COMPUTATIONAL STUDIES OF PROTEIN POSTTRANSLATIONAL MODIFICATION: GLYCOSYLATION OF OLIGOPROLINE AND COLLAGEN PEPTIDES

EMMANUEL B. NAZIGA
BSc Honours, University of Port Harcourt, 2004
PGD, African Institute for Mathematical Sciences, 2006
MSc, Ecole Normale Supérieure de Lyon, 2007
MSc, Sapienza - Università di Roma, 2007

A Thesis Submitted to the
School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements of the Degree

DOCTOR OF PHILOSOPHY

Department of Chemistry and Biochemistry
The University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

Dedication

For Monne, Nasibezor, Tuka, Tonbari, Nubari, Tordi, Eyefaa and Dumle

Abstract

Glycosylation is the most complex posttranslational modification of proteins and has consequences on protein structure and function. In particular, the hydroxyproline (Hyp) rich glycoproteins (HRGPs) of plants are heavily glycosylated. On the other hand, glycosylation has not been observed in animal collagen despite the high occurrence of Hyp residues. This thesis uses computational chemistry to provide molecular level information about the structural effects of Hyp glycosylation to help understand the biological implications of the modification and explain the lack of glycosylation in animals. Initially, the nature of the glycosidic linkage between Hyp and galactose was determined. The theoretical results were validated by comparing to the recent experimental data, which helped understand other experimental observations. Subsequently, contiguous and non-contiguous glycosylation of a nonaproline oligopeptide was considered, which revealed that contiguous glycosylation increases the stability of the all *trans* polyproline II (PPII) conformation, while non-contiguous glycosylation leads to loss of PPII content. Sophisticated modeling suggested that this difference arises since peptide–solvent interactions stabilize the PPII conformation in the contiguously glycosylated peptide, while sugar–peptide backbone interactions that stabilize the *cis* conformations of some residues are stronger in the non-contiguously glycosylated peptide. Finally, the effects of Hyp glycosylation on the collagen triple helix were assessed, where it was determined that glycosylation makes the monomeric state more stable and hence hinders triple helix formation, which agrees with experimental results and highlights that the synergy between computation and experiments is necessary to understand complex glycosylation in nature.

Acknowledgements

I would like to thank my supervisor Dr. Stacey D. Wetmore for giving me the opportunity to study for a PhD under her supervision and for all I have learnt over the course of the last three and half years. Thank you to the members of my supervisory committee, Dr. Rene Boere, Dr. Kenneth Vos, and Dr. Paul Hazendonk for their support and guidance throughout this period. I also thank Dr. Frank Schweizer my experimental collaborator, Dr. Peter Kusalik for been the external examiner for my thesis defense and Dr. Peter Dibble for chairing my thesis examination committee.

Funding for my studies came from the School of Graduate Studies of the University of Lethbridge and Natural Sciences and Engineering Research Council (NSERC) for which I am very grateful. Additionally, I thank URACIL and WestGrid-Compute/Cacul Canada for computer resources.

Thank you to Dr. Lesley Rutledge, Dr. Andrea Millen and Jennifer Kellie for introducing me to research in the Wetmore lab and that welcome lunch at the Zoo on my first day on campus. I would also like to appreciate the support I received from all other members of the Wetmore lab including Dr. Purshotam Sharma, Mohadeseh Majdi Yazdi, Stefan Lenz and Katie Wilson.

Over the past three plus years I have come to make many friends who enriched my time in Lethbridge. I would therefore like to thank Ibukun, Saturday, Buchi, Okeke, Blessing, Chuchu, Stephen, Kenneth, Jonathan, Ijay, Dupe, Sandrine, Joyce, Michel, Dawn, Michelle, Femi, Titi and Chris for your friendship. Thanks too to the entire membership of the Redeemed Christian Church of God, Lethbridge.

Lastly but by no means lease, I am very grateful to my family for all their support, prayers and love.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

°	degree
Å	angstrom
ABMD	adaptively biased molecular dynamics
Ac	acetyl
AG	arabinogalactan
AGP	arabinogalactan protein
Ala	alanine
Ara	arabinose
Asn	asparagine
Azp	azidoproline
BSSE	basis set soperposition error
CD	circular dichroism
CG	conjugate gradient
CMP	collagen model peptide
CV	collective variable
DFT	density functional theory
ER	endoplasmic recticulum
Flp	4R-fluoroproline
Gal	galactose
Glc	glucose
Gly	glycine
HRGP	hydroxyproline rich glycoprotein
HT-REMD	Hamiltonian-temperature replica exchage molecular dynamics
Hyl	hydroxylysine
Hyp	4R-hydroxyproline
LJ	Lennard Jones
LMOD	low mode
Lys	lysine
Man	mannose
MD	molecular dynamics
MM	molecular mechanics
MM-RISM	molecular mechanics-reference interaction site model
NMR	nuclear magnetic resonance
NPT	constant number of particles, pressure and temperature
ns	nanosecond
NVT	constant number of particles, volume and temperature
OH	hydroxyl
OMe	methoxy
PDB	protein databank
PMF	potential of mean force

PPI	polyproline I
PPII	polyproline II
Pro	proline
PRP	proline rich protein
ps	picoseconds
PTM	post-translational modification
RDF	radial distribution function
REMD	replica exchange molecular dynamics
SD	steepest decent
SE	Schrodinger equation
Ser	serine
Thr	threonine
T_m	melting temperature

Chapter 1. Introduction

1.1 General Background

Proteins are essential for the normal functioning of all living organisms. They are involved in virtually every kind of cellular activity, including biochemical catalysis, immune response and cell–cell communication. As such, significant resources at the cellular level are devoted to protein generation. Briefly, protein biosynthesis requires transcription of genetic code to yield information about the amino acid units that compose a particular protein, and subsequent translation of this data to build a polypeptide chain, which usually folds into a distinct, well-defined three dimensional structure.[1] It is in this so-called native structure that proteins perform their far-reaching functions.

For some proteins, posttranslational modifications (PTMs) can take place prior to, or after folding. PTMs are specific chemical alterations to the amino acids in the protein polypeptide chain. Example PTMs include disulphide bridges, hydroxylation, glycosylation, methylation and phosphorylation. Glycosylation, which involves covalent attachment of a carbohydrate molecule, is the most structurally complex protein PTM and has been implicated in many cellular processes including signalling and protein folding.[2-3]

However, experimental determination of the three dimensional structure of glycopeptides by X-ray crystallography can be more challenging than those of unglycosylated proteins due to the inherent conformational flexibility of the covalently-linked carbohydrate moieties.[4-5] Additionally, while nuclear magnetic resonance (NMR) spectroscopy is extensively used to study solution chemistry,[6] interpretation of experimental results can become complicated by fast conformational dynamics in glycoproteins which affects fitting procedures used in data analysis.[7-8] As a consequence, detailed molecular information regarding how glycosylation affects protein structure and function is often unavailable.

In order to address the significant gap in our knowledge regarding the role of glycosylation PTMs, this thesis will use advanced molecular modeling techniques to study how glycosylation affects the structure of smaller peptide fragments. In what follows, a brief overview of carbohydrate structure is given, the importance of protein glycosylation and putative functions are summarized, and currently identified carbohydrate–protein linkages are described. Thereafter, the specific type of glycosylation (glycopeptides) considered in the present thesis will be discussed.

1.2. Oligosaccharides

Carbohydrates are comprised of monosaccharides (Figure 1.1), which are derived from polyhydroxy aldehydes or ketones. The hexoses ($C_6H_{12}O_6$) have 16 isomers, including galactose (Gal), glucose (Glc), mannose (Man), and arabinose (Ara). Cyclization of the open form leads to either a six membered pyranose or a five membered furanose depending on the position of the attacking hydroxyl (OH) group. In either case, a new chiral carbon is generated and an anomeric predisposition (α or β) that is dictated by the orientation of the OH group.

Polysaccharides are formed by a condensation reaction between two monosaccharide OH groups, which results in a glycosidic linkage. Due to the availability of several OH groups on each saccharide, there are a variety of positions for carbohydrate chain growth, which enables the formation of highly complex structures and contributes to the rich and diverse functionality of glycoproteins. An example of disaccharide formation involving an α(1–4) linkage between α-D-galactose and a methyl–α–D–galactose is illustrated in Figure 1.2.

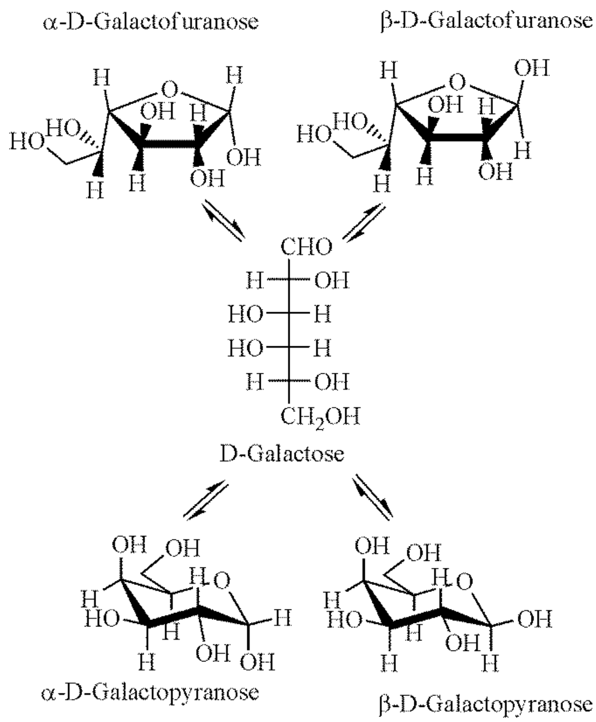

Figure 1.1. Cyclic isomers of D-galactose: cyclization can occur at different positions on the open chain, giving rise to a five-membered furanose structure (top) or a six-membered pyranose form (bottom). Two anomers (α or β) of each ring are possible.

1.3. Protein Glycosylation

It has been determined that over half of eukaryotic proteins undergo post–translational glycosylation, indicating that this type of PTM has far reaching consequences on protein structure and function.[9] In particular, glycosylation has been implicated in protein folding and stability.[10-14] For example, a recent publication revealed that all three asparagine (Asn) residues in the predicted glycosylation sites (positions 127, 284 and 516) of a disulfide isomerase (PDIA2) are modified, and that mutation of Asn284 to glutamine (Gln) increases dimer formation in PDIA2.[11] In addition, it has been demonstrated that glycosylation reduces the efficacy of proteolytic degradation enzymes.[3,15] Specifically, the collagen binding activity of a bacterial parasite is reduced upon deglycosylation, indicating consumption by proteolytic enzymes.[15] Protein-carbohydrate recognition is another important function associated with glycosylation.[16-18] This has been recently demonstrated by the binding specificity of Hevein, a model lectin, to the core Man(GlcAc)$_2$ trisaccharide that is very prevalent in eukaryotic glycoproteins.[19] These very diverse functions of glycosylation are dictated in part by the structural complexity of the attached glycans.

