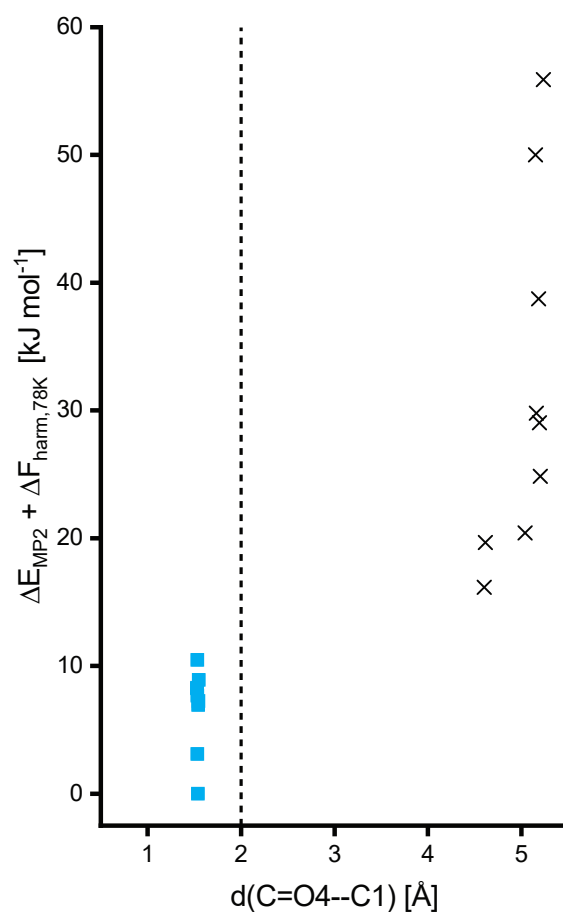
Figure S7: Energy hierarchies of reoptimized (PBE0-D3/6-311+G(d,p)) **4Ac** galactosyl cations at RI-MP2/CBS level of theory with a harmonic correction at 78 K. The energies are plotted as a function of the C=O4–C1-distance. Square marks highlight structures bearing the same ring pucker as the lowest-energy structure, while other puckers are pictured as crosses. C4-acetyl dioxolenium-type structures are shown in blue, while oxocarbenium-type structures are represented in black.
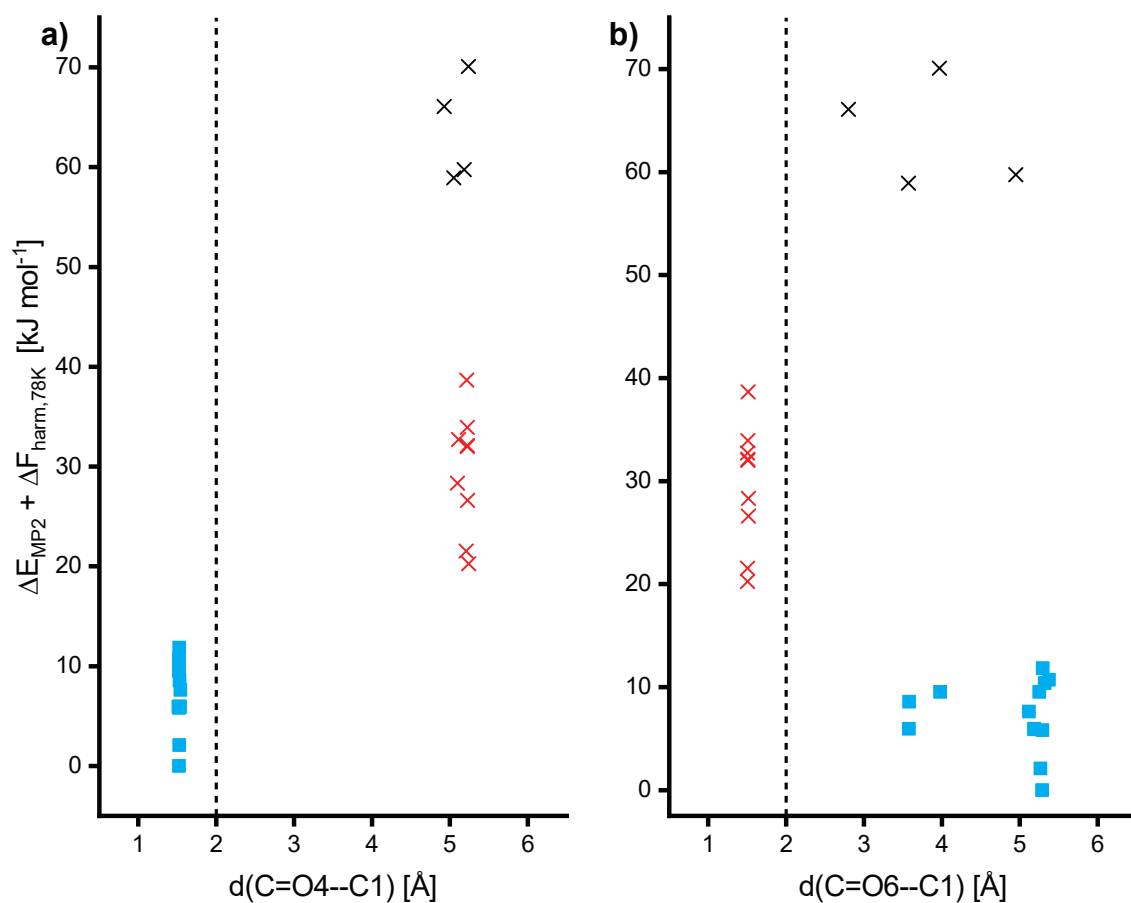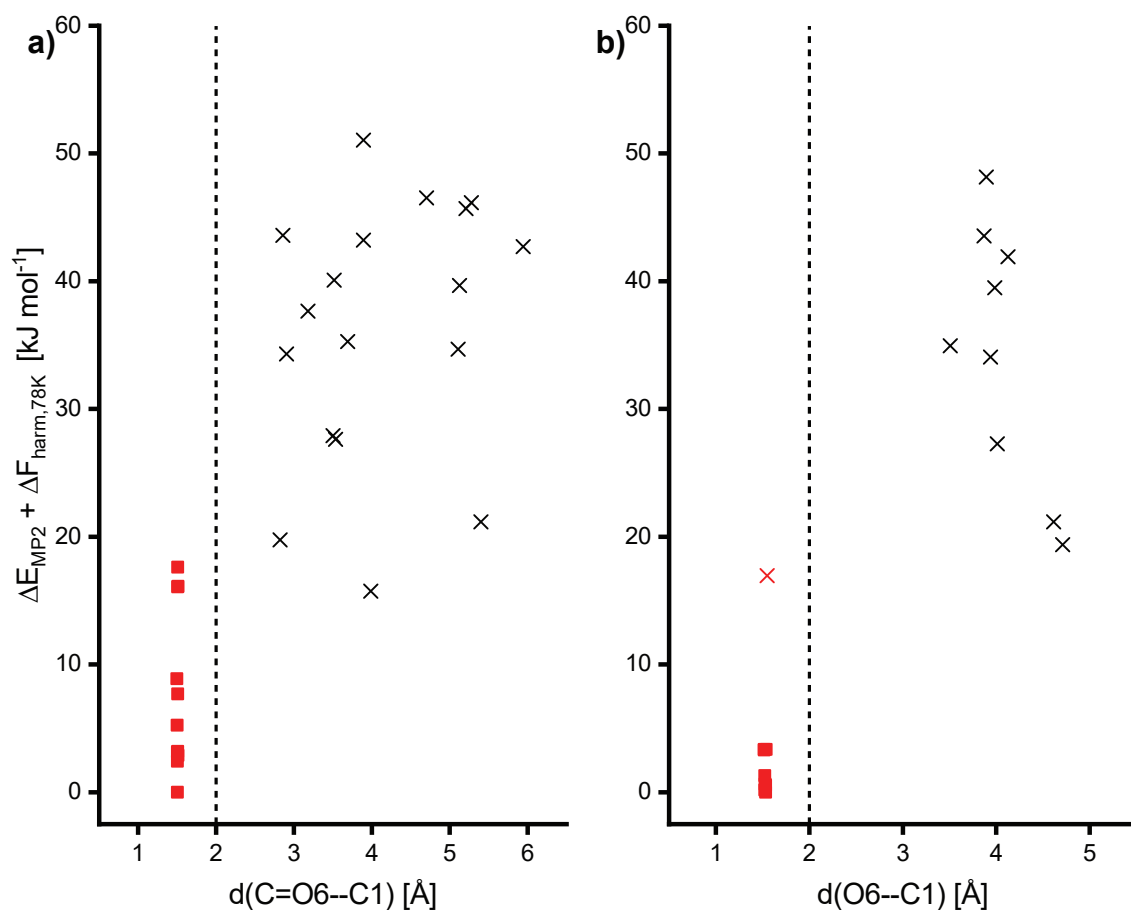
Figure S8: Energy hierarchies of reoptimized (PBE0-D3/6-311+G(d,p)) (a) **6Ac** and (b) **Bn** galactosyl cations at RI-MP2/CBS level of theory with a harmonic correction at 78 K. The energies are plotted as a function of the (a) C=O6–C1- or the (b) O6–C1-distance. Square marks highlight structures bearing the same ring pucker as the lowest-energy structure, while other puckers are pictured as crosses. C6-acetyl dioxolenium-type and C6-benzyl oxonium-type structures are shown in red, while oxocarbenium-type structures are represented in black.

