

# Structural Aspects of the O-glycosylation Linkage in Glycopeptides via MD Simulations and Comparison with NMR Experiments

Aysegül Turupcu,<sup>[a]</sup> Matthias Diem,<sup>[a]</sup> Lorna J. Smith,<sup>[b]</sup> and Chris Oostenbrink<sup>\*[a]</sup>

A powerful conformational searching and enhanced sampling simulation method, and unbiased molecular dynamics simulations have been used along with NMR spectroscopic observables to provide a detailed structural view of O-glycosylation. For four model systems, the force-field parameters can accurately predict experimental NMR observables (J couplings and NOE's). This enables us to derive conclusions based on the generated ensembles, in which O-glycosylation affects the

peptide backbone conformation by forcing it towards to an extended conformation. An exception is described for  $\beta$ -GalNAc-Thr where the  $\alpha$  content is increased and stabilized via hydrogen bonding between the sugar and the peptide backbone, which was not observed in the rest of the studied systems. These observations might offer an explanation for the evolutionary preference of  $\alpha$ -linked GalNAc glycosylation instead of a  $\beta$  link.

## 1. Introduction

Glycosylation is a co- and posttranslational modification (PTM) of proteins which is widely observed. It is the most diverse form of PTM, since different glycosidases and glycosyltransferases within the ER and Golgi apparatus give rise to different patterns of protein glycosylation in each cell line. This variety contributes to the production of glycoproteins with different and specific functions. It has been shown that glycans have important biological roles; such as correct folding of proteins, recognition events important for development, differentiation of a particular cell, tissue, or organism.<sup>[1]</sup>

N- and O-glycosylation are the two most common types of glycosylation where glycans are attached to proteins via either the nitrogen of Asn (N-linked) or the oxygen of Ser/Thr (O-linked). In some limited cases O-glycans can be attached to modified hydroxyproline and hydroxylysine and in a very rare event, O-glycosidic linkage of  $\alpha$ -glucose to tyrosine is observed in glycogen containing eukaryotic cells.<sup>[2–4]</sup> The consensus sequence for N-glycosylation is Asn-X-Thr/Ser where X is any amino acid different than Proline. On the other hand, O-glycosylation does not have a defined sequence motif in the protein. The most common O-glycosylation type in mammals is the mucin-type glycan or O-GalNAc glycans where the first

carbohydrate residue is a conserved N-acetylgalactosamine (GalNAc) which is covalently  $\alpha$ -linked to the side chain hydroxyl substituent of serine or threonine.<sup>[5]</sup> They are found on many secreted and membrane-bound glycoproteins in eukaryotes. There are also non-mucin types of O-linked sugars in mammalian cells including  $\alpha$ -linked fucose and mannose;  $\beta$ -linked xylose and N-acetylglucosamine (GlcNAc), and  $\alpha$  or  $\beta$ -linked galactose and glucose. Examples of these unusual non-mucin O-linked sugars are O-fucose found in epidermal growth factor (EGF) domains; O-GlcNAc on cytosolic and nuclear proteins and O-mannose in bovine peripheral nerve  $\alpha$ -dystroglycan.<sup>[6,7]</sup>

The importance of O-glycosylation is documented through its effect on the properties of the proteins such as increasing the viscosity in mucins or increasing the solubility of the proteins in venoms.<sup>[6,8–10]</sup> In addition to its direct effect on the protein, O-glycosylation also has a notable role in protein-ligand interaction; for example in many species binding between oocyte and spermatozoon is orchestrated by O-linked oligosaccharides found on the zona pellucida protein 3 (ZP3) in the fertilization event.<sup>[11]</sup> Also, O-glycosylation provides a protecting barrier over epithelial surfaces against chemical, physical, and microbial agents, protects from the proteolytic cleavage,<sup>[12]</sup> increases the stability of the protein<sup>[13,14]</sup> even with the shortest O-GalNAc on interleukin-2.<sup>[7]</sup> In disease states, alterations of O-glycans on the cell surface occur which enable cancer cells to be differentiated, and make surface O-glycans biomarkers in directed therapeutic approaches.<sup>[15,16]</sup>

However, there is a lack of reliable information on the structure of glycosylated systems for several reasons. Most of the experimentally solved proteins are recombinantly expressed in bacteria such as *E.coli* lacking the glycosylation machinery. Furthermore, biomolecular structures have undergone extensive manipulation of oligosaccharides before X-ray crystallography or NMR spectroscopy because of their inherent flexibility and high degree of coordination with water. Even though 70% of all proteins are modified by glycans in human cells, only 3 to

[a] A. Turupcu, M. Diem, Prof. Dr. C. Oostenbrink  
Department of Material Sciences and Process Engineering, Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences Vienna, Vienna, Austria  
E-mail: chris.oostenbrink@boku.ac.at

[b] Assoc. Prof. L. J. Smith  
Department of Chemistry, University of Oxford, Oxford, U.K.

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cphc.201900079>

© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA.  
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

4% of the structures in the PDB carry glycan chains.<sup>[17,18]</sup> Since glycans hamper the crystal growth, in eukaryotic expression systems these units are cleaved, explaining why 80% of the available X-ray structures with glycans show only one or two residues of the glycan units. Also, structures from X-ray crystallography offer the crystalline form of glycosylated systems, lacking a more diverse solution representation. NMR can offer structures in solution; however, these represent averages of simultaneously occurring conformers, limiting the amount of structural information. The complexity of the glycosylation creates obstacles while studying them experimentally. A computational approach opens the way by offering a platform which can be controlled; therefore, a direct reasoning on the effect of a model system can be made. Among the computational methods, molecular dynamics simulation emerges as a powerful tool for the modeling of glycosylated systems by offering a detailed spatial and temporal resolution.

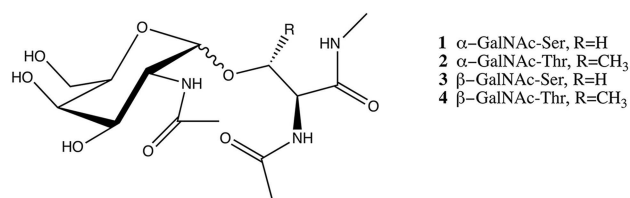
In this context, there is an ongoing effort in developing carbohydrate force fields within different force-field families. Examples are CHARMM,<sup>[19]</sup> GLYCAM06,<sup>[20]</sup> GROMOS 53A6-GLYC,<sup>[21,22]</sup> GROMOS 56A6CARBO,<sup>[23,24]</sup> MARTINI<sup>[25]</sup> (coarse-grained), and a recent polarizable DRUDE forcefield for carbohydrates.<sup>[26]</sup> A review of some of these force fields can be found in ref. [27]. Recently, Corzana et al.<sup>[28–30]</sup> studied O-glycosylated dipeptides through combined NMR and with NMR restrained MD simulations using the AMBER force field. Also, Mallajosyula et al.<sup>[31]</sup> studied 14 glycopeptides (without unglycosylated forms) with Hamiltonian replica exchange (HREX) to validate the CHARMM carbohydrate force field. They used 2D biasing potentials<sup>[32]</sup> which have been previously studied in the content of peptidic<sup>[33,34]</sup> and oligosaccharide systems.<sup>[35]</sup> The GROMOS force field for carbohydrates 53A6GLYC was recently validated for N-glycans and cyclodextrin.<sup>[36,37]</sup>

In this work, unbiased MD and local elevation with umbrella sampling (LEUS) simulations have been applied on four glycopeptides which are models of the most common mucin type O-linkages alongwith their unglycosylated forms. Since the exact effects of these glycans on peptides are unclear, establishment of a reliable computational setup can help to shed light on the resulting structural ensembles. Therefore, we first focused on the more studied systems where there is enough information to compare and validate. To test our force field parameters, the NMR spectroscopy studies of the  $\alpha$ -GalNAc-Ser/Thr and  $\beta$ -GalNAc-Ser/Thr glycopeptides from ref. [28,29] were used as reference values.

## 2. Methods

### 2.1. MD Simulation Settings

In this study, we have focused on the most common O-glycosylation type by using the four model systems represented in Figure 1. Systems 1 and 2 are also known as Tn antigen. All MD simulations were performed using the GROMOS11 biomolecular simulation package (<http://www.gromos.net>)<sup>[38]</sup> and the 54A8 GROMOS force field.<sup>[39]</sup> For compatibility with the protein



**Figure 1.** Graphical representation of the glycopeptides used in this study. The ends of the peptide part are patched with N-acetyl and N-methyl groups at the N- and C-terminus, respectively in order to be compatible with the experimental composition.

force field minor modifications to the original 53A6glyc carbohydrate parameter set were applied.<sup>[36,40]</sup> Initial structures of the studied units were modeled in the molecular operating environment (MOE)<sup>[41]</sup> by setting their glycosidic dihedral angles to their respective free-energy minima, which have been previously reported.<sup>[36]</sup> A short energy minimization was applied using the steepest descent algorithm. The compounds were placed in a periodic cubic water box with simple point charge (SPC) water<sup>[42]</sup> molecules and initialized with a 1.4 nm minimum distance of the solute to the box walls. With position restraints on the solute atoms, the system was further relaxed by a steepest descent minimization before the production run. Then, the systems were equilibrated with initial random velocities generated from a Maxwell-Boltzmann distribution at 60 K then heated up to 300 K in five discrete steps. While heating up the system, position restraints on the solute atoms were reduced from  $2.5 \times 10^4$  to  $0.0 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ .