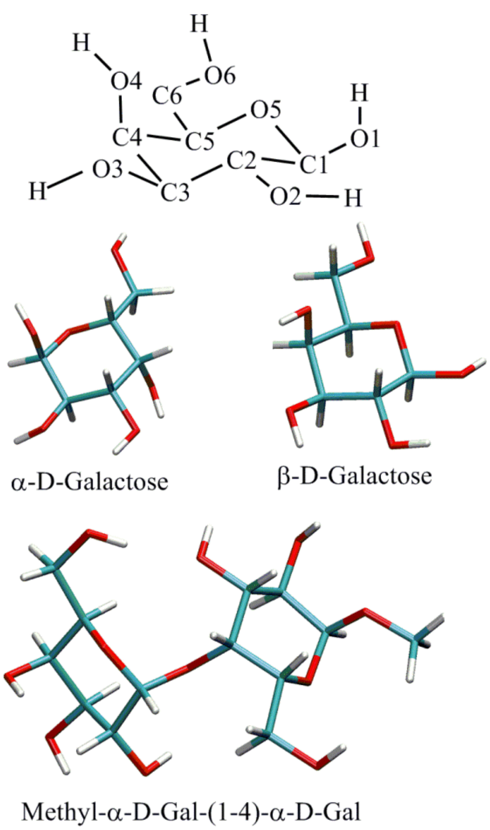

Figure 1.2. Top: Chemical sketch of a galactose monomer with atomic numbering. Middle: Models of α- and β-D-galactose monosaccharides. Bottom: Example disaccharide with an α(1–4) glycosidic linkage.

1.3.1. Types of Carbohydrate–Protein Covalent Linkages

Two major modalities exist for oligosaccharide–protein covalent association, namely N or O-linked glycosylation depending on whether the moieties are bridged by a nitrogen or oxygen atom (Figure 1.3). In N-glycosylation, which takes place in the endoplasmic reticulum (ER) of the cell, a fully synthesized oligosaccharide ($Glc_3Man_9GlcNAc_2$) is transported to the unfolded protein.[20] This type of glycosylation is sequence specific and the oligosaccharide is only attached to asparagine residues contained in an Asn–X–Ser/Thr amino acid sequence, where X is any residue other than proline (Pro). Subsequently, the core oligosaccharides are processed by glucosidases, mannosidases and/or glycosyltransferases to add or remove sugar monomers as neccessary.[20] Glucosidases and mannosidases affect the removal of some monomers of the core saccharides in the Golgi apparatus, while other sugars are added by the action of various glycosyltransferases.[20]

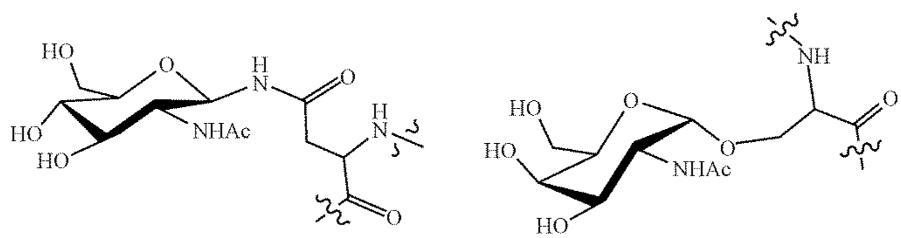

Figure 1.3. Examples of naturally occurring N-acetylglucosamine glycosidic linkage to asparagine (left) and serine (right).

Recently, high resolution crystal structures for a series of N-glycosylated compounds (Figure 1.4) were obtained using neutron diffraction.[21] In contrast to previous X-ray structure determinations,[22] hydrogen atom positions could be located in the neutron diffraction study. This allowed for detailed examination of intramolecular hydrogen bonding in the compound. Contrary to previous literature proposals, strong hydrogen bonding between adjacent exocyclic hydroxyl groups were not observed. Additionally, as expected the $^4C_1$ chair conformation was observed to be the preferred for the sugar. In agreement with previous computational and crystal structures,[22] the key glycosidic linkage dihedral angle (ϕ_N, Figure 1.4) has a preferred value of –90°.

O-glycosylation, on the other hand, usually takes place in the cell cytoplasm after protein folding.[20] In this case, a core monosaccharide is attached to the protein by a specific glycosyltransferase. While there is no consensus regarding the preferred sequence for O-glycosylation, the linkage mainly occurs to serine (Ser), threonine (Thr) and lysine (Lys) residues in animals.[23] In plants, a prominent glycosylation involves β–glycosidic linkages of (2*S*,4*R*)–hydroxyproline (Hyp) to either L–arabinose or D–galactose, particularly in hydroxyproline rich glycoproteins (HRGPs).[24] The degree of glycosylation ranges from mono or disaccharide units (as observed in collagen)[23] to highly complex polysaccharides (as observed in HRGPs).[25] The large variation in the final size of the attached glycans underscores the complexity of carbohydrate–protein interactions and the heterogeneity of glycan function. Additionally, glycosylation can be contiguous, where all consecutive Hyp residues in a given protein segment are glycosylated, or non-contiguous, which allows for unmodified Hyp residues in-between the modified variants.[26] To begin to understand these complex structure–function relationships, this thesis will focus on the structural details and implications of O-glycosidic linkages to hydroxyproline.

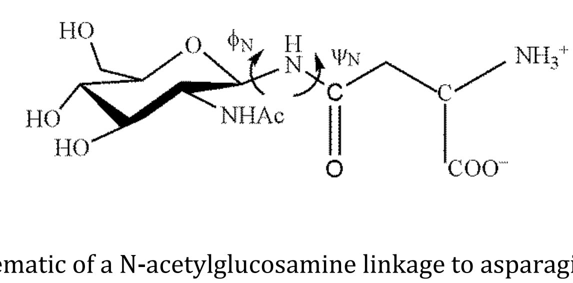

Figure 1.4. Schematic of a N-acetylglucosamine linkage to asparagine determined by neutron diffraction.

1.4. Hydroxyproline-O-Glycosylation

Glycosylation of Hyp is of particular interest since Pro analogs exhibit properties unique among the proteinogenic amino acids. Specifically, Pro residues have limited rotation about the ϕ (∠(C–C$^{\alpha}$–N–C), Figure 1.5) dihedral angle because their side chains are fused to the peptide backbone. As a result, there is a reduction in the energy difference between the *cis* (ω = 0°) and *trans* (ω = 180°) isomers, which makes both conformations nearly isoenergetic according to many experimental and computational results, [27-31] and leads to a higher *cis* isomer content compared with other amino acids.

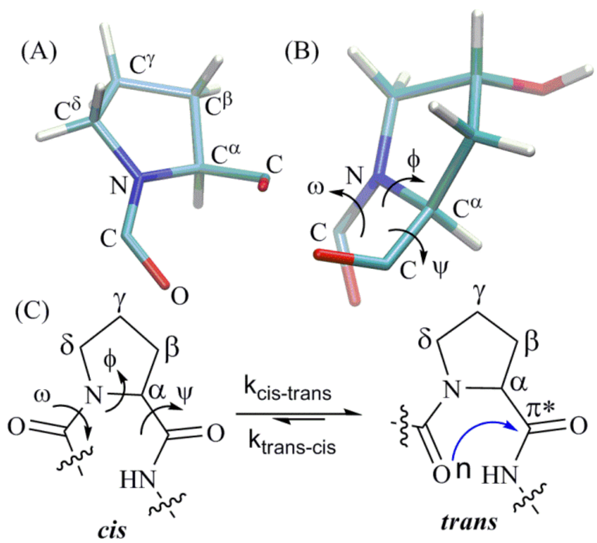

Figure 1.5. (A) A proline residue with atomic labels, (B) model of a hydroxyproline residue and (C) *cis*–*trans* isomerization of proline, including important dihedral angle definitions: ϕ (∠(C–C$^{\alpha}$–N–C), ψ (∠(N–C$^{\alpha}$–C–N), ω (∠(C$^{\alpha}$ –C–N– C$^{\alpha}$) (left) and the n → π* interaction (right).

Due to the above structural feature, Pro and derivatives perform some distinct functions in proteins and peptides. For instance, pH jump kinetic studies show that prolyl *cis*/*trans* isomerization is often the rate-determining step in the folding pathways of many peptides (proteins),[27,32] Pro residues induce turn structures that are crucial in protein–protein interactions,[33-34] and facilitate the formation of extended helical structures such as the polyproline II (PPII) helix (Figure 1.6).

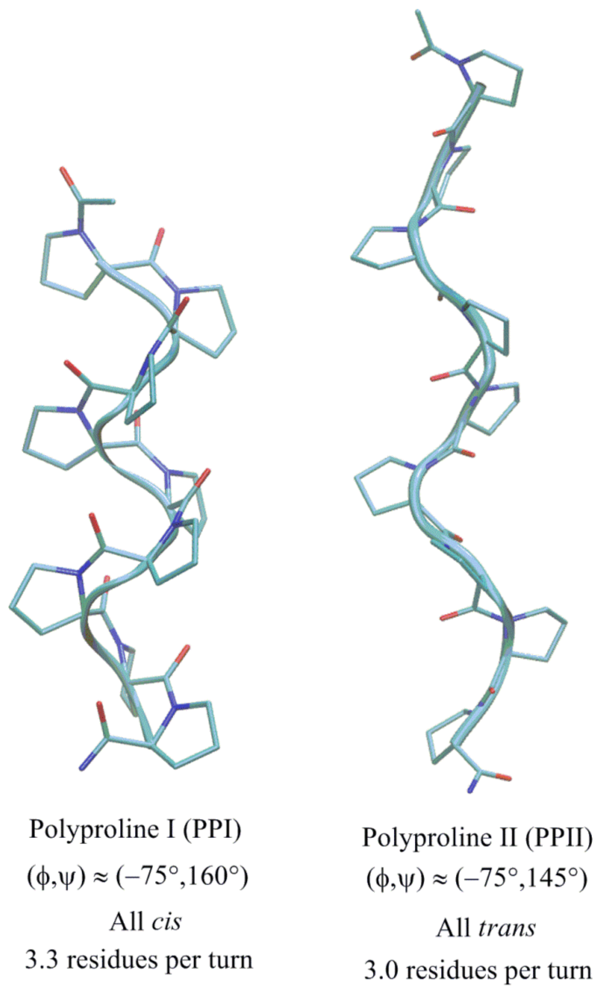

Figure 1.6. Polyproline I (PPI) and polyproline II (PPII) helices of a nonaproline model compound (hydrogen atoms removed for clarity).

The PPII structure is observed in Pro rich regions of proteins and play an important role in the stability of structural proteins, such as collagens[35-36] and HRGPs,[37-38] which are known to require the PPII-helical conformation to function. In fact, the left–handed PPII conformation with $(\phi,\psi) \approx (-75°,146°)$ (Figure 1.6) and all *trans* ω dihedral angles has emerged as an important secondary structure. In particular, this conformation which occurs even in non-Pro segments of peptides and proteins, is dominant in unfolded proteins and oligoproline segments, and has been identified from crystal structures as a prominent conformation of ligands for signalling proteins.[39-40] The related PPI conformation occurs when $(\phi,\psi) \approx (-75°,160°)$ and all *cis* ω dihedral angles are adopted, is typically present at high temperatures or in the presence of certain solvents, such as methanol and 1-propanol.[41]

The properties of Pro-based proteins described above depend on the relative stability of the *cis* and *trans* isomers of Pro, which are in turn governed by several factors,[42-43] including an $n \rightarrow \pi^*$ interaction between the oxygen lone pair in the prolyl *N*-terminal amide C=O and the antibonding orbital of the C-terminal C=O (Figure 1.5C).[27,42,44-48] This $n \rightarrow \pi^*$ interaction occurs when the lone pair on the oxygen (n) donates electron density to the antibonding (π^*), which occurs only in the trans isomer of Pro and derivatives. Additionally, inductive and stereoelectronic effects influence the cis/trans stability of Pro derivatives.[48-53] For example, naturally occurring 4-hydroxylation and attachment of other electron-withdrawing groups have a further influence on *cis*–*trans* isomerization by affecting the $n \rightarrow \pi^*$ interaction through changes in the conformation of the pyrrolidine ring (puckering) and the prolyl backbone ψ ($\angle$(N–C$^{\alpha}$–C–N) dihedral angle.[54] In plant and animal proteins, the irreversible transfer of a hydroxyl group to the C$^{\gamma}$ atom of Pro is catalysed by a prolyl 4–hydroxylase to give (2*S*,4*R*)–hydroxyproline (or Hyp),[55-57] which is the most prevalent PTM. The (2*S*,4*S*)–hydroxyproline (hyp) stereoisomer is not very common, but has been extracted from several species of fungi and the cyanobacteria *Lyngbya majuscula*.[58]

Therefore, hydroxylation leads to increased stability of the *trans* isomer that is characteristic of the PPII conformation and positions the residue for glycosylation.