Tables

Table S1: GA parameters used in initial search.

	Parameter	Value
	Distance_cutoff_1	1.2
Molecule	Distance_cutoff_2	2.15
	Rmsd_cutoff_uniq	0.25
	Popsize	10
	Prob_for_crossing	0.95
	Prob_for_mut_pyranosering	0.6
GA settings	Prob_for_mut_torsion	0.8
	Fitness_sum_limit	1.2
	Selection	Roulette wheel
	Max_mutations_torsion	3
	max_iter	30

Table S2: Number of generated and selected structures in the employed multistep procedure.

Glycosyl cation	All generated structures	Unique structures	Selected oxocarbenium-type	Selected dioxolenium- or oxonium-type
46Ac	412	300	4	29
4Ac	685	378	9	10
6Ac	287	208	18	11
Bn	313	249	9	10

Lines of Code

Listing 1: *Python* script that prints bond length, angle, dihedral angle or ring pucker, depending on the number of atom labels it is called with.

```python
##############################################################
###     Script to get bond distance, angle, dihedral angle or pucker
    depending on
###     how many atom labels the script is called with.
###             python scriptname.py filename.xyz C1 C2 C3 C4 C5 O
##############################################################

from __future__ import division
import sys
from math import sqrt
import numpy as np

def get_coordinates_atom(a): #get the coordinates of a distinct atom
    searchfile = open(inputfile, "r")
    ListLines = list(searchfile)
    line = ListLines[a]
    x = float(line.split()[1].split('\n')[0])
    y = float(line.split()[2].split('\n')[0])
    z = float(line.split()[3].split('\n')[0])
    A = np.array([x,y,z])
    searchfile.close()
    return A

def vector(a,b): #get the vector between two atoms
    A = get_coordinates_atom(a)
    B = get_coordinates_atom(b)
    v = B-A
    return v

def bond_length(a,b): #calculate the bond length (call with atomic
    coordinates (two arguments))
    v = vector(a,b)
    L = sqrt(v[0]**2+v[1]**2+v[2]**2)
    return L

def norm(a): #calculate the norm of a vector (equal to bond length, but
    is called with only one argument)
    L = sqrt(a[0]**2+a[1]**2+a[2]**2)
    return L

def scalar_product(a,b): #calculate the scalar product of two vectors
```

```python
    S = a[0]*b[0]+a[1]*b[1]+a[2]*b[2]
    return S


def angle(a,b): #calculates the angle between two bonds/vectors using
    the scalar product (for further uses in functions)
    alpha = np.arccos(scalar_product(a,b)/(norm(a)*norm(b)))*180./np.pi
    return alpha


def molecular_angle(a,b,c): #calculates the angle between three atoms
    v1 = vector(b,a) #in order to get the right result, both vectors
        need to have the same origin
    v2 = vector(b,c)
    alpha = np.arccos(scalar_product(v1,v2)/(norm(v1)*norm(v2)))*180./np
        .pi
    return alpha


def cross_product(a,b): #cross product of a plane is the normal vector
    of the plane
    x = a[1]*b[2]-a[2]*b[1]
    y = a[2]*b[0]-a[0]*b[2]
    z = a[0]*b[1]-a[1]*b[0]
    v = np.array([x,y,z])
    return v


def dihedral_angle(a,b,c,d): #angle between two planes (each plane is
    defined by two vectors)
    v1 = vector(b,a)
    v2 = vector(b,c)
    w1 = vector(c,b)
    w2 = vector(c,d)