The production simulations were performed at a constant temperature of 300 K and a constant pressure of 1 atm using a weak coupling scheme<sup>[43]</sup> for both temperature and pressure with coupling times  $\tau_T = 0.1 \text{ ps}$  and  $\tau_P = 0.5 \text{ ps}$ , respectively with an isothermal compressibility of  $4.575 \times 10^{-4} \text{ kJ}^{-1} \text{ mol nm}^3$ . Newton's equations of motion were integrated using the leapfrog scheme<sup>[44]</sup> with a time step of 2 fs. The SHAKE algorithm<sup>[45]</sup> was used to maintain the bond distances at their optimal values. Long-range electrostatic interactions beyond a cutoff of 1.4 nm were truncated and approximated by a generalized reaction field<sup>[46]</sup> with a relative dielectric permittivity of 61.<sup>[47]</sup> Non-bonded interactions up to a distance of 0.8 nm, were computed at every time step using a pairlist<sup>[48]</sup> that was updated every 5 steps. Interactions up to 1.4 nm, were computed at pairlist updates and kept constant in between.

The GROMOS++ software<sup>[49]</sup> is used for time series analysis. A geometrical criterion was used to identify hydrogen bonds if a hydrogen-acceptor distance is smaller than 0.25 nm and the donor-hydrogen-acceptor angle is larger than  $135^\circ$ . Secondary structure propensities are determined as described in Ref. [50], using the definitions in Table S4 in SI.

### 2.2. Parametrization

The backbone parameters of nonglycosylated threonine were reparametrized to better reproduce the experimental J-values

and secondary structure propensities ( $\alpha$ ,  $\beta$  and  $P_{ii}$ ) from Grdadolnik et al.<sup>[51]</sup> A Monte Carlo search scheme was used in combination with Hamiltonian reweighing to identify new dihedral backbone parameters with the closest agreement with experiment.<sup>[50]</sup> The reason to parametrization specifically the Thr backbone parameters was the strong dependence of the  $J$ -value on the backbone length and the secondary structure, which was not seen for Ser (Table S3, in SI). Updated parameters for the 54A8 GROMOS force field can be found in Table S3.

### 2.3. Creating Biased Potentials with LE and Sampling with US

For each system, unbiased MD simulations were carried out for 100 ns after equilibration. In addition to unbiased simulations, an enhanced sampling method, local elevation with umbrella sampling (LEUS)<sup>[52,53]</sup> was applied. To ensure a near-to-complete sampling along both O-glycosidic linkage and peptide backbone, two 2D LE potentials are built separately for the O-glycosidic linkage ( $\phi_s$ ,  $\psi_s$ ) with  $t_{LE}=100$  ns and for the peptide backbone ( $\phi_p$ ,  $\psi_p$ ) with  $t_{LE}=10$  ns. In the US phase, the LE biased potentials were frozen and sampling was applied by using both 2D potentials by saving trajectories every 0.1 ps for 100 ns to achieve statistical efficiency as discussed in ref. [52]. In the LEUS method dihedral angles are binned in  $N_g=36$  bins, a biasing potential width of  $\sigma=360^\circ/N_g$  was used with a force constant increment of  $c=0.005$  kJ mol<sup>-1</sup>.

The unbiased probability of any property  $Q$  can be obtained from the LEUS (biased) simulations through reweighing:

$$P(Q^o) = \frac{\langle \delta(Q - Q^o) \exp[U_{LEUS}(Q)/k_B T] \rangle}{\langle \exp[U_{LEUS}(Q)/k_B T] \rangle} \quad (1)$$

where  $\langle \rangle$  indicates an ensemble average of the biased LEUS simulation,  $U_{LEUS}(Q)$  is the biasing energy at a particular value of  $Q$ ,  $\delta$  is the Kronecker delta function,  $k_B$  is the Boltzmann constant and  $T$  is the absolute temperature. The corresponding free energies can be obtained from the calculated probabilities,

$$G(Q) = -k_B T \ln P(Q) \quad (2)$$

For glycosylated systems, two free energy maps,  $G(\phi_s, \psi_s)$  for the glycosidic linkage and  $G(\phi_p, \psi_p)$  for the protein backbone were created from the LEUS simulations after reweighing of the biased energy with eqs. 1 and 2. To compare the effect of glycosylation, free-energy maps  $G(\phi_p, \psi_p)$  of the unglycosylated systems were also created. The global minimum of each map represents the lowest free energy with the highest probability of the state which is set to 0 kJ/mol and the colormap is drawn using a 5 kJ/mol contour. Detailed explanation for the construction of the free-energy maps can be found in Ref. [36].

### 2.4. <sup>3</sup>J-coupling Constants and NOE Calculations

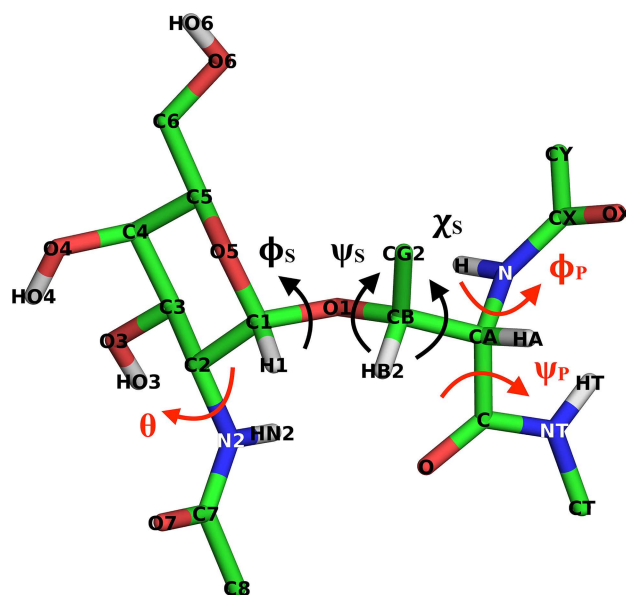
Simulations are compared with NOE data and <sup>3</sup>J-coupling constants. Aliphatic carbons atoms are treated as united atoms in the GROMOS force field. Therefore, virtual atomic positions for prochiral CH<sub>2</sub> ( $C\beta$  in Ser), for CH ( $C\alpha$  and  $C\beta$  in Thr) and pseudo atomic positions for CH<sub>3</sub> ( $C1$  in GalNAc and  $C\gamma$  in Thr) were used to calculate interproton distances. For the NOE analysis, since experimental NOE distances represent an average over space and time, and for small molecules the NOE intensity is proportional to  $r^{-6}$ , averaging is performed as  $\langle r^{-6} \rangle^{-1/6}$ .<sup>[54]</sup> More elaborate approaches to compute the NMR spectra explicitly from simulation trajectories were shown to deviate by at most 8–9% from this inverse sixth power assumption.<sup>[55,56]</sup>

<sup>3</sup>J-coupling can be related to a torsional angle through the Karplus relation. At particular angles, variations in the Karplus relation parameters (a, b, c) lead to differences of up to 3 Hz even though the experimental uncertainties are generally lower than 0.1 Hz.<sup>[57–59]</sup>

<sup>3</sup>J<sub>H<sub>NH</sub>α</sub>, <sup>3</sup>J<sub>H<sub>α</sub>H<sub>β</sub></sub> and <sup>3</sup>J<sub>H<sub>NH</sub>H<sub>2</sub></sub> coupling constants are related to the dihedral angles of  $\phi_s$ ,  $\chi_s$  and  $\theta_{N-Acetyl}$  (Figure 2), and were calculated from MD and LEUS simulations using the following Karplus relations.<sup>[60–62]</sup> Note that for Thr there is one <sup>3</sup>J<sub>H<sub>α</sub>H<sub>β</sub></sub> while for Ser there are two possible values.

$$\begin{aligned} {}^3J_{H_{NH}\alpha} &= 6.51 \cos^2 \theta - 1.76 \cos \theta + 1.60, \\ \theta &= (CX - N - C\alpha - C) - 60^\circ \end{aligned} \quad (3)$$

$$\begin{aligned} {}^3J_{H_{NH}H_2} &= 9.6 \cos^2 \theta - 1.51 \cos \theta + 0.99, \\ \theta &= (C7 - N2 - C2 - C1) + 60^\circ \end{aligned} \quad (4)$$



**Figure 2.** The molecular structure of  $\alpha$ -GalNAc-Thr is represented along with atom names as an example for the rest of the glycopeptides. Dihedral angles are defined as  $\phi_s = O5-C1-O1-C\beta$ ,  $\psi_s = C1-O1-C\beta-C\alpha$ ,  $\chi_s = N-C\alpha-C\beta-O1$ ,  $\phi_p = CX-N-C\alpha-C$ ,  $\psi_p = N-C\alpha-C-NT$ ,  $\theta = O7-C7-N2-C1$ .

$${}^3J_{\text{HaH}\beta} = 9.5\cos^2\theta - 1.6\cos\theta + 1.80, \quad (5)$$

$$\theta = (\text{N} - \text{C}\alpha - \text{C}\beta - \text{O}1) - 120^\circ \text{ for Thr}$$

$${}^3J_{\text{HaH}\beta 2} = 9.5\cos^2\theta - 1.6\cos\theta + 1.80, \quad (6)$$

$$\theta = (\text{N} - \text{C}\alpha - \text{C}\beta - \text{O}1) \text{ for Ser}$$

$${}^3J_{\text{HaH}\beta 3} = 9.5\cos^2\theta - 1.6\cos\theta + 1.80, \quad (7)$$

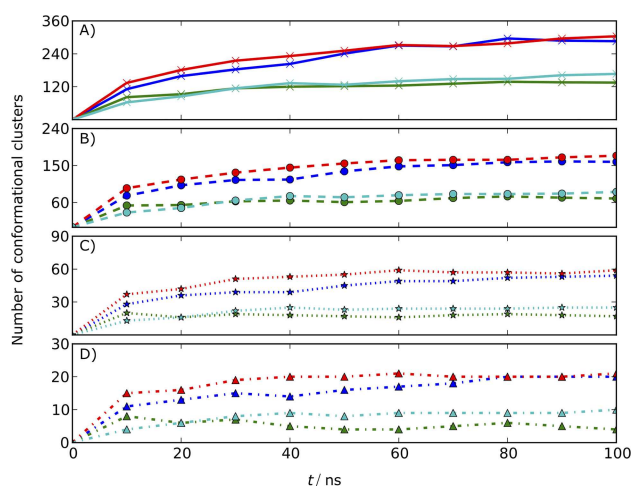
$$\theta = (\text{N} - \text{C}\alpha - \text{C}\beta - \text{O}1) - 120^\circ \text{ for Ser}$$

where the notation (A–B–C–D) denotes the value of the dihedral angle defined by atoms A, B, C and D.