Covalent addition of complex sugar moieties to Hyp occurs in HRGPs that form part of plant cell walls,[24-25] and the glycans have been implicated in various functions of the protein.[59-60] HRGPs can be divided into three main classes: extensins, proline rich proteins (PRP) and arabinogalactan glycoproteins (AGPs).[61-63] Extensins are known to be important for cell wall assembly.[60,64] PRPs and extensins are mostly glycosylated with arabinose sugars, although PRPs are much less glycosylated compared to extensins.[24,65] AGPs are associated with several cell functions, where their possible function as structural glycoproteins has been recently demonstrated.[65-66] Apart from their biological functions, APGs form a major component of the gum extract from plants that are used as emulsifiers, binders and stabilizers in the food, mining and pharmaceutical industries.[67-68] They are highly glycosylated with arabinogalactan (AG) polysaccharides[69] consisting of between 30 to 150 monosaccharides.[25,63] The AG units consist of (1–3)-linked β–D–galactosyl backbones, with branching through (1–6)–linked β–D–galactose units that are terminated by arabinose or rhamnose sugars.[65-66,68] The carbohydrate–protein attachment in AGPs occurs via O-galactosyl linkages (Figure 1.7), and is the primary focus of this thesis.

AGPs are generally considered as either classical or non-classical.[70-71] Classical AGPs have a protein core consisting mainly of hydroxyproline as well as alanine, serine and glycine amino acids, while non-classical AGPs exhibit a large variation in the residues of the core protein. The high Hyp content of AGPs suggests that the protein components are helical due to the propensity of Pro derivatives to adopt the PPII conformation. Indeed, several investigations[59,66] have indicated that AGPs (and HRGPs in general) adopt the PPII conformation. This is based on circular dichroism (CD) measurements, which show a minimum in the measured molar ellipticity at around 205 nm and a maximum at around 225 nm, which are characteristic of the PPII conformation. Nevertheless, more heterogeneous portions of HRGPs exist, which have been proposed to permit kinks in the protein structure and thereby allow interactions with other molecules.[59]

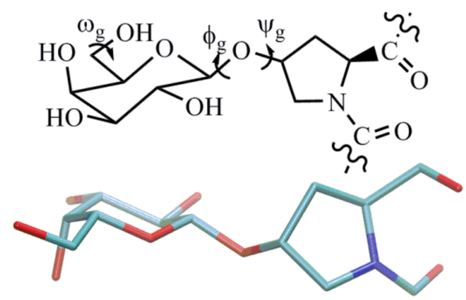

Figure 1.7. Sketch of a galactosylated hydroxyproline residue (top) and a model glycopeptide (bottom). Hydrogen atoms removed for clarity.

As mentioned in Section 1.1, it is experimentally difficult to obtain information regarding the three dimensional structure of HRGPs. In fact, no crystal structure of a glycosylated HRGP has been resolved to date. This lack of structural data is mainly due to the complexity of polysaccharides attached to biological HRGPs that complicate NMR spectra, the experimental difficulties associated with structure elucidation of the highly mobile glycans via X-ray crystallography, which is additionally complicated by the complex glycans shielding the protein backbone.[61] Therefore, a synthetic approach, involving controlled development and characterization of glycosylated Hyp compounds is used to provide valuable information regarding this important plant protein.

Along these lines, Schweizer and coworkers synthesized model glycosylated Hyp peptides[72-74] and provided quantitative data that supports previous claims that glycosylation increases the stability of the PPII structure. First, in their initial study, a model O-glycosylated (2*S*,4*R*/*S*)–Hyp residue was used to show that glycosylation slightly enhances the *trans* isomer population of the (2*S*,4*S*)–Hyp peptide. This was based on measured $K_{cis/trans}$ of 2.4±0.1 for unglycosylated (2*S*,4*S*)–Hyp compared to 2.9±0.3 (2.9±0.1) for the α(β) glycosylated variants in D_2O at 24.8°C. An identical value for $K_{cis/trans}$ (2.4±5%) was previous measured by Bretscher et al[48] for (2*S*,4*S*)–Hyp at in D_2O 25°C. Nuclear Overhauser (nOe) inter-proton distance measurements by Schweizer and coworkers showed that distant contacts exist between the galactose and Hyp, which implies that glycosylation could have a stabilizing effect on longer peptides.[73] Second, when a capped nonaproline was studied, a significantly higher melting temperature (T_m) was observed for the contiguously glycosylated peptide (70°C) compared to a fully hydroxylated peptide (38°C).[74] T_m which correlates with the stability of the PPII conformation was obtained by taking the first derivative of the change in molar ellipticity at 225 nm over a temperature range (see Chapter 2 for details). However, the structural/molecular basis of the reported observations could not be fully explained. Therefore, more studies are required to explain these experimental findings.

Despite the lack of specific structural information as described above, two AGP structural models have been proposed in the literature: the “wattle blossom” model[75] and the “twisted hairy rope” model (Figure 1.8).[76] The “wattle blossom” model entails a globular glycan pointing away from the helical core protein. In contrast, the “twisted hairy rope”, describes an extended peptide backbone with glycans arranged along, and interacting with, the backbone. However, recent work in the Kieliszewski lab[65-66] has revealed the primary sequence of select AGP polysaccharides, and shown that the glycan structure is not as complex as generally assumed. Specifically, the AG polysaccharides extracted from fusion glycoproteins expressed in tobacco cells were found to have 14–22 monomeric sugars per polysaccharide.[65] Molecular mechanics minimizations revealed a compact glycan structure that interacts with the peptide backbone, although in a more complex manner than the simple "twisted hairy rope" model implies.

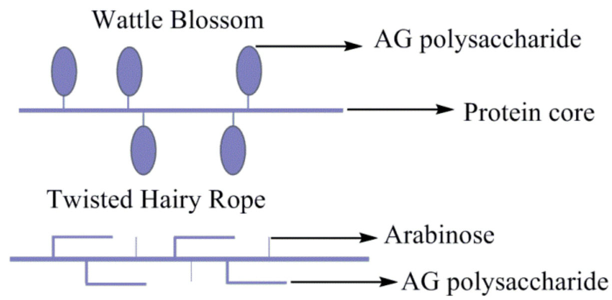

Figure 1.8. Proposed structural models for arabinogalactan proteins. The "wattle blossom" model depicts a glycopeptide structure having globular sugars that point away from the peptide backbone, while the "twisted hairy rope" model has the sugars in a parallel-like fashion and interacting with the protein backbone.

Clearly, further information about the structure of the Hyp-O-glycosidic linkage and how Hyp glycosylation affects protein (peptide) structure are required. This thesis aims to help alleviate this lack of information in the literature. This is done by carrying out detailed systematic molecular modeling studies of the naturally occurring Hyp-O-galactosylated peptides that model the HRGPs of plants as well as examine the natural the lack of glycosylation in the Hyp rich collagen of animals. The next section will consider glycosylation of collagen, a Hyp-rich protein that occurs in animals and plays a structural role similar to the HRGPs of plants described in this section.

1.5. Collagen O-Glycosylation

Collagen consists of three protein (peptide) strands that adopt the PPII conformation and are wound around each other to form a triple helix and held together by interstrand (C=O···H–N) hydrogen bonds (Figure 1.9).[77-80] Most collagen monomer strands exhibit a distinctive X–Y–glycine (Gly) amino acid sequence repeat (X, Y are other amino acid residues) and each residue in the repeat has a specific function. For example, it is well known that Gly residues are necessary for collagen stability. Melting temperatures from experimental mutation studies,[81-82] as well as energetics from computational studies[83-87] where Gly is replaced by other amino acids show significant triple helix destabilization. This is caused by a disruption of interstrand hydrogen bonds by the bulkier amino acids. In fact, some of the most severe consequences of collagen mutation occur when Gly is replaced by another amino acid.[88] This is because as evidenced by crystal structures, the close packing of the triplex helix requires glycine, which is the least sterically demanding amino acid, at every third position, in order to accommodate the three residues per turn necessary for the PPII structure and triple helical association. The X and Y positions in the triple helix sequence repeat have been found to be occupied by proline (Pro) derivatives in a large number of collagens.[89-90] Indeed, the X position frequently contains the unmodified Pro residue. This is not surprising since Pro is naturally predisposed to the PPII conformation due to its unique structure.

Like the HRGPs, collagens are structural proteins whose biological function is closely associated with maintaining a triple helical structure. There is a high level of Hyp occupation of the Y position in the collagen tripeptide repeat, which has been shown to correlate with the stability of synthetic collagen model peptides (CMPs). Consequently, numerous experimental (T_m and ΔG measurements) and computational studies have examined the effect of proline hydroxylation on the stability of the collagen quaternary structure.[48,50,77,81-83,91-111]

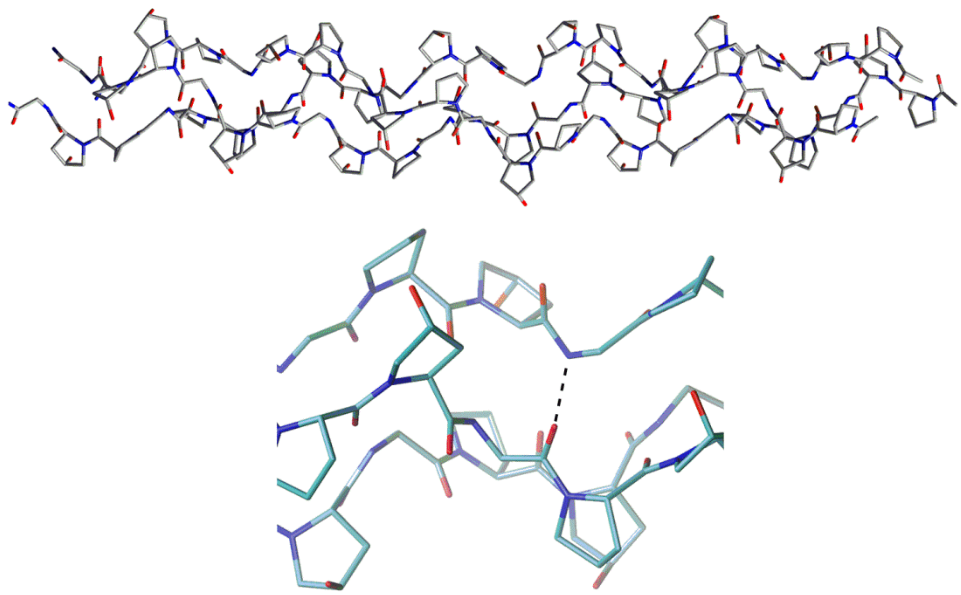

Figure 1.9. (Top) A model of the collagen triple helix and a closed up view (bottom) showing an interstrand backbone–backbone (C=O···H–N) hydrogen bond (dashed lines). Hydrogen atoms are not shown for clarity.