    n1 = cross_product(v1,v2)
    n2 = cross_product(w1,w2)
    phi = angle(n1,n2)          #dihedral angle is equal to the angle
        between the two normal vectors of the two planes
    A = scalar_product(-w2,n1) #triple product to determine whether the
        dihedral angle is positive or negative
    if A < 0:                   #(cosinus function in angle function is
        not able to differentiate between positive and negative)
        return -phi
    else:
        return phi


def pucker_dihedrals(a,b,c,d,e,f): #creates a list with the dihedral
    angles in a pyranosering
    dh1 = int(dihedral_angle(a,b,c,d))
    dh2 = int(dihedral_angle(b,c,d,e))
```

```python
    dh3 = int(dihedral_angle(c,d,e,f))
    dh4 = int(dihedral_angle(d,e,f,a))
    dh5 = int(dihedral_angle(e,f,a,b))
    dh6 = int(dihedral_angle(f,a,b,c))
    dh_list = [dh1, dh2, dh3, dh4, dh5, dh6]
    return dh_list


# Dictionary with a list of the 38 canonical sugar puckers and their set
    of dihedrals
# 0,1: chairs; 2-7: boats; 8-13:skewboats; 14-25:halfchairs; 26-37:
    envelopes
dict_dh = {'1C4': [60, -60, 60, -60, 60, -60],
                    '4C1': [-60, 60, -60, 60, -60, 60],
                    '1,4B': [0, 60, -60, 0, 60, -60],
                    'B2,5': [60, 0, -60, 60, 0, -60],
                    '0,3B': [60, -60, 0, 60, -60, 0],
                    'B1,4': [0, -60, 60, 0, -60, 60],
                    '2,5B': [-60, 0, 60, -60, 0, 60],
                    'B0,3': [-60, 60, 0, -60, 60, 0],
                    '1S5': [30, 30, -60, 30, 30, -60],
                    '0S2': [60, -30, -30, 60, -30, -30],
                    '3S1': [30, -60, 30, 30, -60, 30],
                    '5S1': [-30, -30, 60, -30, -30, 60],
                    '2S0': [-60, 30, 30, -60, 30, 30],
                    '1S3': [-30, 60, -30, -30, 60, -30],
                    '1H2': [45, -15, 0, -15, 45, -60],
                    '3H2': [60, -45, 15, 0, 15, -45],
                    '3H4': [45, -60, 45, -15, 0, -15],
                    '5H4': [15, -45, 60, -45, 15, 0],
                    '5H0': [0, -15, 45, -60, 45, -15],
                    '1H0': [15, 0, 15, -45, 60, -45],
                    '4H5': [-15, 45, -60, 45, -15, 0],
                    '0H5': [0, 15, -45, 60, -45, 15],
                    '0H1': [-15, 0, -15, 45, -60, 45],
                    '2H1': [-45, 15, 0, 15, -45, 60],
                    '2H3': [-60, 45, -15, 0, -15, 45],
                    '4H3': [-45, 60, -45, 15, 0, 15],
                    '1E': [30, 0, 0, -30, 60, -60],
                    'E2': [60, -30, 0, 0, 30, -60],
                    '3E': [60, -60, 30, 0, 0, -30],
                    'E4': [30, -60, 60, -30, 0, 0],
                    '5E': [0, -30, 60, -60, 30, 0],
                    'E0': [0, 0, 30, -60, 60, -30],
                    '4E': [-30, 60, -60, 30, 0, 0],
                    'E5': [0, 30, -60, 60, -30, 0],
                    '0E': [0, 0, -30, 60, -60, 30],
```

```
                               'E1': [-30, 0, 0, 30, -60, 60],
                               '2E': [-60, 30, 0, 0, -30, 60],