### 3. Results and Discussion

Unbiased MD and LEUS simulations were analyzed in terms of average NOE-derived atom-atom distances,  ${}^3J_{\text{HNH}\alpha}$ ,  ${}^3J_{\text{HaH}\beta}$  and  ${}^3J_{\text{HNH}2}$  couplings, intramolecular hydrogen bonding occurrences, peptide backbone and sugar glycosidic angle distributions.

Overall, we think the LEUS simulations are more appropriate to compare to, due to a more complete sampling. Convergence of the LEUS simulations is shown by performing a conformational clustering over time (see Figure 3). First, an all atom root-



**Figure 3.** Time course of the number of conformational clusters of studied systems 1:  $\alpha$ -D-GalNAc-Ser (red), 2:  $\alpha$ -D-GalNAc-Thr (blue), 3:  $\beta$ -D-GalNAc-Ser (green) and 4:  $\beta$ -GalNAc-Thr (cyan) with a combined weight of 99% (A), 95% (B), 75% (C) and 50% (D). Each point in the curves of the panels represents the total number of conformational clusters that make up 99%, 95%, 75% and 50% of the trajectory sampled up to the corresponding time point. A levelling of the curves implies that no new conformations are sampled.

mean-square difference (RMSD) matrix was calculated after fitting of atomic coordinates of the backbone atoms of the peptide and the ring atoms of the sugar. Then, a conformational clustering<sup>[63]</sup> was performed on a set of 10000 glycopeptide structures taken at 10 ps intervals from the simulation, using a 0.1 nm cutoff to define similar structures.

Figure 3 shows the development over time of the number of clusters that is needed to capture 99%, 95%, 75% and 50% of the trajectory. Convergence of these curves indicate that no

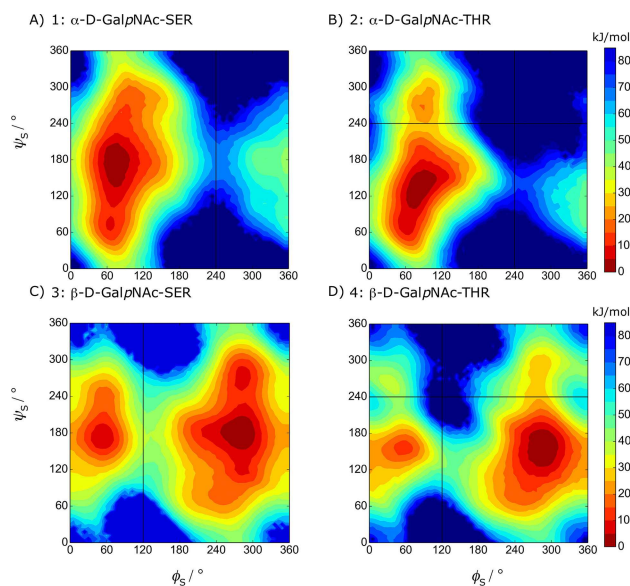
new conformations are sampled anymore. This occurs for the Thr systems (2 and 4) after about 50 ns and for the Ser systems (1 and 3) after about 60 ns, suggesting a reasonably complete sampling of conformational space. Therefore, all the analyses that are calculated from LEUS simulations are shown in the main text. To show the power of the enhanced sampling, the same analysis from the unbiased MD simulations is presented in the supplementary information. Ensemble averages from LEUS simulations were calculated by reweighing to the unbiased ensemble while for unbiased MD simulations they are calculated by averaging over the whole simulation trajectory. Error estimates are reported as the standard deviation over averages obtained from dividing the trajectories in four equally long blocks.

While the secondary structure propensities from unbiased MD simulations capture the LEUS simulations, overall substantial differences are seen in system 2 and system 4. In system 2 unbiased MD simulation shows 0.50  $\beta$ , 0.35  $P_{\text{II}}$  propensity while in LEUS simulations populations of these regions were calculated as 0.38 and 0.51, respectively. This might result from the fact that hydrogen bonding between the amide proton of the N-acetyl group of GalNAc (HN2) and the oxygen of the carbonyl group of C terminus (O) is stabilizing the  $\beta$  conformation with 8% occurrence among the  $\beta$  conformation and this H-bonding was not captured in the  $P_{\text{II}}$  conformation (data not shown). Therefore, in LEUS simulations the true population might have been captured better by crossing the energetic barrier associated with breaking the H-bond. Another difference was seen in system 4, with 0.11  $\alpha$  and 0.49  $P_{\text{II}}$  content in unbiased MD simulations while these were 0.18 and 0.36 in LEUS simulations. The increase in the  $P_{\text{II}}$  content in unbiased MD can be attributed to the hydrogen bonding between the amide proton at the N-terminus (H) and ring oxygen (O5) with an occurrence of 15% in this region, which is not observed in the  $\beta$  region. In addition to this, the  $\alpha$  content of all glycosylated systems were reduced with respect to the unglycosylated peptides except for system 4 where it had 0.11 and 0.18  $\alpha$  content in unbiased MD and LEUS simulations, respectively. H-bonding was observed between the amide proton at the N-terminus (H) and the ring oxygen (O5), with 13% occurrence in the  $P_{\text{II}}$  region, while it is not observed in the  $\beta$  region in MD simulations. The increase in the  $\alpha$  propensity of system 4 can be explained with the stabilization due to HT-O5 hydrogen bonding, which is not observed in the other systems.

#### 3.1. Free-Energy Landscape

Visual inspection of the glycosidic free-energy maps (Figure 4) reveals that the  $\alpha$ -linked systems (1 and 2) present significant minima for  $\phi_5 \approx 60^\circ$  ( $g^+$ ). In contrast, the  $\beta$ -linked systems (3 and 4) present significant minima for both  $\phi_5 \approx 60^\circ$  ( $g^+$ ) and  $\phi_5 \approx 300^\circ$  ( $g^-$ ) in which the latter one is the lowest energetic state. These observations agree with the exo-anomeric effect.<sup>[64–66]</sup>

Further investigation of the free-energy maps reflects the additional steric hindrance upon an additional methyl group in Thr on the conformational preferences of the glycosidic linkage.



**Figure 4.** Free-energy maps  $G(\phi_s, \psi_s)$  of the glycosidic dihedral angles from the LEUS simulations. Contour maps are drawn with 5 kJ/mol spacing starting from the global minimum energy which is set to 0 kJ/mol. The regions that were never visited are shown in dark blue and the corresponding unbiased free energies are represented in the color maps.

For  $\alpha$ -linked Ser,  $\psi_s$  can span almost all of  $0^\circ$  to  $360^\circ$  when  $\phi_s \approx 60^\circ$ ; in Thr the additional methyl group splits the energy landscape at  $\psi_s \approx 240^\circ$ . This difference between system 1 and 2 has an important implication in biological recognition, since antibodies exhibit different affinities towards glycopeptides having  $\alpha$ -linked Ser or  $\alpha$ -linked Thr.<sup>[67]</sup> Our free-energy landscapes can capture this difference with the lowest minima at  $\phi_s \approx 60^\circ$  while  $\psi_s \approx 60^\circ$  and  $120^\circ$  for system 1 and 2, respectively. This was recently confirmed experimentally to be the main conformers in solution and in the bound state.<sup>[68]</sup> For  $\beta$ -linked systems, a similar difference is seen for  $\beta$ -linked Thr (system 4) where the upper region with  $\phi_s \approx 300^\circ$  and  $\psi_s \approx 270^\circ$  is no longer thermally accessible (higher than  $10 k_B T$ ). In contrast,  $\beta$ -linked Ser exhibits a thermally accessible state at this region.

The free-energy maps of the protein backbone dihedral for unglycosylated and for the glycosylated systems are shown in Figure 5. One can clearly see the effect of glycosylation on the protein backbone from these free-energy maps where the  $\alpha_R$

region becomes an energetically less favorable conformational state, except for system 4.

## 3.2. $^3J$ Coupling Constants

### 3.2.1. $^3J_{\text{HNH}\alpha}$ Couplings

For all the systems 1 to 4 as well as Ser and Thr without glycosylation, Table 1 shows the experimental  $^3J_{\text{HNH}\alpha}$  couplings and the computed values from the LEUS simulations (see Table S5 for unbiased MD results). Binned  $\phi_p$  preferences are also reported to trace the contribution to the J-value.