Although there is still debate in the literature regarding the exact mechanism by which Hyp stabilizes collagen, many consequences of hydroxylation have been established.[48,103,112] Particularly, Hyp in the X position greatly reduces collagen stability if the Y position is not occupied by another Hyp, with Hyp only observed in the Y position in vertebrate collagen.[113] Hyp is destabilizing in the X position since it prefers the C^{γ}-exo pucker which has the backbone dihedral angles that are not optimal for the X position in the triple helix.[89] However, certain collagen obtained from deep sea hypothermal vent worms have Hyp in the X position and still maintain stable triple helices.[114] A major difference between this kind of collagen and most animal collagen is the high content of glycosylated threonine (Thr) residues. For example, synthetic triple helices made from the Ac–(Gly–Pro–Thr)$_{10}$-NH_2 and Ac–(Gly–Pro–Thr(β–Galactose))$_{10}$–NH_2 peptides, based on cuticle collagen of a deep sea hydrothermal vent worm, were developed by Bann and co-workers to investigate Thr glycosylation.[94] It was concluded that triple helix formation requires glycosylation since a high melting temperature (T_m = 41°C) was measured using CD spectroscopy for the glycosylated peptide, while the unmodified compound was unable to form a triple helix. Possible explanations of how glycosylation of threonine leads to viable collagen triple helices include restricting procollagen conformational space and favourable inter-strand dipole-dipole interactions, as well as stabilization due to (intra)intermolecular hydrogen bonding between the sugar moiety and the peptide backbone.

Interestingly, lysine (Lys) residues in human collagen can be further processed into hydroxylysine (Hyl), which can be glycosylated with a galactose–glucose disaccharide.[115] This is in contrast to the Hyp-O-glycosylation observed in HRGPs. In fact, Hyp glycosylation has not been observed in any animal protein, regardless of the similar structural roles played by collagen and HRGPs and the abundance of Hyp residues in animal collagen. It is therefore important to understand the reason for this conspicuous absence and, hence, the structural consequences of Hyp glycosylation in collagen.

Given the experimental difficulties associated with crystallizing biological collagen, synthetic collagen model peptides (CMPs) are usually employed to study collagen structure–function relationships.[99,116-118] In this light, a CMP with three galactosylated Hyp residues was developed by Wennemmers and co-workers and used to show that glycosylation modestly decreases the T_m (7°C) in this compound.[99] However, only the effect of a single glycosylation per strand was considered. Therefore, to provide a comprehensive picture of the effects of Hyp glycosylation on collagen triple helix stability, CMPs with varying degrees of glycosylation were synthesized by Schweizer and co-workers.[119] In particular, the Ac–(POG)$_7$–NH$_2$, Ac–(POG)$_3$–(PO$^\beta$G)–(POG)$_3$–NH$_2$, Ac–(POG)$_3$–(PO$^\alpha$G)–

$(POG)_3$–NH_2, Ac–[(POG)–($PO^{\beta}G$)]$_3$-(POG)–NH_2 and Ac–($PO^{\beta}G$)$_7$–NH_2 model collagen peptides were considered, where P = Pro, O = Hyp and O^{α}(O^{β}) = α(β)-galactosylated Hyp. The experimental results indicate that a single β-glycosylation per strand ($(POG)_3$–($PO^{\beta}G$)–$(POG)_3$) decreases the melting temperature of the CMP from 39.2±1°C to 37.1±1°C, while the triple helix no longer forms after three or more β-glycosylations per strand ([(POG)–($PO^{\beta}G$)]$_3$–(POG) and ($PO^{\beta}G$)$_7$). Furthermore, a single α-glycosylation per strand ($(POG)_3$–($PO^{\alpha}G$)–$(POG)_3$) prevents helix formation. Unfortunately, the reason for the loss of the triple helical structure upon glycosylation is not clear from this study. However, steric clashes due to the acylation of Hyp residues in a (Pro-Hyp-Gly)$_{10}$ triple helix to give [Pro–Hyp(C(O)CH_3)–Gly]$_{10}$ has been proposed in the literature to be responsible for the observed T_m decrease for the acylated triple helix,[107] and this could provide an explanation for the glycosylated Hyp peptides described above.

1.6. Molecular Modeling

Computational modeling has been successfully applied to various areas of science including weather and biochemical systems. This approach provides information that is complementary to experimental studies and aids in the understanding of many phenomena, such as the consequences of covalent attachment of bulky addition products to DNA[120] or antibody recognition of small molecules.[121] In the case of glycopeptides, modeling is even more relevant because obtaining three dimensional structures by the usual experimental machinery can be more difficult than for unglycosylated peptides (proteins) due to the high mobility of glycans, difficulties associated with crystallization and the structural complexity of (the sugar component) this class of proteins.[5,22,122] As mentioned earlier, no crystal structure (X-ray or NMR) exists for HGRPs. Likewise, there is limited structural data for collagen, especially regarding the covalently-linked carbohydrates, where no crystal structure exists to the best of my knowledge. However, the type and anomericity of the glycan attached to both collagen and HRGPs is known, which allows for the application of molecular modeling to yield specific information about glycopeptide structure and sugar–peptide interactions.

Both HRGPs and collagen have a fairly rigid PPII backbone that is connected to highly flexible sugars, and therefore require molecular dynamics to adequately sample their conformational (potential energy) surfaces. This is in contrast to routine geometry optimizations which guide molecules towards the nearest local minimum. Molecular dynamics (MD) entails the propagation of the degrees of freedom of the molecular system over time, and is rooted in statistical mechanics principles that establish the equivalence of time and ensemble averages. In MD, evolution of the molecular system is based on Newton's equation:

$$F_i = m_i a_i = -\nabla U$$

Where F_i is the force acting on particle m_i and U is the interaction potential. If initial positions and velocities are known, then the force can be calculated and applied to propagate the system. The accuracy of the information obtained from MD primarily depends on the applied potential function. Apart from the MM approximation, the potential, U, can be obtained from first principles based on quantum mechanics (QM). This approach is based on obtaining an approximate solution to the Schrödinger equation (see Chapter 2 for methodological details).

The two approaches (MM and QM) outlined above was used to derive the interaction potential for molecular systems considered in this thesis. The force field method is an approximation of the true potential, but has the advantage of being computationally inexpensive, which allows for the simulation of large systems of biological interest. Additionally, it has been shown in the literature that careful parameterization and application of MM forcefields yields accurate structural and energetic information in comparison to experiments.[122-124] On the other hand, the DFT method is in principle more accurate than MM since the electronic structure is used to derive the interaction potential. However, computational cost increases very steeply with the number of degrees of freedom in the molecular system.

Both approaches have been used in the computational studies of structural properties of Pro–based models in the literature.[8,103,125-130] For example, DFT and other *ab intio* calculations on 4–hydroxyproline and 4–flouroproline provided evidence for experimental observations[131] that 4R–electronegative substituents induce *exo* puckering, while 4S substituents induce *endo* puckering.[103] Additionally, 4R–electronegative elements were shown to stabilize the *trans* conformation, with the stability increasing with electronegativity of the substituent,[103] while 4S–substituents destabilize the *trans* conformation unless a hydrogen bond forms between the 4S–substituent and the C–terminal carbonyl, such as for 4S–hydroxyproline. Song and Kang used B3LYP calculations to examine the transition between the *exo* and *endo* ring conformations of Hyp via an intermediate that adopts an envelope pucker with the backbone N at the top of the envelope.[125] By studying the structure and amide isomerization of N–acetyl–4–hydroxyproline models using DFT, *ab initio* and MM methods, Aliev *et al* noted that a tetrahedral geometry of the nitrogen atom at the transition state is an important structural feature of this *cis*/*trans* transformation.[8,126] Calculations have revealed that a hydrogen bond between the 4S–hydroxyl group and the peptide backbone could explain an increase in the *trans* to *cis* isomer ratio in $CDCl_3$,[127] indicating that water solvent is important for determining the backbone structural preference under biological conditions. Similarly, studies on a [(4S)NHAc–Pro]–OCH_3 model with a 4S N–acetyl group found very high *trans* isomer stability due to strong interactions with the acetamide at C^{α} and the C–terminal carbonyl group of the methyl ester.[128] Finally, Teklebrhan *et al* used B3LYP calculations to determine that at least implicit solvation effects are necessary to correctly calculate the *cis-trans* distribution in glucosyl 3S–hydroxy–5–hydroxymethylproline hybrids since intramolecular hydrogen bonding interactions are modulated by the presence of solvent.[129-130]

In terms of elucidating the structural details of protein–sugar glycosidic linkages, Avenoza and co-workers have used MM and DFT calculations to examine O-glycosylation of model glycopeptides and explain differences in the conformation of an α-linked sugar to Ser and Thr.[132-134] Similarly, Ali *et al* studied asparagine N-glycosylation with N-acetylated glucose, and revealed that the N-acetyl group is important for the rigidity of the glycosidic linkage.[22] However, there is no study in the literature that examines the nature of the hydroxyproline O-glycosidic linkage, which is of importance to understanding the structure of plant HRGPs.

Several computational studies have examined the properties of oligoproline peptides.[135-141] Many of these recent works have concentrated on the energetics for the conversion between the PPI and PPII conformations of oligopeptides. For instance, MM calculations showed that the heterogeneity observed by fluorescence measurements of the end–to–end distance distributions of Pro oligomers is the result of *cis* conformations that occur at several positions in the peptides.[140] The study utilized a F–(Pro)$_x$-Trp peptide (F is a fluorophore that is quenched by a tryptophan (Trp) residue) to measure the end–to–end distance of the oligoprolines. This implies that oligoprolines do not purely adopt the PPII conformation in water. Similar conclusions were drawn using MM-based free energy methods regarding oligoproline peptides.[139] Another study utilized *ab initio* calculations to consider the effects of the end groups on the PPII helix, where it was found that neutral termini (capping with CH_3 ($COCH_3$)and NH_2 ($CONH_2$) groups at the N and C-termini, respectively) favoured the PPII over the PPI conformation, while the reverse is true for the positively charged N–terminus (protonated, $CONH_2^+$) and a negatively charged C–terminus (deprotonated, COO^-).[141]

Other computational studies have been undertaken to aid the understanding of collagen-like peptides,[7,84-85,105,108,110-111,142-148] including a very recent computational study of collagen fibrils.[149] However, these studies broadly concentrated on specific amino acid substitutions (e.g. Gly to Ala) and do not consider the structural consequences of collagen glycosylation. With these previous studies in mind, this thesis aims to provide more information about the structural implications of Hyp glycosylation to fill this void in the literature.

1.7. Thesis overview

The next chapter (Chapter 2) of this thesis will describe the molecular modeling methodologies utilized in this work, as well as provide a summary of how melting temperatures are obtained from CD spectroscopy. Subsequently, in Chapter 3, the nature of the glycosidic linkage is characterized, in order to obtain more information about the effect of Hyp–O–glycosylation on peptide structure. Particularly, MD and DFT are used to determine the conformation of O–galactose linkages in Ac–(4*R*/*S*)–Hyp–OMe compounds. The results are compared to experimental data and used to explain experimental findings that show glycosylation leads to increased stability of the *trans* isomer in the 4*S*–stereoisomer of Hyp compared to an unglycosylated compound, while no measurable increase in *trans* isomer stability is observed in the 4*R*–stereoisomer.

In Chapter 4, the effect of contiguous hydroxylation and glycosylation of a nonaproline peptide are considered as a model for contiguous HRGP glycosylation. Using MD techniques such as replica-exchange MD (REMD), adaptively-biased MD (ABMD) and Hamiltonian-temperature REMD (HT–REMD), the conformations of an unmodified, contiguously-hydroxylated and glycosylated oligopeptide are comprehensively sampled. The results show that both PTMs stabilize the PPII conformation, in accordance with experimental findings.[74] This data also explains the higher melting temperatures measured for the glycosylated compound compared to a hydroxylated nonaproline.