                               'E3': [-60, 60, -30, 0, 0, 30]}


def pucker_match(mylist,P,N): #searches the canonical pucker that
    matches the best with input pucker coordinates
    for V in range(61):        #V is deviation that is needed for finding
        a match (iterates from 0 to value that is needed for match)
        for x in range(N):    #N is the number of canonical puckers 38
            through which is iterated for finding a match
            if mylist[x][0] in range(P[0]-V,P[0]+V):    #find match for
                the first coordinate
                if mylist[x][1] in range(P[1]-V,P[1]+V):#find match for
                    the second coordinate and so on
                    if mylist[x][2] in range(P[2]-V,P[2]+V):
                        if mylist[x][3] in range(P[3]-V,P[3]+V):
                            if mylist[x][4] in range(P[4]-V,P[4]+V):
                                if mylist[x][5] in range(P[5]-V,P[5]+V):
                    #              print("Deviation = " + str(V))  #
                                prints deviation that is needed to find a
                                match
                                    return mylist[x]


def pucker_multimatch(mydict,mylist,P,N):   #function for finding a
    distinct number of closest matching puckers
    x = pucker_match(mylist,P,N)
    print(find_pucker(mydict, mylist,P,N))
    for i in range(1,3):                        #at the moment the three
        closest matching puckers are searched
        if x in mydict.values():            #number could be increased
            if more matches are neede
            mylist.remove(x)
        x = pucker_match(mylist,P,N-i)
        print(find_pucker(mydict, mylist,P,N-i))


def find_pucker(mydict,mylist,P,N):     #function to print the
    designation of the matching pucker
    v = pucker_match(mylist,P,N)
    return list(mydict.keys())[list(mydict.values()).index(v)]


#arguments that are read in from console
inputfile = str(sys.argv[1]) #coordinate filename


try:
    a1 = int(sys.argv[2])+1     #C1 label
    a2 = int(sys.argv[3])+1     #C2 label
```

```python
    a3 = int(sys.argv[4])+1      #C3 label
    a4 = int(sys.argv[5])+1      #C4 label
    a5 = int(sys.argv[6])+1      #C5 label
    a6 = int(sys.argv[7])+1      #O label
    P = pucker_dihedrals(a1,a2,a3,a4,a5,a6) #creates the coordinate list
        of your input pyranosering
    list_dh = list(dict_dh.values())          #uses the dictionary to
        create a list with the canonical puckering coordinates
    print("Pucker of pyranosering is " + str(find_pucker(dict_dh,list_dh
        ,P,38)))
except IndexError:
    if 'a5' in locals():
        print("Specify two coordinates for calculating bond distances,
            three for angles, four for dihedral angles and six for
            puckers")
    elif 'a3' and 'a4' in locals():
        print("Dihedral Angle is " + str(dihedral_angle(a1,a2,a3,a4)))
    elif 'a3' in locals():
        print("Angle is " + str(molecular_angle(a1,a2,a3)))
    else:
        print("Bond Distance is " + str(bond_length(a1,a2)))
```