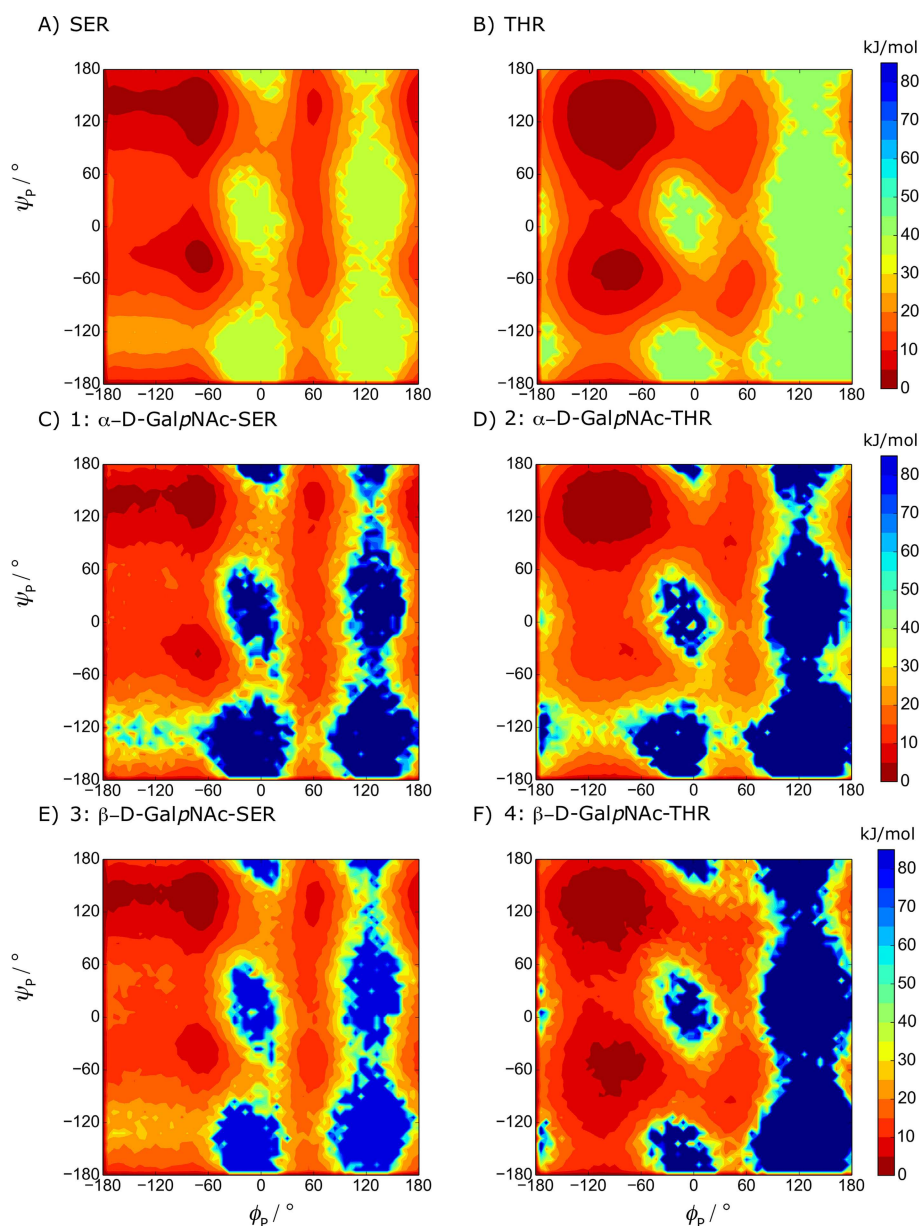
Since the  $^3J_{\text{HNH}\alpha}$  coupling constant reflects only the  $\phi_p$  torsion angle, one can not distinguish secondary structure preferences. Therefore, propensities were calculated and reported for all the systems in Table 1 for LEUS simulations and in Table S5 for unbiased MD simulations.

First, we investigated the Ser and Thr aminoacids with an acetyl group at the N-terminus and a N-methyl group at the C-terminus to check the protein backbone model without glycosylation. Experimental propensities from Grdadolnik et al.,<sup>[51]</sup> obtained from fitting ATR-Absorbance and Raman data, give exceedingly low populations of alpha conformations with 3% for Ser and 4% for Thr, compared to studies from random coil analyses in whole proteins. The reason of the low  $\alpha$  content is the fact that the system which consists of single amino acid can not form hydrogen bonding; therefore, can not form helical shapes. The sensitivity of the  $^3J_{\text{HNH}\alpha}$  value of Thr can also be evidenced with the different experimental studies displaying variation according to length of the backbone; 7.4 Hz in the shortest peptide, 7.9 Hz in the pentapeptide (Table S3). In preliminary simulations of the smallest peptides, it turned out that in particular, the conformational preferences of Thr and its  $^3J_{\text{HNH}\alpha}$  couplings did not agree with the available dipeptide experimental data<sup>[51,69]</sup> (Table S1). The calculated J coupling constant with the original backbone parameters had a 1.1 Hz deviation from the experimental value, which was greater than the deviation of Ser with 0.7 Hz. This deviation in Thr can be explained by looking at the calculated propensities where it prefers mostly an  $\alpha$  conformation instead of  $\beta$  with the 54A8 parameter set. For this reason, new backbone torsional parameters for Thr are introduced and these parameters are used for  $\alpha/\beta$ -GalNAC-Thr (system 2 and 4). 54A8 and updated

**Table 1.** Experimental and calculated  $^3J_{\text{HNH}\alpha}$  coupling constants with propensities from LEUS simulations after reweighing.  $\phi_p$  distributions are calculated to track the contribution of each conformation to the J value.

Sys.	Exp. <sup>a</sup> $^3J_{\text{HNH}\alpha}$	LEUS $^3J_{\text{HNH}\alpha}$	LEUS Propensities <sup>b</sup>		$P_{\parallel}$	UC	LEUS $\phi$ Preferences		
			$\alpha$	$\beta$			$[-180^\circ, -100^\circ]$	$[-100^\circ, 0^\circ]$	$[0^\circ, 180^\circ]$
SER	7.0	$6.3 \pm 0.01$	0.17	0.24	0.49	0.10	45.0%	38.0%	17.0%
THR	7.4	$7.4 \pm 0.01$	0.16	0.25	0.44	0.15	37.4%	48.2%	14.4%
1	6.2	$6.5 \pm 0.2$	0.11	0.31	0.50	0.08	37.2%	56.5%	6.3%
2	8.8	$7.7 \pm 0.1$	0.02	0.38	0.51	0.09	23.3%	75.1%	1.6%
3	6.6	$6.3 \pm 0.1$	0.07	0.28	0.56	0.09	31.6%	63.9%	4.5%
4	7.4	$7.7 \pm 0.1$	0.18	0.29	0.36	0.17	43.8%	54.7%	1.5%

<sup>a</sup> Experimental values for Ser & Thr from Ref. [69] and rest are from [28,29], uncertainties not supplied; <sup>b</sup> Propensities are computed according to the definitions in Figure S2. UC stands for unclassified.

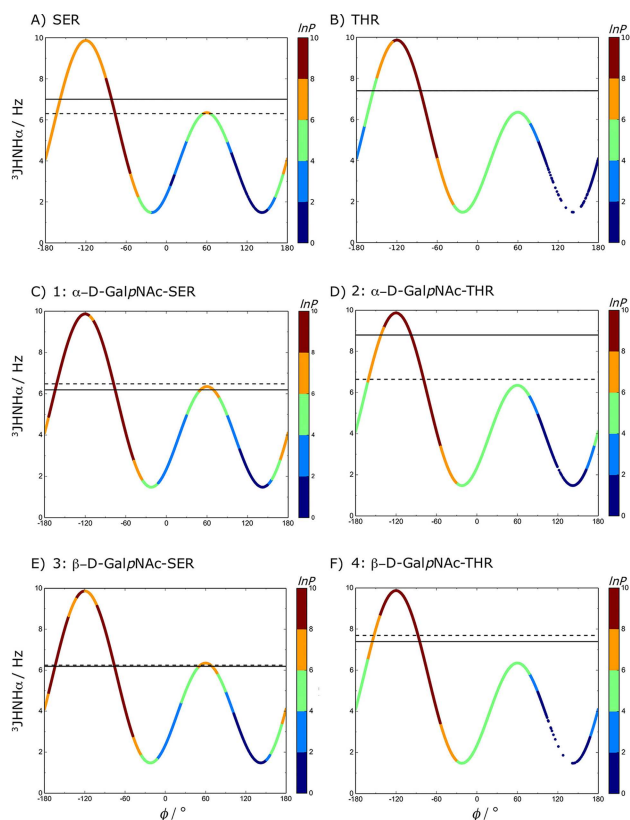


**Figure 5.** Free-energy maps  $G(\phi_p, \psi_p)$  of the protein backbone dihedral angles from the LEUS simulations. Colors are explained in Figure 3.

backbone dihedral angle parameters ( $k$ ,  $\theta^\circ$  and  $m$ ) are presented in Table S2. Altering the backbone dihedral angle parameters in threonine shifted the propensities towards the experimental target values with an increase in  $\beta$  population from 0.09 to 0.24. Although the experimental  $\beta$  propensity has still not been met (0.58), the deviation in the  $^3J_{\text{HNH}\alpha}$  coupling was reduced to 0.01 Hz.