Having determined the structural consequences of contiguous glycosylation in Chapter 4, Chapter 5 concentrates on what happens when the oligoproline peptide is non-contiguously glycosylated, as is the case in many natural HRGPs.[26] It is found that the degree of glycosylation plays an important role in the conformational preferences of the peptide. In particular, intramolecular hydrogen-bonding interactions control *cis* and *trans* conformers of certain peptide residues, and thus the PPII content, which is supported by experimental data.[119]

In Chapter 6, the effects of Hyp–O–glycosylation on the collagen triple helix are examined. Models with one, three and seven glycosylations per strand are included in the

MD calculations. Electronic structure effects were also considered using DFT on a reduced, but informative, model of the (un)glycosylated peptides. Both methods correctly model collagen structure and provide valuable information about this PTM. Additionally, energetics reveal why the triple helix does not form when more than one glycosylation per strand is implemented.

Concluding remarks are made in Chapter 7, which tie together the body of work carried out in this thesis and highlight the contributions to the literature on hydroxyproline glycosylation. Future studies that can be carried out to further explore the ideas reported in this thesis are also provided.

1.8. References[a]

1. Voet, D.; Voet, J., G.; Pratt, C., W., *Fundamentals of Biochemistry*. John Wiley and Sons: New York, 2002.

2. Dwek, R. A., Glycobiology: Toward understanding the function of sugars. *Chemical Reviews* **1996,** *96*, 683–720.

3. Imperiali, B., Protein glycosylation: The clash of the titans. *Acc. Chem. Res.* **1997,** *30*, 452–459.

4. Fadda, E.; Woods, R. J., Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today* **2010,** *15*, 596–609.

5. DeMarco, M. L.; Woods, R. J., Structural glycobiology: A game of snakes and ladders. *Glycobiology* **2008,** *18*, 426–440.

6. Bross-Walch, N.; Kuhn, T.; Moskau, D.; Zerbe, O., Strategies and tools for structure determination of natural products using modern methods of NMR spectroscopy. *Chemistry & Biodiversity* **2005,** *2*, 147–177.

7. Aliev, A. E.; Courtier-Murias, D., Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly. *J. Phys. Chem. B* **2010,** *114*, 12358–12375.

[a] ACS referencing style was implemented throughout this thesis.

8. Aliev, A. E.; Courtier-Murias, D., Conformational analysis of L-prolines in water. *J. Phys. Chem. B* **2007,** *111*, 14034–14042.

9. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta-Gen. Subj.* **1999,** *1473*, 4–8.

10. Yao, J.; Hong, W.; Huang, J.; Zhan, K.; Huang, H.; Hong, M., N-glycosylation Dictates Proper Processing of Organic Anion Transporting Polypeptide 1B1. *Plos One* **2012,** *7*, e52563.

11. Walker, A. K.; Soo, K. Y.; Levina, V.; Talbo, G. H.; Atkin, J. D., N–linked glycosylation modulates dimerization of protein disulfide isomerase family A member 2 (PDIA2). *Febs Journal* **2013,** *280*, 233–243.

12. Grimaldi, S.; Robbins, J.; Edelhoch, H., Interaction of carbohydrate and protein in thyroxine binding globulin. *Biochemistry* **1985,** *24*, 3771–3776.

13. Cheng, S.; Morrone, S.; Robbins, J., Effect of deglycosylation on the binding and immunoreactivity of human thyroxine-binding globulin. *J. Biol. Chem.* **1979,** *254*, 8830–8835.

14. Cheng, S. M.; Edwards, S. A.; Jiang, Y. D.; Grater, F., Glycosylation Enhances Peptide Hydrophobic Collapse by Impairing Solvation. *ChemPhysChem* **2010,** *11*, 2367–2374.

15. Tang, G. Y.; Ruiz, T.; Mintz, K. P., O-Polysaccharide Glycosylation Is Required for Stability and Function of the Collagen Adhesin EmaA of Aggregatibacter actinomycetemcomitans. *Infect. Immun.* **2012,** *80*, 2868-2877.

16. Lee, Y. C., Biochemistry of carbohydrate-protein interaction. *FASEB J.* **1992,** *6*, 3193–3200.

17. Weis, W. I.; Drickamer, K., Structural Basis of Lectin-Carbohydrate Recognition. *Annu. Rev. Biochem.* **1996,** *65*, 441–473.

18. Petrescu, A. J.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R., A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology* **1999,** *9*, 343–352.

19. Hernandez-Gay, J. J.; Arda, A.; Eller, S.; Mezzato, S.; Leeflang, B. R.; Unverzagt, C.; Canada, F. J.; Jimenez-Barbero, J., Insights into the Dynamics and Molecular Recognition Features of Glycopeptides by Protein Receptors: the 3D Solution Structure of Hevein Bound to the Trisaccharide Core of N-Glycoproteins. *Chem.--Eur. J.* **2010,** *16*, 10715-10726, S10715/10711-S10715/10713.

20. Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G., Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **1998,** *33*, 151–208.

21. Cioci, G.; Srivastava, A.; Loganathan, D.; Mason, S. A.; Pérez, S.; Imberty, A., Low-Temperature Neutron Diffraction Structures of N-Glycoprotein Linkage Models and Analogues: Structure Refinement and Trifurcated Hydrogen Bonds. *J. Am. Chem. Soc.* **2011,** *133*, 10042–10045.

22. Ali, M. M. N.; Aich, U.; Varghese, B.; Perez, S.; Imberty, A.; Loganathan, D., Conformational preferences of the aglycon moiety in models and analogs of GlcNAc-Asn linkage: Crystal structures and ab initio quantum chemical calculations of N-(beta-D-glycopyranosyl)haloacetam ides. *J. Am. Chem. Soc.* **2008,** *130*, 8317–8325.

23. Spiro, R. G., Characterization and quantitative determination of hydroxylysince-linked carbohydrate units of several collagens. *J. Biol. Chem.* **1969,** *244*, 602–612.

24. Lamport, D. T. A., Hydroxyproline-O-glycosidic linkage of the plant cell wall glycoprotein extensin. *Nature (London)* **1967,** *216*, 1322–1324.

25. Clarke, A. E.; Anderson, R. L.; Stone, B. A., Form and function of arabinogalactan-proteins. *Phytochemistry* **1979,** *18*, 521-540.

26. Kieliszewski, M. J., The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **2001,** *57*, 319–323.

27. Brandts, J. F.; Halvorson, H. R.; Brennan, M., Consideration of the possibility that the slow step in protein denaturation is due to cis-trans isomerism of proline residues. *Biochemistry* **1975,** *14*, 4953–4963.

28. MacArthur, M. W.; Thornton, J. M., Influence of proline residues on protein conformation. *J. Mol. Biol.* **1991,** *218*, 397–412.

29. Stein, R. L., Mechanism of enzymatic and nonenzymatic prolyl cis-trans isomerization. *Adv. Protein. Chem.* **1993,** *44*, 1–24.

30. Kang, Y. K.; Jhon, J. S.; Park, H. S., Conformational preferences of proline oligopeptides. *J. Phys. Chem. B* **2006,** *110*, 17645–17655.

31. Shoulders, M. D.; Kotch, F. W.; Choudhary, A.; Guzei, I. A.; Raines, R. T., The Aberrance of the 4S Diastereomer of 4-Hydroxyproline. *J. Am. Chem. Soc.* **2010,** *132*, 10857–10865.

32. Fischer, G.; Schmid, F. X., The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell. *Biochemistry* **1990,** *29*, 2205–2212.

33. Halab, L.; Gosselin, F.; Lubell, W. D., Design, synthesis, and conformational analysis of azacycloalkane amino acids as conformationally constrained probes for mimicry of peptide secondary structures. *Biopolymers* **2000,** *55*, 101–122.

34. Halab, L.; Lubell, W. D., Effect of Sequence on Peptide Geometry in 5-tert-Butylprolyl Type VI β-Turn Mimics. *J. Am. Chem. Soc.* **2002,** *124*, 2474–2484.

35. Berg, R. A.; Prockop, D. J., Thermal transition of a nonhydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen. *Biochem. Biophys. Res. Commun.* **1973,** *52*, 115-120.

36. Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B., Amino acid propensities for the collagen triple-helix. *Biochemistry* **2000,** *39*, 14960–14967.

37. Vanholst, G. J.; Varner, J. E., Reinforced Polyproline-II Conformation in a Hydroxyproline-Rich Cell-Wall Glycoprotein from Carrot Root. *Plant Physiol.* **1984,** *74*, 247–251.

38. Shpak, E.; Barbar, E.; Leykam, J. F.; Kieliszewski, M. J., Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in Nicotiana tabacum. *J. Biol. Chem.* **2001,** *276*, 11272–11278.

39. Doig, A. J.; MacArthur, M. W.; Stapley, B. J.; Thornton, J. M., Structures of N-termini of helices in proteins. *Protein Sci.* **1997,** *6*, 147–155.

40. Stapley, B. J.; Creamer, T. P., A survey of left-handed polyproline II helices. *Protein Sci.* **1999,** *8*, 587–595.

41. Kakinoki, S.; Hirano, Y.; Oka, M., On the stability of polyproline-I and II structures of proline oligopeptides. *Polymer Bulletin (Heidelberg, Ger.)* **2005,** *53*, 109–115.

42. Zimmerman, S. S.; Scheraga, H. A., Stability of Cis, Trans, and Nonplanar Peptide Groups. *Macromolecules* **1976,** *9*, 408–416.

43. Zimmerman, S. S.; Scheraga, H. A., Influence of local interactions on protein structure. I. Conformational energy studies of N-acetyl-N'-methylamides of Pro-X and X-Pro dipeptides. *Biopolymers* **1977,** *16*, 811–843.

44. MacArthur, M. W.; Thornton, J. M., Influence of Proline Residues on Protein Conformation. *J. Mol. Biol.* **1991,** *218*, 397–412.

45. Stein, R. L., *Advances in Protein Chemistry* **1993,** *44*, 1–24.

46. Zimmerman, S. S.; Scheraga, H. A., Influence of local interactions on protein structure. I. Conformational energy studies of N-acetyl-N'-methylamides of Pro-X and X-Pro dipeptides. *Biopolymers* **1977,** *16*, 811–843.

47. Hodges, J. A.; Raines, R. T., Energetics of an n → π* Interaction that Impacts Protein Structure. *Organic Letters* **2006,** *8*, 4695–4697.

48. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T., Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **2001,** *123*, 777–778.

49. Eberhardt, E. S.; Loh, S. N.; Hinck, A. P.; Raines, R. T., Solvent effects on the energetics of prolyl peptide bond isomerization. *J. Am. Chem. Soc.* **1992,** *114*, 5437–5439.

50. DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L., Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations. *J. Am. Chem. Soc.* **2002,** *124*, 2497–2505.

51. Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L., Fluoroprolines as tools for protein design and engineering. *Angew. Chem.-Int. Edit.* **2001,** *40*, 923–925.

52. Cadamuro, S. A.; Reichold, R.; Kusebauch, U.; Musiol, H.-J.; Renner, C.; Tavan, P.; Moroder, L., Conformational properties of 4-mercaptoproline and related derivatives. *Angew. Chem.-Int. Edit.* **2008,** *47*, 2143–2146.

53. Kuemin, M.; Sonntag, L.-S.; Wennemers, H., Azidoproline containing helices: Stabilization of the polyproline II structure by a functionalizable group. *J. Am. Chem. Soc.* **2007,** *129*, 466–467.