Listing 2: Gaussian parser for Fafoom written in *Python*.

```python
'''Wrapper for Gaussian'''

from __future__ import division
import glob
import os, sys
import subprocess

from utilities import sdf2gjf

hartree2eV = 27.21138602

class GaussianObject():
    '''Create and handle Gaussian objects.'''
    def __init__(self, commandline, memory, fafoompath, name_inputfile="
        gaussian_molecule", chargemult="1 1", nprocs=32):
        """Initialize the GaussianObject.

        Args(required):
            commandline
            memory
            fafoompath
        Args(optional):
```

```python
            chargemult (default="1 1")
            nprocs (default=32)
        Raises:
            KeyError: if the commandline, memory or fafoompath is not
                defined
        """
        self.commandline = commandline
        self.name_inputfile = name_inputfile
        self.memory = memory
        self.chargemult = chargemult
        self.nprocs = nprocs
        self.fafoompath = fafoompath

    def generate_input(self, sdf_string):
        """Create input files for Gaussian.
        Args:
            sdf_string (str)
        """
        gjf_string = sdf2gjf(sdf_string)
        coord = gjf_string.split('\n')
        string1 = '%chk='+str(self.name_inputfile)+'\n'
        string2 = '%mem='+str(self.memory)+'\n'
        string3 = '%nprocs='+str(self.nprocs)+'\n'
        string4 = '# '  +str(self.commandline)+'\n'
        string5 = '\n'
        string6 = str(self.name_inputfile)+'\n'
        string7 = '\n'
        string8 = str(self.chargemult)+'\n'
        with open('gaussian_molecule.gjf', 'w') as f:
            f.write(string1)
            f.write(string2)
            f.write(string3)
            f.write(string4)
            f.write(string5)
            f.write(string6)
            f.write(string7)
            f.write(string8)
            f.write('\n'.join(coord))

    def run_gaussian(self, execution_string):
        """Run Gaussian (will automatically write output to '
            gaussian_molecule.log').
        After each run the content of the 'gaussian_molecule.log is
            appened to 'results.out'.
        The optimized geometry is written to 'gaussian_molecule.xyz'.
```

```python
    Warning: this function uses subprocessing to invoke the run.
    The subprocess's shell is set to TRUE.
    Args:
        execution_string (str): e.g. Gaussian or for parallel
            version
        /the/complete/path/to/gaussian
    Raises:
        OSError: if gaussian_molecule.gjf not present in the working
            directory
    """
    success = False
    if os.path.exists('gaussian_molecule.gjf') is False:
        raise OSError("Required input file not present.")
    gaussian = subprocess.Popen(
        execution_string+str(" gaussian_molecule.gjf"),
        stdout=subprocess.PIPE, shell=True)
    gaussian.wait()

    datafile = open("gaussian_molecule.log", "r")
    data = datafile.read()
    datafile.close()

    results = open("results.out", "a")
    results.write(data)
    results.close()

    searchfile = open("gaussian_molecule.log", "r")

    s0 = "Normal termination of Gaussian"
    s = "SCF Done"
    not_conv = True
    for line in searchfile:
        if s0 in line:
            not_conv = False
    searchfile.close()
    if not_conv:
        not_converged = open("not_converged_Gaussian.dat", "w")
        not_converged.close()

    else:
        searchfile = open("gaussian_molecule.log", "r")
        for line in searchfile:
            if s in line:
                energy_tmp = float(line.split(" ")[7])
        searchfile.close()
        self.energy = energy_tmp
```

```python
        os.system('python ' + str(self.fafoompath) + '/
            get_geometry_gaussian.py gaussian_molecule.log')

        with open('gaussian_molecule.xyz', 'r') as f:
            self.xyz_string_opt = f.read()
        f.close()
        success = True
    return success

def get_energy(self):
    """Get the energy of the molecule.

    Returns:
        energy (float) in eV
    Raises:
        AttributeError: if energy hasn't been calculated yet
    """
    if not hasattr(self, 'energy'):
        raise AttributeError("The calculation wasn't performed yet."
            )
    else:
        return hartree2eV*self.energy

def get_xyz_string_opt(self):
    """Get the optimized xyz string.

    Returns:
        optimized xyz string (str)
    Raises:
        AttributeError: if the optimization hasn't been performed
            yet
    """
    if not hasattr(self, 'xyz_string_opt'):
        raise AttributeError("The calculation wasn't performed yet."
            )
    else:
        return self.xyz_string_opt

def clean(self):
    """Clean the working direction after the gaussian calculation
        has been
    completed.
    """
    for f in glob.glob("gaussian_molecule.*"):
        os.remove(f)
```