Next, we turn our attention to the glycosylated systems. A maximum deviation between the computed and measured  $^3J_{\text{HNH}\alpha}$ -coupling values of 1.1 Hz was observed for system 1. The rest of the  $^3J_{\text{HNH}\alpha}$  couplings for the glycosylated systems were in agreement with the experimental values with a deviation of 0.3 Hz. Compared to system 1, system 2 has a large experimental  $^3J_{\text{HNH}\alpha}$ -coupling of 8.8 Hz<sup>[28]</sup> which has also been

reported by other groups in the 8.3–9.2 Hz range. Coltart et al.<sup>[16]</sup> report a value of 6.7–7.0 Hz for system 1 which is still smaller than the values reported for system 2. This increase in  $^3J_{\text{HNH}\alpha}$ -coupling of  $\alpha$ -linked O-glycosylation in threonine as compared to serine does not seem to be completely captured in our simulations, even though we do see a pronounced reduction of the  $\alpha$ -conformation towards the  $\beta$ -conformations (Table 1) for this system (see also Figure 5). In Figure 6, we visualize where the large deviation in the  $^3J_{\text{HNH}\alpha}$ -value comes from. In contrast to the Ser cases, the glycosylated Thr does not sample values of  $\phi_p$  in the range  $[-180^\circ, -120^\circ]$  as much, and as a result, the  $^3J_{\text{HNH}\alpha}$  value does not increase to the extent observed in the experiments. The systems with  $\beta$ -link exper-

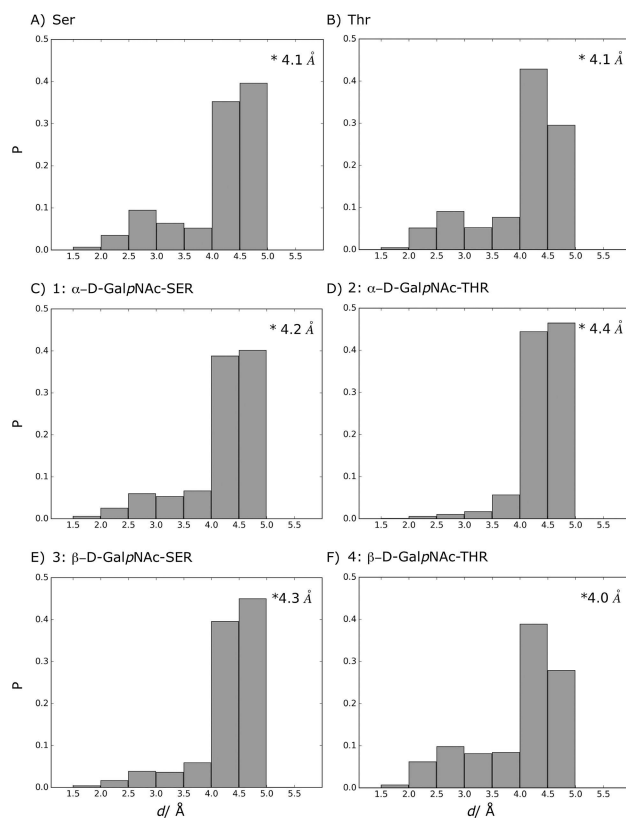


**Figure 6.**  $\phi_p$  vs. calculated  $^3J_{\text{HNH}\alpha}$  couplings from LEUS simulations for all studied systems. Experimental and calculated  $^3J$  values are represented with solid and dashed horizontal lines, respectively. The colors on this Karplus curve indicate the preferred sampling after unbiasing of the LEUS simulations. In the unbiasing procedure, LEUS occurrences ( $P$ ) are binned with  $6^\circ$  grid spacing. Negative values of  $\ln P$  set to zero.

imentally show a smaller effect of glycosylation, which we seem to represent more accurately.

Evidence for the extended conformation of system 4 was further checked with the analysis of the HT–H distance ( $d(\text{H}_T, \text{H})$ ) which was shown to be at 4.4 Å through NMR studies.<sup>[70]</sup> For all systems  $d(\text{H}_T, \text{H})$  distributions were calculated from LEUS simulations and reported as populations in Figure 7. Upon glycosylation the average value of  $d(\text{H}_T, \text{H})$  increases, except for system 4 which exhibits a decrease due to an increase in the population of the bins from 1.5 Å to 3.5 Å.

The notable preference for  $\alpha$ -helical conformations at the expense of more extended  $P_{\text{II}}$  conformations may offer an explanation why  $\beta$ -linked GalNAc is not observed in naturally occurring O-glycosylations, which are mostly observed in extended ( $P_{\text{II}}$ ) regions of the protein. There are  $\beta$ -linked O-glycosylation examples, but they are seen with GlcNAc (N-acetylglucosamine), Gal (Galactose), Glc (Glucose) or Xyl (Xylose).



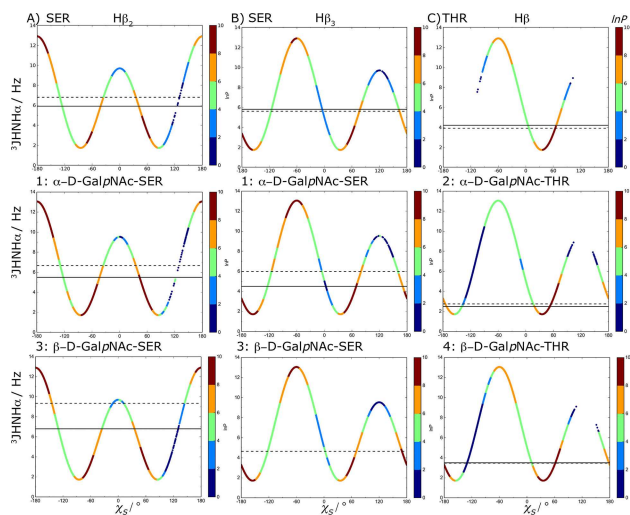
**Figure 7.** Calculated  $\text{H}_T, \text{H}$  distance (Å) distributions from LEUS simulations after unbiasing. Overall average unbiased distance of each system is represented in the upper-right corner of every panel. NMR studies reported that extended backbone conformation shows  $d(\text{H}_T, \text{H})$  at 4.4 Å and 4.3 Å for  $\beta$  and  $P_{\text{II}}$ , respectively.<sup>[70]</sup>

### 3.2.2. $^3J_{\text{H}\alpha\text{H}\beta}$ Couplings

Table 2 compares multiple measured  $^3J_{\text{H}\alpha\text{H}\beta}$  values to the computed values from LEUS simulations; for unbiased MD simulation results see Table S6. Note that Ref. [16] measured  $^3J_{\text{H}\alpha\text{H}\beta}$  in a pentapeptide (STTAV). As this seems to be the most relevant available data for the unglycosylated systems, we also include the corresponding values for the glycosylated systems. For system 4, we see the effect of more complete sampling in LEUS, where the  $J$ -value drops from 6 Hz in plain MD to 3.5 Hz in the LEUS simulations, in agreement with experimental values 3.5 Hz<sup>[29]</sup> or 4.6 Hz.<sup>[16]</sup> Our simulations reproduce the increase in the experimental  $J$ -value from  $\alpha$  to  $\beta$ -linked Thr with about 0.5 Hz deviation from the averaged experimental values ( $\alpha$ -linked 2.4 Hz;  $\beta$ -linked 4.0 Hz). The population of  $\chi_5$  that is in the  $[-120^\circ, 0^\circ]$  region increases by 9%. This increase is also depicted in Figure 4, where this region changes its color from green to orange showing an increase in the population. Since this region corresponds to a higher  $J$ -value, increase in this population results in a higher  $J$ -value for system 4 compared to system 2. The same effect is seen for systems 1 and 3  $\beta$ -linked GalNAc shows an increase from 6.7 to 9.4 Hz when compared to the  $\alpha$ -linked analogue. Experimentally, this increase is also observed, albeit considerably less pronounced

in Refs. [28,29] than in [16]. For these systems the increase in  $^3J_{\text{H}\alpha\text{H}\beta}$  comes from the doubling of the anti population ( $[-180^\circ, -120^\circ]$ ,  $[120^\circ, 180^\circ]$ ). This region of the Karplus curve corresponds to the steepest and most sensitive edges, resulting in large deviations from a small difference in the dihedral angle, which probably explains the deviation from the experimental values.

Overall, for this J-value, the Thr systems agree excellently with the measured data, while larger deviations are observed for the Ser systems. From Figure 8 we learn that the average



**Figure 8.**  $\chi_5$  vs. calculated  $^3J_{\text{H}\alpha\text{H}\beta}$  coupling from LEUS simulations for all studied systems. Experimental and calculated average  $^3J_{\text{H}\alpha\text{H}\beta}$  values are represented with solid and dashed horizontal lines, respectively. Only the dipeptide experimental values from Refs. [28,29] are compared here. The colors indicate the preferred sampling after unbiasing of the LEUS simulations.

values obtained in the simulations are the result of sampling three conformations with lower and higher J-values than observed in the experiment. A slight shift in the conformational preferences towards the gauche  $+/-$  conformations could improve the agreement. Furthermore, it becomes clear from Figure 8 that the  $\chi_5$  angle of the Thr systems are more restricted than the Ser systems. While  $\chi_5$  can take all three conformations with highest populations (represented in red) in the Ser

systems, in the Thr systems conformations around  $+60^\circ$  are preferred over the other conformations.