54. Shoulders, M. D.; Hodges, J. A.; Raines, R. T., Reciprocity of Steric and Stereoelectronic Effects in the Collagen Triple Helix. *J. Am. Chem. Soc.* **2006,** *128*, 8112–8113.

55. Prockop, D. J.; Kivirikko, K. I., Collagens - Molecular-biology, Disseases, and Potentials for Therapy. *Annu. Rev. Biochem.* **1995,** *64*, 403–434.

56. Tryggvason, K.; Risteli, J.; Kivirikko, K. I., Seperation of Proplyl 3-Hydroxylase and 4-Hydroxylase activities and 4-Hydroxyproline Requirement for Synthesis of 3-Hydroxyproline. *Biochem. Biophy. Res. Commun.* **1977,** *76*, 275–281.

57. Gorres, K. L.; Raines, R. T., Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **2010,** *45*, 106–124.

58. Khashimova, Z. S., Hydroxyproline-containing plant proteins. *Chem. Nat. Compd.* **2003,** *39*, 229–236.

59. Ferris, P. J.; Woessner, J. P.; Waffenschmidt, S.; Kilz, S.; Drees, J.; Goodenough, U. W., Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. *Biochemistry* **2001,** *40*, 2978–2987.

60. Velasquez, S. M.; Ricardi, M. M.; Dorosz, J. G.; Fernandez, P. V.; Nadra, A. D.; Pol-Fachin, L.; Egelund, J.; Gille, S.; Harholt, J.; Ciancia, M.; Verli, H.; Pauly, M.; Bacic, A.; Olsen, C. E.; Ulvskov, P.; Petersen, B. L.; Somerville, C.; Iusem, N. D.; Estevez, J. M., O-Glycosylated Cell Wall Proteins Are Essential in Root Hair Growth. *Science* **2011,** *332*, 1401–1403.

61. Ellis, M.; Egelund, J.; Schultz, C. J.; Bacic, A., Arabinogalactan-Proteins: Key Regulators at the Cell Surface? *Plant Physiol.* **2010,** *153*, 403–419.

62. Showalter, A. M., Structure and function of plant cell wall proteins. *Plant Cell* **1993,** *5*, 9–23.

63. Nothnagel, E. A., Proteoglycans and related components in plant cells. In *International Review of Cytology - a Survey of Cell Biology, Vol 174*, Academic Press Inc: San Diego, 1997; Vol. 174, pp 195–291.

64. Cannon, M. C.; Terneus, K.; Hall, Q.; Tan, L.; Wang, Y. M.; Wegenhart, B. L.; Chen, L. W.; Lamport, D. T. A.; Chen, Y. N.; Kieliszewski, M. J., Self-assembly of the plant cell wall requires an extensin scaffold. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 2226–2231.

65. Tan, L.; Varnai, P.; Lamport, D. T. A.; Yuan, C. H.; Xu, J. F.; Qiu, F.; Kieliszewski, M. J., Plant O-Hydroxyproline Arabinogalactans Are Composed of Repeating Trigalactosyl Subunits with Short Bifurcated Side Chains. *J. Biol. Chem.* **2010,** *285*, 24575–24583.

66. Tan, L.; Qiu, F.; Lamport, D. T. A.; Kieliszewski, M. J., Structure of a hydroxyproline (Hyp)-arabinogalactan polysaccharide from repetitive Ala-Hyp expressed in transgenic Nicotiana tabacum. *J. Biol. Chem.* **2004,** *279*, 13156–13165.

67. Ellis, M.; Egelund, J.; Schultz, C. J.; Bacic, A., Arabinogalactan-Proteins: Key Regulators at the Cell Surface? *Plant Physiology* **2010,** *153*, 403–419.

68. Showalter, A. M., Arabinogalactan-proteins: structure, expression and function. *Cellular and Molecular Life Sciences* **2001,** *38*, 1399–1417.

69. Pope, D. G., Relationship betweeen hydroxyproline containing proteins secreted into cell wall and medium by suspension cultured acer-psudoplantanus cells. *Plant Physiol.* **1977,** *59*, 894-900.

70. Mau, S. L.; Chen, C. G.; Pu, Z. Y.; Moritz, R. L.; Simpson, R. J.; Bacic, A., *Plant J.* **1995,** *8*, 269–281.

71. Du, A.; Simpson, R. J.; Clarke, A. E.; Bacic, A., Molecular characterization of a stigma-specific gene encoding an arabinogalactan-protein (AGP) from Nicotiana alata. *Plant J.* **1996,** *9*, 313–323.

72. Owens, N. W.; Braun, C.; O'Neil, J. D.; Marat, K.; Schweizer, F., Effects of glycosylation of (2S,4R)-4-hydroxyproline on the conformation, kinetics, and thermodynamics of prolyl amide isomerization. *J. Am. Chem. Soc.* **2007,** *129*, 11670–11671.

73. Owens, N. W.; Lee, A.; Marat, K.; Schweizer, F., The Implications of (2S,4S)-Hydroxyproline 4-O-Glycosylation for Prolyl Amide Isomerization. *Chem.-Eur. J.* **2009,** *15*, 10649–10657.

74. Owens, N. W.; Stetefeld, J.; Lattova, E.; Schweizer, F., Contiguous O-Galactosylation of 4(R)-Hydroxy-L-proline Residues Forms Very Stable Polyproline II Helices. *J. Am. Chem. Soc.* **2010,** *132*, 5036–5042.

75. Fincher, G. B.; Stone, B. A.; Clarke, A. E., Arabinogalactan proteins - structure, biosynthesis, and function. *Annu. Rev. Plant Physiol. Plant Molec. Biol.* **1983,** *34*, 47–70.

76. Qi, W.; Fong, C.; Lamport, D. T. A., Gum arabic glycoprotein is a twisted hairy rope - a new model based on O-galactosylhydroxyproline as a polysaccharide attachment site. *Plant Physiol.* **1991,** *96*, 848–855.

77. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M., Crystal structure and molecular structure of a collagen-like peptide at 1.9 Angstrom resolution. *Science* **1994,** *266*, 75–81.

78. Ramachandran, G. N.; Kartha, G., Structure of Collagen. *Nature* **1954,** *174*, 269-270.

79. Ramachandran, G. N.; Kartha, G., Structure of Collagen. *Nature* **1955,** *176*, 593–595.

80. Rich, A.; Crick, F. H. C., Structure of Collagen. *Nature* **1955,** *176*, 915–916.

81. Bachinger, H. P.; Davis, J. M., Sequence Specific Thermal-stability of the Collagen Triple Helix. *Int. J. Biol. Macromol.* **1991,** *13*, 152–156.

82. Bachinger, H. P.; Morris, N. P.; Davis, J. M., Thermal-stability and Folding of the Collagen Triple Helix and the Effects of Mutations in Osteogenesis Imperfecta on the Triple Helix of Collagen. *Am. J. Med. Genet.* **1993,** *45*, 152–162.

83. Doig, A. J., Statistical Thermodynamics of the Collagen Triple-Helix/Coil Transition. Free Energies for Amino Acid Substitutions within the Triple-Helix. *J. Phys. Chem. B* **2008,** *112*, 15029–15033.

84. Lee, K. H.; Kuczera, K.; Holl, M. M. B., Effect of osteogenesis imperfecta mutations on free energy of collagen model peptides: A molecular dynamics simulation. *Biophys. Chem.* **2011,** *156*, 146-152.

85. Lee, K. H.; Holl, M. M. B., Free Energy Simulation to Investigate the Effect of Amino Acid Sequence Environment on the Severity of Osteogenesis Imperfecta by Glycine Mutations in Collagen. *Biopolymers* **2011,** *95*, 401–409.

86. Marlowe, A. E.; Singh, A.; Yingling, Y. G., The effect of point mutations on structure and mechanical properties of collagen-like fibril: A molecular dynamics study. *Mater. Sci. Eng. C-Mater. Biol. Appl.* **2012,** *32*, 2583–2588.

87. Gautieri, A.; Vesentini, S.; Redaelli, A.; Buehler, M. J., Osteogenesis imperfecta mutations lead to local tropocollagen unfolding and disruption of H-bond network. *RSC Adv.* **2012,** *2*, 3890-3896.

88. Rauch, F.; Glorieux, F. H., Osteogenesis imperfecta. *Lancet* **2004,** *363*, 1377–1385.

89. Shoulders, M. D.; Raines, R. T., Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009,** *78*, 929–958.

90. Ramshaw, J. A. M.; Shah, N. K.; Brodsky, B., Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J. Struct. Biol.* **1998,** *122*, 86–91.

91. Berisio, R.; De Simone, A.; Ruggiero, A.; Improta, R.; Vitagliano, L., ., Role of side chains in collagen triple helix stabilization and partner recognition. *J. Pep. Sci.* **2009,** *15*, 131–140.

92. Vitagliano, L.; Berisio, R.; Mazzarella, L.; Zagari, A., Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **2001,** *58*, 459–465.

93. Fields, G. B., Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.* **2010,** *8*, 1237–1258.

94. Bann, J. G.; Peyton, D. H.; Bachinger, H. P., Sweet is stable: glycosylation stabilizes collagen. *FEBS Lett.* **2000,** *473*, 237–240.

95. Benzi, C.; Improta, R.; Scalmani, G.; Barone, V., Quantum mechanical study of the conformational behavior of proline and 4R-hydroxyproline dipeptide analogues in vacuum and in aqueous solution. *J. Comput. Chem.* **2002,** *23*, 341–350.

96. Berisio, R.; De Simone, A.; Ruggiero, A.; Improta, R.; Vitagliano, L., Role of side chains in collagen triple helix stabilization and partner recognition. *J. Peptide Sc.* **2009,** *15*, 131–140.

97. Brodsky, B.; Thiagarajan, G.; Madhan, B.; Kar, K., Triple-helical peptides: An approach to collagen conformation, stability, and self-association. *Biopolymers* **2008,** *89*, 345–353.

98. Chelberg, M. K.; McCarthy, J. B.; Skubitz, A. P. N.; Furcht, L. T.; Tsilibary, E. C., Characterization of a Synthetic Peptide from Type-IV Collagen that Promotes Melanoma Cell-Adhesion, Spreading, and Motility. *J. Cell Biol.* **1990,** *111*, 261–270.

99. Erdmann, R. S.; Wennemers, H., Functionalizable Collagen Model Peptides. *J. Am. Chem. Soc.* **2010,** *132*, 13957–13959.

100. Erdmann, R. S.; Wennemers, H., Importance of Ring Puckering versus Interstrand Hydrogen Bonds for the Conformational Stability of Collagen. *Angew. Chem.-Int. Edit.* **2011,** *50*, 6835–6838.

101. Fraser, R. D. B.; Macrae, T. P.; Suzuki, E., Chain Conformation in the Collagen Molecule. *J. Mol. Biol.* **1979,** *129*, 463–481.

102. Horng, J.-C.; Raines, R. T., Stereoelectronic effects on polyproline conformation. *Protein Sci.* **2006,** *15*, 74–83.

103. Improta, R.; Benzi, C.; Barone, V., Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* **2001,** *123*, 12568–12577.

104. Improta, R.; Berisio, R.; Vitagliano, L., Contribution of dipole-dipole interactions to the stability of the collagen triple helix. *Protein Sci.* **2008,** *17*, 955–961.

105. Improta, R.; Mele, F.; Crescenzi, O.; Benzi, C.; Barone, V., Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 2. A Quantum Mechanical/Molecular Mechanical Study of (Proline-Proline-Glycine)n Polypeptides. *J. Am. Chem. Soc.* **2002,** *124*, 7857–7865.