Acknowledgment

The elaboration of this master thesis would not have been possible without the support of my supervisors, co-workers, fellow students, friends and family.

First of all, I would like to thank Prof. Dr. Kevin Pagel, who gave me the opportunity to work on an interesting topic for my master thesis in his research group. I am most thankful for the inspiring discussions leading to fruitful results.

I am most grateful to Prof. Dr. Dietrich Volmer for his willingness to be the first examiner of this master thesis.

I would like to thank my supervisor Eike Mucha. He taught me how to use the helium droplet machine. Whenever I had trouble, his skills and expertise led very quickly to excellent results. Furthermore, I am much obliged to the whole helium droplet team. We have spent numerous beamtimes together, often until midnight, to get valuable results. Thanks to you, the long shifts were never boring.

I would also like to express my gratefulness to Prof. Dr. Gerard Meijer and Prof. Dr. Gert von Helden and the whole Department of Molecular Physics of the Fritz Haber Institute of the Max Planck Society in general for giving me the opportunity to carry out my master thesis in such an excellent environment, both in terms of people and research. Particularly, I would like to thank Carla Kirschbaum, Lilian Maas, Łukasz Polewski, Michele Pugini, Rayoon Chang, Márkó Grabarics, Maike Lettow, Christian Manz, Eike Mucha, Dr. Andreas Zappe and Dr. Daniel Thomas of the research groups of Kevin Pagel and Gert von Helden for many fruitful discussions.

The acquirement of the results would not have been possible without the expertise of Dr. Wieland Schöllkopf and Sandy Gewinner of the FHI Facility Free Electron Laser. I would like to thank you for operating the laser and providing high-energy IR photons.

I am grateful to Prof. Dr. Mateusz Marianski for helping us with the theoretical calculations. He helped me to deepen my existing knowledge of computational chemistry so that I can now acquire most theoretical results independently. He and Dmitrii Maksimov, the current main developer of Fafoom, encouraged me to learn the programming language *Python*.

I am thankful to Sooyeon Moon and Alonso Pardo for synthesizing the building blocks and performing the flow chemistry experiments of glycosylation reactions to determine the stereochemical outcome of the reactions. I'm grateful to Prof. Dr. Peter Seeberger and Dr. Kerry Gilmore for providing the necessary expertise to discuss the experimental results.

Furthermore, I would like to thank the proofreaders of this work, Eike Mucha, Lilian Maas, Carla Kirschbaum, Christian Manz, Julius Stückrath and Ralf Greis, for finding numerous typing and comma placement errors.

Last but not least, I would like to thank my family for the long lasting mental and financial support. In particular, I would like to thank my mother Flora and my brother Ralf, who always supported and encouraged me to do the things that I wanted.