### 3.2.3. $^3J_{\text{H}\text{N}\text{H}2}$ Couplings

Finally, we consider the  $^3J_{\text{H}\text{N}\text{H}2}$ -values that result from the N-acetyl group in GalNAc. The results are summarized in Table 3

<b>Table 3.</b> Experimental and calculated $^3J_{\text{H}\text{N}\text{H}2}$ coupling constants from LEUS simulations after reweighing. $\theta$ distributions are calculated to track the contribution of each conformation to the J value.					
Sys.	Exp. <sup>a</sup> $^3J_{\text{H}\text{N}\text{H}2}$	LEUS $^3J_{\text{H}\text{N}\text{H}2}$	LEUS Preferences		
			$[-180^\circ, -120^\circ]$ $[120^\circ, 180^\circ]$	$[-120^\circ, -60^\circ]$ $[60^\circ, 120^\circ]$	$[-60^\circ, 60^\circ]$
1	9.2	$10.7 \pm 0.2$	98.9%	1.1%	0.1%
2	9.5	$10.8 \pm 0.1$	99.2%	0.8%	0.0%
3	9.6	$11.4 \pm 0.1$	100.0%	0.0%	0.0%
4	9.6	$11.3 \pm 0.1$	99.9%	0.1%	0.0%

<sup>a</sup> Experimental values from ref. [28,29], uncertainties not supplied.

and Figure 9 for LEUS simulations and Table S7 for MD simulations. No significant differences between plain MD and LEUS were observed. The simulations seem to consistently overestimate the  $^3J_{\text{H}\text{N}\text{H}2}$ -values. As can be seen in Figure 9, we do sample the correct dihedral angles, but because these regions correspond to the steep areas of the Karplus curve, small deviations in the angle (or the Karplus constants) lead to large deviations in the J-value. Comparison of the  $\alpha$  and  $\beta$  systems shows that there is a slight increase in the calculated J-values for  $\beta$  systems (0.7 Hz) which can be explained by looking at Figure 9 where the gauche conformations are not accessible in the  $\beta$  systems (which corresponds to a lower J-value) causing an increase in the overall J-value.

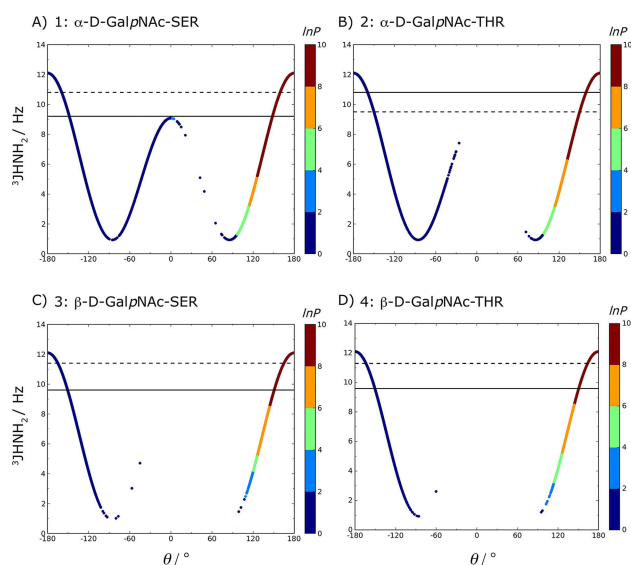
### 3.3. NOE Analysis

As the systems consist of small molecules for which the molecular tumbling is fast, we used  $r^{-6}$  averaging to compare distances observed in the simulations to those derived from NOE intensities. To check this hypothesis, the rotational relaxation time of the studied systems were calculated from the

<b>Table 2.</b> Experimental and calculated $^3J_{\text{H}\alpha\text{H}\beta}$ coupling constants from LEUS simulations after reweighing. $\chi_5$ distributions are calculated to track the contribution of each conformation to the J value.							
Sys.	Exp. $^3J_{\text{H}\alpha\text{H}\beta}$	Exp. $^3J_{\alpha\text{H}\beta 2}$	LEUS $^3J_{\text{H}\alpha\text{H}\beta}$	LEUS $^3J_{\alpha\text{H}\beta 2}$	LEUS $\chi_5$ Preferences		
					$[-180^\circ, -120^\circ]$ $[120^\circ, 180^\circ]$	$[-120^\circ, 0^\circ]$	$[0^\circ, 120^\circ]$
SER	5.8 <sup>a</sup>	5.9 <sup>a</sup>	$6.8 \pm 0.08$	$5.6 \pm 0.04$	38.5%	28.7%	32.8%
THR	4.2 <sup>a</sup> ; 5.0 <sup>a</sup>		$3.9 \pm 0.01$		0.0%	13.1%	89.9%
1	5.5 <sup>b</sup> ; 5.2 <sup>a</sup>	4.5 <sup>b</sup> , 3.3 <sup>a</sup>	$6.7 \pm 2.0$	$6.0 \pm 0.8$	38.7%	34.1%	27.2%
2	2.5 <sup>c</sup> ; 2.3 <sup>a</sup>		$2.8 \pm 0.1$		7.3%	0.7%	92.0%
3	6.8 <sup>c</sup> ; 8.0 <sup>a</sup>	- <sup>c</sup> , 5.5 <sup>a</sup>	$9.4 \pm 1.1$	$4.7 \pm 1.3$	66.4%	19.3%	14.3%
4	3.5 <sup>c</sup> ; 4.6 <sup>c</sup>		$3.5 \pm 0.5$		9.8%	9.5%	80.7%

<sup>a</sup> Experimental values from ref. [16]; <sup>b</sup> from ref. [28]; <sup>c</sup> from ref. [29], uncertainties not supplied.





**Figure 9.**  $\theta$  vs. calculated  $^3J_{\text{HNH}_2}$  couplings from LEUS simulations after reweighing for glycosylated systems. Experimental and calculated average  $^3J_{\text{HNH}_2}$  values are represented with solid and dashed horizontal lines, respectively. The colors indicate the preferred sampling after unbiased of the LEUS simulations.

**Table 4.** Comparison of experimental NOE data with  $r^{-6}$  averaged distances in Å obtained from unbiased MD and LEUS simulations (reweighted) for systems 1–4.

NOE	Exp. <sup>a</sup>	MD	MD violation	LEUS	LEUS violation
<b>1: <math>\alpha</math>-D-GalNAc-Ser</b>					
d(H <sub>T</sub> ,H)	2.9	3.4 ± 0.02	0.5	3.4 ± 0.2	0.5
d(H <sub>A</sub> ,H <sub>T</sub> )	2.3	2.3 ± 0.01	0	2.3 ± 0.06	0
d(H <sub>A</sub> ,H)	2.9	2.6 ± 0.02	0	2.7 ± 0.2	0
d(H <sub>T</sub> ,H <sub>B1</sub> )	2.5	3.5 ± 0.01	1.0	3.2 ± 0.2	0.7
d(H <sub>T</sub> ,H <sub>B2</sub> )	2.6	3.2 ± 0.06	0.6	3.1 ± 0.2	0.5
d(H <sub>1</sub> ,H <sub>A</sub> )	3.9	3.8 ± 0.01	0	3.7 ± 0.6	0
<b>2: <math>\alpha</math>-D-GalNAc-Thr</b>					
d(H <sub>T</sub> ,H)	2.8	3.8 ± 0.03	1.0	4.0 ± 0.6	1.2
d(H <sub>A</sub> ,H <sub>T</sub> )	2.4	2.2 ± 0.3	0	2.2 ± 0.6	0
d(H <sub>A</sub> ,H)	2.9	2.8 ± 0.01	0	2.7 ± 0.03	0
d(H <sub>T</sub> ,H <sub>B</sub> )	2.8	3.2 ± 0.01	0.4	2.8 ± 0.1	0
d(H <sub>1</sub> ,H <sub>B</sub> )	3.5	3.0 ± 0.2	0	3.4 ± 0.4	0
d(H <sub>1</sub> ,H <sub>N2</sub> )	3.3	4.0 ± 0.06	0.6	3.4 ± 0.3	0.1
<b>3: <math>\beta</math>-D-GalNAc-Ser</b>					
d(H <sub>T</sub> ,H)	2.9	3.4 ± 0.03	0.5	3.6 ± 0.2	0.7
d(H <sub>A</sub> ,H <sub>T</sub> )	2.2	2.3 ± 0.01	0.1	2.3 ± 0.00	0.1
d(H <sub>A</sub> ,H)	2.6	2.6 ± 0.04	0	2.7 ± 0.1	0.1
d(H <sub>T</sub> ,H <sub>B1</sub> )	2.6	3.4 ± 0.01	0.8	3.4 ± 0.2	0.8
d(H <sub>T</sub> ,H <sub>B2</sub> )	2.8	3.3 ± 0.04	0.5	3.4 ± 0.3	0.6
<b>4: <math>\beta</math>-D-GalNAc-Thr</b>					
d(H <sub>T</sub> ,H)	3.0	3.4 ± 0.02	0.4	3.1 ± 0.1	0.1
d(H <sub>A</sub> ,H <sub>T</sub> )	2.3	2.3 ± 0.01	0	2.3 ± 0.06	0
d(H <sub>A</sub> ,H)	2.8	2.7 ± 0.01	0.1	2.7 ± 0.1	0.1

<sup>a</sup> Experimental values from ref. [28,29], uncertainties not supplied.

autocorrelation function of the Legendre polynomials. Three vectors were defined; one from the C $\alpha$  atom to the center of mass of the sugar moiety, a second one from the peptide backbone atoms N to C, and the last one as the cross product of the two. For each of these vectors, the auto-correlation function is computed and fitted to an exponential function of

the form  $Ae^{-k/\tau} + c$ . After fitting, the decay constant was calculated as  $\tau = 0.5$  ns which is at least a factor 10 faster than for typical proteins.