106. Jakobsche, C. E.; Choudhary, A.; Miller, S. J.; Raines, R. T., n -> pi* Interaction and n)(pi Pauli Repulsion Are Antagonistic for Protein Stability. *J. Am. Chem. Soc.* **2010,** *132*, 6651–6653.

107. Jenkins, C. L.; McCloskey, A. I.; Guzei, I. A.; Eberhardt, E. S.; Raines, R. T., O-acylation of hydroxyproline residues: effect on peptide-bond isomerization and collagen stability. *Biopolymers* **2005,** *80*, 1–8.

108. Klein, T. E.; Huang, C. C., Computational investigations of structural changes resulting from point mutations in a collagen-like peptide. *Biopolymers* **1999,** *49*, 167–183.

109. Pálfi, V. K.; Perczel, A., How stable is a collagen triple helix? An ab initio study on various collagen and β-sheet forming sequences. *J. Comput. Chem.* **2008,** *29*, 1374–1386.

110. Streeter, I.; de Leeuw, N. H., Atomistic Modeling of Collagen Proteins in Their Fibrillar Environment. *J. Phys. Chem. B* **2010,** *114*, 13263–13270.

111. Suarez, E.; Diaz, N.; Suarez, D., Entropic Control of the Relative Stability of Triple-helical Collagen Peptide Models. *J. Phys. Chem. B* **2008,** *112*, 15248–15255.

112. Barone, V.; Orlandini, L., Methyl addition to acetylene and ethylene from a density functional approach. *Chem. Phys. Lett.* **1995,** *246*, 45–52.

113. Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J., Synthesis and physical properties of (hydroxyproline-proline-glycine)10 - hydroxyproline in the X-positon decreases the melting temperature of the collagen triple helix. *Arch. Biochem. Biophys.* **1982,** *219*, 198–203.

114. Bann, J. G.; Bachinger, H. P., Glycosylation/Hydroxylation-induced Stabilization of the Collagen Triple Helix: 4-trans-Hydroxyproline in the Xaa Position can Stabilize the Triple Helix. *J. Biol. Chem.* **2000,** *275*, 24466-24469.

115. Spiro, R. G., Characterization and Quantitative Determination of the Hydroxylysine-linked Carbohydrate Units of Several Collagens. *J. Biol. Chem.* **1969,** *244*, 602.

116. Myllyharju, J.; Kivirikko, K. I., Collagens and collagen-related diseases. *Ann. Med.* **2001,** *33*, 7–21.

117. Fields, G. B., Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.* **2010,** *8*, 1237–1258.

118. Fields, C. G.; Mickelson, D. J.; Drake, S. L.; McCarthy, J. B.; Fields, G. B., Melanoma cell adhesion and spreading activities of a synthetic 124-residue triple helical mini-collagen. *J. Biol. Chem.* **1993,** *268*, 14153–14160.

119. S. Bommagani, E.B. Naziga, N. W. Owens, E. Lattová, J. D. O'Neil, S. D. Wetmore and F. Schweizer, The Effects of (2*S*,4*R*)-4-Hydroxyproline Glycosylation on the Stability of the Collagen Triple Helix, submitted for publication **2013**.

120. Manderville, R. A.; Omumi, A.; Rankin, K. M.; Wilson, K. A.; Millen, A. L.; Wetmore, S. D., Fluorescent C-Linked C-8-Aryl-guanine Probe for Distinguishing syn from anti Structures in Duplex DNA. *Chem. Res. Tox.* **2012,** *25*, 1271–1280.

121. Yuan, M.; Na, Y.; Li, L. L.; Liu, B.; Sheng, W.; Lu, X. N.; Kennedy, I.; Crossan, A.; Wang, S., Computer-Aided Molecular Modeling Study on Antibody Recognition of Small Molecules: An Immunoassay for Triazine Herbicides. *J. Agric. Food Chem.* **2012,** *60*, 10486–10493.

122. Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B., Effects of Glycosylation on Peptide Conformation: A Synergistic Experimental and Computational Study. *J. Am. Chem. Soc.* **2004,** *126*, 8421–8425.

123. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29*, 622–655.

124. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 10541–10545.

125. Song, I. K.; Kang, Y. K., Conformational Preference and Cis-Trans Isomerization of 4(R)-Substituted Proline Residues. *J. Phys. Chem. B* **2006,** *110*, 1915–1927.

126. Aliev, A. E.; Bhandal, S.; Courtier-Murias, D., Quantum Mechanical and NMR Studies of Ring Puckering and cis/trans-Rotameric Interconversion in Prolines and Hydroxyprolines. *J. Phys. Chem. A* **2009,** *113*, 10858–10865.

127. Shoulders, M. D.; Kotch, F. W.; Choudhary, A.; Guzei, I. A.; Raines, R. T., The Aberrance of the 4S Diastereomer of 4-Hydroxyproline. *J. Am. Chem. Soc.* **2010,** *132*, 10857–10865.

128. Kuemin, M.; Nagel, Y. A.; Schweizer, S.; Monnard, F. W.; Ochsenfeld, C.; Wennemers, H., Tuning the cis/trans Conformer Ratio of Xaa-Pro Amide Bonds by Intramolecular Hydrogen Bonds: The Effect on PPII Helix Stability. *Angew. Chem.-Int. Edit.* **2010,** *49*, 6324–6327.

129. Zhang, K.; Teklebrhan, R. B.; Schreckenbach, G.; Wetmore, S.; Schweizer, F., Intramolecular Hydrogen Bond-Controlled Prolyl Amide Isomerization in Glucosyl 3'(S)-Hydroxy-5'-hydroxymethylproline Hybrids: Influence of a C-5'-Hydroxymethyl Substituent on the Thermodynamics and Kinetics of Prolyl Amide Cis/Trans Isomerization. *J. Org. Chem.* **2009,** *74*, 3735–3743.

130. Teklebrhan, R. B.; Zhang, K. D.; Schreckenbach, G.; Schweizer, F.; Wetmore, S. D., Intramolecular Hydrogen Bond-Controlled Prolyl Amide Isomerization in Glucosyl 3(S)-Hydroxy-5-hydroxymethylproline Hybrids: A Computational Study. *J. Phys. Chem. B* **2010,** *114*, 11594-11602.

131. Eberhardt, E. S.; Panisik, N., Jr.; Raines, R. T., Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. *J. Am. Chem. Soc.* **1996,** *118*, 12261–12266.

132. Corzana, F.; Busto, J. H.; Engelsen, S. B.; Jimenez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Avenoza, A., Effect of beta-O-glucosylation on L-Ser and L-Thr diamides: A bias toward alpha-helical conformations. *Chem.-Eur. J.* **2006,** *12*, 7864–7871.

133. Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; Asensio, J. L.; Jimenez-Barbero, J.; Peregrina, J. M.; Avenoza, A., New insights into alpha-GalNAc-Ser motif: Influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **2006,** *128*, 14640–14648.

134. Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; García de Luis, M.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A., Serine versus Threonine Glycosylation: The Methyl Group Causes a Drastic Alteration on the Carbohydrate Orientation and on the Surrounding Water Shell. *J. Am. Chem. Soc.* **2007,** *129*, 9458–9467.

135. Tanaka, S.; Scheraga, H. A., Calculation of Conformational Properties of Oligomers of L-Proline. *Macromolecules* **1974,** *7*, 698–705.

136. Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A., Polyproline II helix conformation in a proline-rich environment: A theoretical study. *Biophys. J.* **2004,** *86*, 731–742.

137. Moradi, M.; Babin, V.; Roland, C.; Sagui, C., A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *J. Chem. Phys.* **2010,** *133*, 125104–125118.

138. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., A Statistical Analysis of the PPII Propensity of Amino Acid Guests in Proline-Rich Peptides. *Biophys. J.* **2011,** *100*, 1083–1093.

139. Moradi, M.; Lee, J.-G.; Babin, V.; Roland, C.; Sagui, C., Free Energy and Structure of Polyproline Peptides: An Ab Initio and Classical Molecular Dynamics Investigation. *Int. J. Quantum Chem.* **2010,** *110*, 2865–2879.

140. Doose, S.; Neuweiler, H.; Barsch, H.; Sauer, M., Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 17400–17405.

141. Kuemin, M.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H., Effects of Terminal Functional Groups on the Stability of the Polyproline II Structure: A Combined Experimental and Theoretical Study. *J. Am. Chem. Soc.* **2009,** *131*, 15474–15482.

142. Mooney, S. D.; Kollman, P. A.; Klein, T. E., Conformational preferences of substituted prolines in the collagen triple helix. *Biopolymers* **2002,** *64*, 63–71.

143. Radmer, R. J.; Klein, T. E., Triple helical structure and stabilization of collagen-like molecules with 4(R)-hydroxyproline in the Xaa position. *Biophys. J.* **2006,** *90*, 578–588.

144. Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S., A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **2005,** *26*, 1612–1616.

145. Bodian, D. L.; Madhan, B.; Brodsky, B.; Klein, T. E., Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry* **2008,** *47*, 5424–5432.

146. Punitha, V.; Raman, S. S.; Parthasarathi, R.; Subramanian, V.; Rao, J. R.; Nair, B. U.; Ramasami, T., Molecular Dynamics Investigations on the Effect of D Amino Acid Substitution in a Triple-Helix Structure and the Stability of Collagen. *J. Phys. Chem. B* **2009,** *113*, 8983–8992.

147. Sundar, R. S.; Gopalakrishnan, R.; Wade, R. C.; Subramanian, V., Structural Basis for the Varying Propensities of Different Amino Acids To Adopt the Collagen Conformation. *J. Phys. Chem. B* **2011,** *115*, 2593–2607.

148. Xu, Y.; Dannenberg, J. J., Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProDAla, and ProProDSer Triads. *J. Am. Chem. Soc.* **2005,** *127*, 14130–14131.

149. Streeter, I.; de Leeuw, N. H., Atomistic Modeling of Collagen Proteins in Their Fibrillar Environment. *J. Phys. Chem. B* **2010,** *114*, 13263–13270.

Chapter 2. Methods

2.1. Introduction

This chapter describes the molecular modeling methods used in this thesis. In particular, the molecular mechanics (MM) methodology is presented, as well as its application in molecular dynamics (MD). Thereafter, several advanced sampling techniques that help overcome kinetic traps on the molecular potential energy surface (PES) will be described. The quantum mechanical (QM) method applied in this thesis (density functional theory (DFT)) is also outlined. Furthermore, the strengths and weaknesses of the computational methodologies and some technical considerations for MD simulations are discussed. A brief description of experimental circular dichroism (CD) spectroscopy used for the derivation melting temperatures (T_m) is also given which correlates with PPII stability that is important for the oligopeptides considered in this thesis.