Publications and Presentations

Peer-Reviewed Journal Articles

[5] M. Marianski*, E. Mucha, <u>K. Greis</u>, S. Moon, A. Pardo, C. Kirschbaum, D. A. Thomas, G. Meijer, G. von Helden, K. Gilmore, P. H. Seeberger, K. Pagel*: Direct Evidence for Remote Participation During Glycosylations Revealed by Cryogenic Vibrational Spectroscopy, *submitted.*

[4] <u>K. Greis</u>, K. Bethke*, J. B. Stückrath, T. T. K. Ingber, S. Valiyaveettil, K. Rademann: One-Pot Synthesis of Xanthate Functionalized Cellulose for the Detection of Micromolar Copper(II) and Nickel(II) Ions, *CLEAN - Soil, Air, Water, accepted.*

[3] <u>K. Greis</u>, Y. Yang, A. J. Canty, R. A. J. O'Hair*: Group 10 Ions in the Formal +1 Oxidation State, *J. Am. Soc. Mass Spectrom.*, DOI: https://doi.org/10.1007/s13361-019-02231-5.

[2] <u>K. Greis</u>, A. J. Canty, R. A. J. O'Hair*: Gas-Phase Reactions of the Group 10 Organometallic Cations, $[(\text{phen})M(CH_3)]^+$ with Acetone: Only Platinum Promotes a Catalytic Cycle via the Enolate $[(\text{phen})Pt(OC(CH_2)CH_3)]^+$, *Z. Phys. Chem.* **2019**, *233(6)*, 845.

[1] K. Bethke, S. Palantöken, V. Andrei, M. Roß, V. S. Raghuwanshi, F. Kettemann, <u>K. Greis</u>, T. T. K. Ingber, J. B. Stückrath, S. Valiyaveettil, K. Rademann*: Functionalized Cellulose for Water Purification, Antimicrobial Applications, and Sensors, *Adv. Funct. Mater.* **2018**, *28*, 1800409.

Poster Presentations

[4] <u>K. Greis</u>, E. Mucha, M. Marianski, S. Moon, K. Gilmore, P. H. Seeberger, K. Pagel: Remote Participation in Glycosylations Revealed by Cryogenic Vibrational Spectroscopy, *Tag der Chemie*, Berlin, **2019**.

[3] <u>K. Greis</u>, Y. Yang, A. J. Canty, R. A. J. O'Hair: Catalytic Decomposition of Acetone and Acetic Acid Mediated by $[(\text{phen})Pt(O_2CCH_3)]^+$, *RACI Victorian Inorganic Chemistry Symposium*, Melbourne, **2018**.

[2] T. T. K. Ingber, J. B. Stückrath, <u>K. Greis</u>, M. Rohloff, K. Bethke, S. Valiyaveettil, A. Fischer, K. Rademann: Mercury Capture by Magnetic Iron Oxide Loaded with Silver Nanoparticles, *ICMAT*, Singapore, **2017**.

[1] <u>K. Greis</u>, J. B. Stückrath, T. T. K. Ingber, K. Bethke, S. Valiyaveettil, K. Rademann: Cellulose Based Chemicals for Heavy Metal Removal Studied by X-Ray-Fluorescence Spectroscopy (XRF), *ICMAT*, Singapore, **2017**.

Oral Presentations

[1] <u>K. Greis</u>: Catalytic Decomposition of Acetone and Acetic Acid Mediated by [(phen)Pt(O$_2$CCH$_3$)]$^+$, *Victorian Mass Spectrometry Symposium*, Melbourne, **2018**.

Statutory Declaration

by Kim Greis, born February 23, 1995 in Luxembourg City

I assert that I have written the submitted Master Thesis *Remote Participation in Glycosylations Revealed by Cryogenic Infrared Spectroscopy* independently and entirely by myself using only the referred sources and support.

Berlin, August 14, 2019

_____

Kim Greis