Trajectories from unbiased MD simulations and LEUS simulations after unbiasing were checked with average interatomic distances, see Table 4. Deviations of the calculated averaged distances from the experimental NOE bounds of less than 1 Å are considered as insignificant. All the calculated averaged NOE distances are within 1 Å except for the HT–H distance in system 2 (see Figure S1 for the labeled system). As the HT to HA NOE seems to be fulfilled, the reason may be found in the distribution of the backbone angle. A shift of this angle towards values of  $-180^\circ$  may be sufficient to fulfill both the NOE and bring the high  $^3J_{\text{HNH}\alpha}$  in closer agreement with experiment (see Table 1 and Figure 6).

NMR studies of O-glycosylation resulting in a strong NOE for d(H<sub>A</sub>,H<sub>T</sub>) along with weak NOE for d(H<sub>T</sub>,H) have been suggested to indicate that the peptide backbone is taking an extended conformation.<sup>[70]</sup> The distance distributions from LEUS simulation after unbiasing are given in Figure 7. The data in Table 2 confirms further that the major populations are the joint  $\beta$  and P<sub>II</sub> conformations. This is also corroborated by the conformational preferences illustrated in Figure 5. Finally, the increased preference for  $\alpha$ -conformations described for system 4 may be reflected by the lack of experimental NOE's for d(H, H<sub>B</sub>), d(H<sub>T</sub>,H<sub>B</sub>) and d(H<sub>T</sub>,H<sub>N</sub>).

### 3.4. Hydrogen Bonding

The occurrence of H-bonds was reweighed to the unbiased ensemble for the LEUS simulations. H-bonds with occurrences larger than 2% are reported in Table 5.

**Table 5.** Hydrogen bond occurrences in LEUS simulations for systems 1–4. Reweighted percentages higher than 2% are reported.

H-bond type	System	System			
		1	2	3	4
sugar-peptide	HN2–O	4.7%	7.3%	–	–
	HN2–OX	5.2%	6.9%	–	–
	H–O5	–	–	2.6%	6.5%
	HT–O5	–	–	–	5.4%
peptide-peptide	HT–OX	2.4%	–	4.0%	–
sugar-sugar	HO4–O6	2.5%	2.7%	–	–

Hydrogen bonding between the amide proton from the N-acetyl group of the GalNAc unit (HN2) and the oxygen atom of carbonyl oxygens (O) or (OX) of Thr and Ser were observed in  $\alpha$ -linked systems but not in  $\beta$ -linked systems. This intermolecular hydrogen bonding pattern seems prominent as it was reported by several other groups as well.<sup>[16,71,72]</sup> The Ser systems (system 1 & 3) had significantly fewer sugar-peptide hydrogen bonds compared to systems with Thr (system 2 & 4) which was also observed in the NMR studies of Ref. [28].

For  $\beta$ -linked GalNAc systems, H-bonds are observed between backbone amide proton (H) or (HT) and ring oxygen

(O5). Also for these H-bonds, a higher occurrence was observed for Thr compared to Ser. Previously in the work of Mallajosyula et al.,<sup>[31]</sup> hydrogen bonding in the GalNAc systems was observed to be very low (<0.1 propensity) while 0.33 hydrogen bond occurrence was observed in the  $\beta$ -GlcNAc-Thr system between HO6 (sugar) and O (Thr). Also for others, a similar increase in hydrogen bonding between  $\beta$ -linked sugar and Thr observed in simulations of  $\beta$ -Glucopyranose,  $\beta$ -N-acetylglucosamine and  $\beta$ -Mannopyranose linked systems.<sup>[31]</sup> As an alternative to direct H-bonding between the sugar and the peptide backbone, water bridged H-bonds have been suggested.<sup>[68]</sup> High occurrences of water bridged H-bonds between H and O7 or OG and O5 were observed with 8 and 11% for system 4 while the corresponding direct H-bonds were not observed or not highly populated.

For the unglycosylated Ser and Thr, the only hydrogen bond (higher than 2%) was found between the amide proton at the C terminus (HT) and the oxygen at the N-terminus (OX), with 2.6 and 2.2%, respectively ( $i, i+2$ ). HT-OX was found in systems 1 and 3 among the glycosylated systems with 2.4 and 4.0%, respectively but not in systems with Thr. In addition to peptide-peptide H-bonding, the only significant intrasugar hydrogen bonding was HO4-O6 which was observed in  $\alpha$ -linked systems with 2.5% and 2.7% for systems 1 and 2.

Overall, the Thr systems display higher occurrence of H-bonding which might be an explanation of the superior efficiency of Thr glycosylation. Also, compared to N-glycosylation where the core-glycan is sticking out from the peptide backbone, O-GalNAc glycosylation of Thr affects the backbone conformation of the peptide through stronger interactions with the peptide which was also observed in the studies of Ref. [73].

## 4. Conclusions

In this work, we studied O-glycosylation by glycosylating the single-residue dipeptides (N-acetyl-Ser/Thr-N-methylamide) as they represent the simplest system. These model systems allowed us to comment on the direct effect of the glycosylation by eliminating the neighbouring residue effect on the backbone conformation.

By using LEUS as an enhanced sampling method it was possible to cover most of the energetically possible states. We have first investigated unglycosylated systems and reparameterized them according to the available experimental data with J-couplings and secondary structure propensities. Subsequently the four model systems with  $\alpha$  and  $\beta$ -linked O-glycosylation were compared to the NMR experimental findings of the same molecules. We conclude that the current forcefield parameters represent the J-values and NOE's without introducing any restraints. For  $^3J_{\text{HNH}\alpha}$  couplings, the biggest deviation was seen in system 2 with 1.1 Hz while the rest were about 0.3 Hz. For  $^3J_{\text{H}\alpha\text{H}\beta}$  couplings an increase in the population of the anti region in system 3 created a big deviation with 2.6 Hz where the Karplus curve is steep causing a significant change from a small difference in the dihedral angle.

Alternative experiments showed large variation in exactly this  $^3J$ -coupling, reducing the deviation with our calculations to

1.2 Hz. Furthermore, the trends in the  $^3J$ -values between systems were captured quite well. For the NOE's the only significant violation was seen in system 2 with 1.2 Å corresponding to the slight deviation in the  $^3J_{\text{HNH}\alpha}$  described above. From the propensities that we have calculated for each system we conclude that O-glycosylation seems to drive the peptide backbone to an extended conformation, with the exception of  $\beta$ -GalNAc-Thr where the  $\alpha$ -helical propensity remains relatively high. In general, Thr systems engage in a closer interaction with the peptide backbone by forming more stable hydrogen bonds.  $\beta$ -GalNAc-Thr displayed an overall shorter  $d(\text{H}_\tau, \text{H})$  distance than the unglycosylated Thr and an increased preference for  $\alpha$ -helical conformations compared to the other glycosylated systems. These slight shifts towards  $\alpha$ -helical conformations may be sufficient to explain the evolutionary preference of  $\alpha$ -linked GalNAc over  $\beta$ -linked GalNAc.

Overall, we conclude that the GROMOS force field describes the O-glycosylated systems very well in relation to the available experimental data from NMR.

## Acknowledgement

This work was supported by the Austrian Science Fund in the context of the Doctoral Program BioToP (Grant W 1224).

## Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** Glycosylation · Molecular dynamics · Glycopeptides · NMR spectroscopy

- [1] J. Eichler, M. Koomey, *Trends Microbiol.* **2017**, *25*, 662–672.
- [2] M. A. Aon, J. A. Curtino, *Biochem. J.* **1985**, *229*, 269–72.
- [3] I. R. Rodriguez, W. J. Whelan, *Biochem. Biophys. Res. Commun.* **1985**, *132*, 829–836.
- [4] K. Zarschler, B. Janesch, M. Pabst, F. Altmann, P. Messner, C. Schaffer, *Glycobiology* **2010**, *20*, 787–798.
- [5] R. D. Cummings, J. M. Pierce, *Handbook of Glycomics*, 1st ed., Academic Press/Elsevier, **2010**.
- [6] P. Van Den Steen, P. M. Rudd, R. A. Dwek, G. Opdenakker, P. V. den Steen, P. M. Rudd, R. A. Dwek, G. Opdenakker, *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 151–208.
- [7] I. Brockhausen, H. Schachter, P. Stanley, *O-GalNAc Glycans*, Cold Spring Harbor Laboratory Press, **2009**.
- [8] D. C. Gowda, E. A. Davidson, *J. Biol. Chem.* **1994**, *269*, 20031–9.
- [9] K. L. Carraway, S. R. Hull, *Glycobiology* **1991**, *1*, 131–8.
- [10] G. J. Strous, J. Dekker, *Crit. Rev. Biochem. Mol. Biol.* **1992**, *27*, 57–92.
- [11] S. Chakravarty, S. Kadunganattil, P. Bansal, R. K. Sharma, S. K. Gupta, *Mol. Reprod. Dev.* **2008**, *75*, 75–88.
- [12] K. E. Coyne, S. E. Hall, S. Thompson, M. A. Arce, T. Kinoshita, T. Fujita, D. J. Anstee, W. Rosse, D. M. Lublin, *J. Immunol.* **1992**, *149*, 2906–13.
- [13] J. Pieper, K. H. Ott, B. Meyer, *Nat. Struct. Biol.* **1996**, *3*, 228–32.
- [14] M. Oh-eda, M. Hasegawa, K. Hattori, H. Kuboniwa, T. Kojima, T. Orita, K. Tomonou, T. Yamazaki, N. Ochi, *J. Biol. Chem.* **1990**, *265*, 11432–5.
- [15] J. Chia, G. Goh, F. Bard, *Biochim. Biophys. Acta Gen. Subj.* **2016**, *1860*, 1623–1639.
- [16] D. M. Coltart, A. K. Royyuru, L. J. Williams, P. W. Glunz, D. Sames, S. D. Kuduk, J. B. Schwarz, X.-T. Chen, S. J. Danishefsky, D. H. Live, *J. Am. Chem. Soc.* **2002**, *124*, 9833–9844.

- [17] R. Apweiler, H. Hermjakob, N. Sharon, *Biochim. Biophys. Acta Gen. Subj.* **1999**, *1473*, 4–8.
- [18] T. Lütteke, C.-W. von der Lieth, *BMC Bioinf.* **2004**, *5*, 69.
- [19] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommelaeghe, T. J. Foster, F. W. Jamison, A. D. MacKerell, *J. Chem. Theory Comput.* **2011**, *7*, 3162–3180.
- [20] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–55.
- [21] L. Pol-Fachin, V. H. Rusu, H. Verli, R. D. Lins, *J. Chem. Theory Comput.* **2012**, *8*, 4681–4690.
- [22] L. Pol-Fachin, H. Verli, R. D. Lins, *J. Comput. Chem.* **2014**, *35*, 2087–2095.
- [23] W. Plazinski, A. Lonardi, P. H. Hünenberger, *J. Comput. Chem.* **2016**, *37*, 354–65.
- [24] K. Panczyk, K. Gaweda, M. Drach, W. Plazinski, *J. Phys. Chem. B* **2018**, *122*, 3696–3710.
- [25] C. A. López, A. J. Rzepliela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger, S. J. Marrink, C. A. López, A. J. Rzepliela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger, S. J. Marrink, *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
- [26] M. C. Small, A. H. Aytenfisu, F.-Y. Lin, X. He, A. D. MacKerell, *J. Comput.-Aided Mol. Des.* **2017**, *31*, 349–363.
- [27] B. L. Foley, M. B. Tessier, R. J. Woods, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 652–697.
- [28] F. Corzana, J. H. Busto, G. Jiménez-Osés, J. L. Asensio, J. Jiménez-Barbero, J. M. Peregrina, A. Avenoza, *J. Am. Chem. Soc.* **2006**, *128*, 14640–14648.
- [29] F. Corzana, J. H. Busto, G. Jiménez-Osés, M. García de Luis, J. L. Asensio, J. Jiménez-Barbero, J. M. Peregrina, A. Avenoza, *J. Am. Chem. Soc.* **2007**, *129*, 9458–9467.
- [30] A. Fernández-Tejada, F. Corzana, J. H. Busto, A. Avenoza, J. M. Peregrina, *J. Org. Chem.* **2009**, *74*, 9305–9313.
- [31] S. S. Mallajosyula, A. D. MacKerell, *J. Phys. Chem. B* **2011**, *115*, 11215–11229.
- [32] M. Yang, A. D. MacKerell, *J. Chem. Theory Comput.* **2015**, *11*, 788–799.
- [33] S. K. Mishra, M. Kara, M. Zacharias, J. Koca, *Glycobiology* **2014**, *24*, 70–84.
- [34] H. S. Hansen, X. Daura, P. H. Hünenberger, *J. Chem. Theory Comput.* **2010**, *6*, 2598–2621.
- [35] S. Kannan, M. Zacharias, *Proteins Struct. Funct. Genet.* **2007**, *66*, 697–706.
- [36] A. Turupcu, C. Oostenbrink, *J. Chem. Inf. Model.* **2017**, *57*, 2222–2236.
- [37] J. Gebhardt, C. Kleist, S. Jakobtorweihen, N. Hansen, *J. Phys. Chem. B* **2018**, *122*, 1608–1626.
- [38] N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger, W. F. van Gunsteren, *Comput. Phys. Commun.* **2012**, *183*, 890–903.
- [39] M. M. Reif, P. H. Hünenberger, C. Oostenbrink, *J. Chem. Theory Comput.* **2012**, *8*, 3705–3723.
- [40] R. D. Lins, P. H. Hünenberger, *J. Comput. Chem.* **2005**, *26*, 1400–1412.
- [41] Molecular Operating Environment (MOE) Chemical Computing Group, 1010 Sherbooke St. West Suite #910; Montreal; QC; Canada; H3A 2R7, **2013**.
- [42] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, H. J. in *Intermol. Forces*, Reidel, Dordrecht, **1981**, pp. 331–342.
- [43] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [44] R. W. Hockney, *Methods Comput. Phys.* **1970**, 136–211.
- [45] J. P. Ryckaert, G. Cicciotti, H. J. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
- [46] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 5451–5459.
- [47] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger, *J. Chem. Phys.* **2001**, *115*, 1125–1136.
- [48] T. N. Heinz, P. H. Hünenberger, *J. Comput. Chem.* **2004**, *25*, 1474–1486.
- [49] A. P. Eichenberger, J. R. Allison, J. Dolenc, D. P. Geerke, B. A. Horta, K. Meier, C. Oostenbrink, N. Schmid, D. Steiner, D. Wang, W. F. van Gunsteren, *J. Chem. Theory Comput.* **2011**, *7*, 3379–3390.
- [50] C. Margreitter, C. Oostenbrink, *J. Chem. Inf. Model.* **2016**, *56*, 1823–1834.
- [51] J. Grdadolnik, V. Mohacek-Grosec, R. L. Baldwin, F. Avbelj, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1794–8.
- [52] H. S. Hansen, P. H. Hünenberger, *J. Comput. Chem.* **2010**, *31*, 1–23.
- [53] T. Huber, A. E. Torda, W. F. van Gunsteren, *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.
- [54] W. F. van Gunsteren, J. R. Allison, X. Daura, J. Dolenc, N. Hansen, A. E. Mark, C. Oostenbrink, V. H. Rusu, L. J. Smith, *Angew. Chem. Int. Ed.* **2016**, *55*, 15990–16010; *Angew. Chem.* **2016**, *128*, 16222–16244.
- [55] C. Peter, X. Daura, W. F. van Gunsteren, *J. Biomol. NMR* **2001**, *20*, 297–310.
- [56] G. Chalmers, J. Glushka, B. Foley, R. Woods, J. Prestegard, *J. Magn. Reson.* **2016**, *265*, 1–9.
- [57] W. F. van Gunsteren, J. R. Allison, X. Daura, J. Dolenc, N. Hansen, A. E. Mark, C. Oostenbrink, V. H. Rusu, L. J. Smith, *Angew. Chem. Int. Ed.* **2016**, *55*, 15990–16010; *Angew. Chem.* **2016**, *128*, 16222–16244.
- [58] X. Daura, W. F. van Gunsteren, A. E. Mark, C. Oostenbrink, S. Riniker, L. Smith, *Angew. Chem. Int. Ed.* **2017**, DOI 10.1002/anie.201702945.
- [59] L. J. Smith, W. F. van Gunsteren, N. Hansen, *J. Phys. Chem. B* **2017**, *121*, 7055–7063.
- [60] G. W. Vuister, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 7772–7777.
- [61] M. Mobli, A. Almond, *Org. Biomol. Chem.* **2007**, *5*, 2243–51.
- [62] A. Demarco, M. Llinás, K. Wüthrich, *Biopolymers* **1978**, *17*, 637–650.
- [63] X. Daura, W. F. van Gunsteren, N. Hansen, A. E. Mark, C. Oostenbrink, S. Riniker, L. Smith, *Angew. Chem. Int. Ed.* **2017**, DOI 10.1002/anie.201702945.
- [64] R. U. Lemieux, S. Koto, *Tetrahedron* **1974**, *30*, 1933–1944.
- [65] S. Pérez, R. H. Marchessault, *Carbohydr. Res.* **1978**, *65*, 114–120.
- [66] H. Thøgersen, R. U. Lemieux, K. Bock, B. Meyer, *Can. J. Chem.* **1982**, *60*, 44–57.
- [67] N. Martínez-Sáez, J. Castro-López, J. Valero-González, D. Madariaga, Compañón, V. J. Somovilla, M. Salvadó, J. L. Asensio, J. Jiménez-Barbero, A. Avenoza, J. H. Busto, G. J. L. Bernardes, J. M. Peregrina, R. Hurtado-Guerrero, F. Corzana, *Angew. Chem. Int. Ed.* **2015**, *54*, 9830–9834; *Angew. Chem.* **2015**, *127*, 9968–9972.
- [68] I. A. Bermejo, I. Usabiaga, I. Compañón, J. Castro-López, A. Insausti, A. Fernández, A. Avenoza, J. H. Busto, J. Jiménez-Barbero, J. L. Asensio, J. M. Peregrina, G. Jiménez-Osés, R. Hurtado-Guerrero, E. J. Cocinero, F. Corzana, *J. Am. Chem. Soc.* **2018**, *140*, 9952–9960.
- [69] F. Avbelj, S. G. Grdadolnik, J. Grdadolnik, R. L. Baldwin, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1272–7.
- [70] K. Wüthrich, *NMR of proteins and nucleic acids*, John Wiley & Sons, New York, **1986**.
- [71] Y. Mimura, Y. Inoue, N. J. Maeji, R. Chujo, *Int. J. Pept. Protein Res.* **1989**, *34*, 363–368.
- [72] Y. Tachibana, K. Monde, S.-I. Nishimura, *Macromolecules* **2004**, *37*, 6771–6779.
- [73] L. Otvos, G. R. Krivulka, L. Urge, G. I. Szendrei, L. Nagy, Z. Q. Xiang, H. C. Ertl, *Biochim. Biophys. Acta* **1995**, *1267*, 55–64.

---

 Manuscript received: January 24, 2019

Revised manuscript received: March 27, 2019

Version of record online: May 6, 2019