2.2. Molecular Mechanics

As mentioned in Chapter 1, the interactions in a molecular system are described by a potential (U) that can be approximated in different ways for practical purposes. In empirical molecular mechanics, the atoms forming the molecular system are typically represented by spheres with partial charges. The spheres are connected via bonds, angles and dihedrals (Figure 2.1), and interact through Coulomb and Lenard-Jones (LJ) forces. Within this representation, U is usually written as a sum of several energy components:

$$U = \sum_{\text{bonds}} U_{\text{bonds}} + \sum_{\text{angles}} U_{\text{angles}} + \sum_{\text{dihedrals}} U_{\text{dihedrals}} + \sum_{\text{nonbonded}} U_{\text{nonbonded}} \qquad 2.1$$

Each component in Equation 2.1 has the following form in the MM calculations using the AMBER/GLYCAM force field reported in this thesis:

$$U_{bonds} = K_b(r-r_0)^2 \tag{2.2}$$

$$U_{angles} = K_\theta(\theta-\theta_0)^2 \tag{2.3}$$

$$U_{dihedral} = \frac{1}{2}K_\varphi(1+\cos(n\varphi+\gamma)) \tag{2.4}$$

where K_b, K_θ and K_φ represent the bond, angle and dihedral force constants (parameters), respectively. The instantaneous bond length and angle are r and θ, respectively, while r_0 and θ_0 are the equilibrium bond and angle parameters. The multiplicity and phase of the dihedral angle (φ) parameters are n and γ, respectively. Non-bonded interactions are included in the potential via Coulomb interactions between point charges and LJ interactions between sites:

$$U_{nonbonded} = \sum_{i,j}\left[\frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}}\right] \tag{2.5}$$

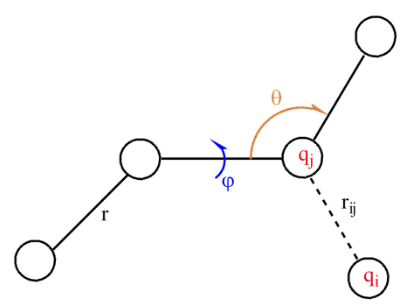

Figure 2.1. Schematic of the ball and stick representation of a molecular system that is used in molecular mechanics based calculations.

where A_{ij} (B_{ij}) are the LJ repulsion-dispersion parameters for each atom pair, q_i and q_j are partial charges that are parameters derived using QM calculations, r_{ij} is the interatomic separation and ε represents the permittivity of free space. To have a complete MM force field for a particular molecular system, all the equilibrium parameters (r_0,θ_0), multiplicity and phase (n and γ), and force constants (K_b, K_θ and K_φ) terms as must be provided. The partial charge on each atom (q_i) as well as the LJ parameters (A_{ij} ,B_{ij}) for all atom pairs are also required.

The bond, angle and dihedral parameters for both the AMBER and GLYCAM force fields are obtained from either experimental data or fitting to quantum mechanical data for small molecules containing representative bonds, angles and dihedrals. In the case of fitting to experimental data, bond parameters for instance are utilized such that the normal mode frequencies obtained using a particular force constant and equilibrium value correspond to experimental vibrational data obtained from infrared (IR) spectroscopy. Alternatively, a quadratic term for the bond length (Equation 2.2) is fitted to energies obtained from a QM scan of the representative bond of a small molecule containing atoms in a similar chemical environment. Such atoms are used to define a particular 'atom type', which is used to build larger molecules. A similar fitting procedure is used to determine angle and dihedral angle parameters.

The atomic partial charges applied in the AMBER and GLYCAM force fields used in this thesis are derived using molecular electrostatic potentials (MEPs) obtained from QM calculations (Hatree-Fock). The particular procedure used in the AMBER/GLYCAM FFs is called the restrained electrostatic potential (RESP) fitting.[1] In this procedure, partial charges that best reproduce the calculated QM electrostatic potential at certain points outside the molecule are assigned to each atom in the molecular system. Using these methods, force field parameters for all proteinogenic amino acids, nucleic acid residues, many sugar monomers, as well as other biological fragments, have been developed for AMBER/GLYCAM. These residues can then be combined to form nucleic acids, oligopeptides or proteins.

In the literature, several research groups have developed parameter sets for different classes of molecules that have been widely applied in biomolecular simulations similar to those studied in the present thesis. The AMBER[2-3] and GLYCAM[4] force fields are used in this thesis and have been extensively applied in the literature to model a wide range of biological molecules, including oligoproline and glycopeptide systems.[4-8] The AMBER (Assisted Model Building with Energy Refinement) force field was primarily developed for protein and nucleic acid systems, but also contains parameters for the simulation of lipids. AMBER uses the exact functional form described in Equations 2.1 to 2.4. The GLYCAM force field also uses Equations 2.1 to 2.4 and was initially designed to complement the AMBER force field by providing parameters for sugars. It has since evolved to be a more complete and stand alone force field.[4] Both force fields were therefore used in this thesis to provide a comprehensive treatment for the glycopeptides studied herein.

Several versions of the AMBER and GLYCAM parameter sets have been developed over the years. The AMBER FF99SB[3] forcefield that is used in this thesis has been shown in the literature to correctly describe the structure and dynamics of oligoproline systems.[3,6] The latest version of GLYCAM force field[4] was used in this thesis, which has been shown to accurately model the structure and dynamics of saccharides and glycoproteins.[7-8]

2.3. Molecular Dynamics (MD) Simulations

2.3.1 Basics of the MD Simulation Method

Given a MM force field, the next step is to generate configurations of the molecular system such that the potential energy surface (PES) is adequately sampled. This is challenging since for a molecular system with 3N degrees of freedom (N ~ thousands for biomolecular systems), the PES is a high dimensional surface. Sampling of the PES can be accomplished in several ways including molecular dynamics (MD) simulations. In MD, the molecular system is evolved based on Newton's equation:

$$F_i = m_i a_i = -\nabla U \tag{2.6}$$

where F_i is the force acting on particle m_i and U is the interaction potential described in Equation 2.1. If the initial positions and velocities of the particles making up the molecular system are known, then the force can be calculated and used to propagate the particles in the system. In practice, starting positions can come from an experimentally determined crystal structure or built using a software program (e.g., the xLeap module of the AMBER suite of programs). In this thesis, both approaches are used depending on the availability of experimental structures. Specifically, the collagen model peptides were derived from a crystal structure (Chapter 6), while starting structures for all other peptides (Chapters 3 to 5) were generated using the xLeap module of the AMBER software suite due to the absence of experimental data. Initial velocities came from a Boltzmann distribution at the desired starting temperature. Most MD simulation programs, including the AMBER suite[9] of programs, update the positions of the particles in the molecular system using a version of the velocity Verlet algorithm[10]

$$r_i(t + \delta t/2) \approx r_i(t) + v_i(t - \delta t/2)\delta t \tag{2.7}$$

$$v_i(t + \delta t/2) \approx v_i(t - \delta t/2) + \frac{f(t)}{m}\delta t \tag{2.8}$$

where δt is the simulation time step, $r_i(t)$ is the position of particle i at time t, and all other parameters are as described in Equation 2.6.

2.3.2. Replica Exchange Molecular Dynamics (REMD)

Propagating the degrees of freedom in a molecular system using Equation 2.7 can adequately sample the PES of some molecules. However, for complex molecules, such as the glycopeptides considered in this thesis, the system can get stuck in kinetic traps that require prohibitively long simulation times to overcome. Thus, it is necessary to utilize techniques that ensure such barriers on the PES are overcome. One such method is replica exchange MD (REMD).[11] In the temperature REMD methodology implemented in this thesis, several copies of the molecular system are simulated at different temperatures. At certain intervals during the simulation, an attempt is made to exchange the coordinates (or temperature) of two replicas. A popular acceptance criterion is the Metropolis scheme, where the acceptance probability of a move between two replicas in the canonical ensemble is given as:

$$P(i \rightarrow j) = e^{(\beta_i - \beta_j)(E_i - E_j)} \qquad 2.9$$

In Equation 2.9, $P(i \rightarrow j)$ is the probability of accepting an exchange between replicas i and j, $\beta = k_bT$ (k_b = Boltzmann constant) and E_i is the energy of replica i. The Metropolis criteria accepts the change if $(\beta_i - \beta_i)(E_i - E_j) < 0$, but uses the probability defined in Equation 2.9 if $(\beta_i - \beta_i)(E_i - E_j) > 0$. To ensure reasonable acceptance rates ($\sim$ 15% or higher), the potential energy distribution of the replicas must contain some overlap.[12] One can ensure this happens by choosing the right number of replicas, where it has been determined that the number of replicas required scales as the square root of the number of degrees of freedom in the molecular system.[13]

The REMD method described above is very general in the sense that specific information about the degrees of freedom that need to be dynamic to overcome kinetic traps is not required. REMD significantly increases the sampling capability of MD simulations and has therefore been applied to a wide variety of biomolecular systems.[14-16] However, in many instances, considerably extensive simulations are still required to adequately sample the PES of complex biomolecules. This is particularly relevant in the case of explicit solvent calculations, where data storage becomes an issue. Specifically, since several replicas of the molecular system are simultaneously simulated, coordinates for each replica have to be periodically stored, which drastically increases storage space requirements. For instance, the REMD calculations reported in Chapter 4 (Subsection 4.3.3) require over 2 Terabytes of storage space, which is about 4 times of the average storage capacity (~500 GB) commonly available on current personal computers. For these reasons, it is necessary to utilize methods that specifically accelerate the dynamics of particular degrees of freedom. In such a case, a collective variable (CV) must be defined that is connected to the interesting conformational change. In the peptides considered in this thesis, the CV is the backbone ω dihedral, which controls *cis* to *trans* isomerization and cannot be sampled over the course of a regular MD simulation. This is the case because the isomerization process requires over 10 kcal mol^{-1} of energy to cross the isomerization barrier.[17-19] Hence, free energy methods that help overcome such barriers were implemented in this thesis.

2.3.3. Adaptively Biased Molecular Dynamics

To overcome the barriers associated with peptide *cis* to *trans* isomerization, adaptively biased molecular dynamics (ABMD)[20-21] simulations were implemented. ABMD is effectively an advanced sampling method that makes use of a time–dependent potential to bias the dynamics of the molecular system. This flattens the PES with respect to the selected CV related to the property under study. The ABMD implementation applied in this thesis is based on the following equations:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i + \frac{\partial}{\partial r_i} U[t|\sigma(r_1, r_2, \ldots\ldots, r_N)] \quad 2.10$$

$$\frac{\partial U(t|\xi)}{\partial t} = \frac{k_B T}{\tau_F} G(\xi|\sigma(r_1, r_2, \ldots\ldots, r_N)) \quad 2.11$$

The first equation (Newton's equation of motion) defines the force acting on particle m_i at position r_i during time t together with a biasing potential U, which evolves according to the second equation. The biasing potential is defined in terms of the CV $\sigma(r_1,r_2,\ldots\ldots,r_N)$, which is dependent on the coordinates of the system. This potential is zero at the beginning of the simulation, but accumulates over time and in principle exactly negates the potential energy in terms of the defined collective variable and thereby flattens the PES. This means that $U(t|\xi)+f(\xi)$ is exactly zero, where $f(\xi)$ is the potential of mean force (PMF) (or the Landau free energy defined as $f(\xi) = -k_B T \ln p(\xi)$) and $p(\xi)$ is the probability density estimate ($p(\xi) = \langle\delta[\xi - \sigma(r_1,r_2,\ldots\ldots,r_N)]\rangle$). The application of ABMD requires the choice of two parameters: the flooding time, τ_F, and the kernel width, $4\Delta\xi$, which is related to the resolution of the kernel $G(\xi)$. As in recent ABMD studies of oligoprolines,[6,22-23] the following collective variable was used in calculations in this thesis (Chapters 4 and 5), which is based on the cosines of the ω dihedral angle:

$$\Omega = \sum_i \cos(\omega_i) \quad 2.12$$

Thus, Ω sums to +9 and −9 for PPI (all *cis* or $\omega = 0°$) and PPII (all *trans* or $\omega = 180°$, Figure 2.2) structures, respectively.

The ABMD methodology can be combined with replica exchange as discussed above to further accelerate conformational sampling.[24] Specifically, two variations of ABMD were implemented in this thesis: (1) a replica exchange variant with each replica at a different temperature and biased by different potentials; and (2) a multiple walker strategy with all replicas at one temperature and biased by the same potential.

The PMFs obtained by applying Equations 2.10 and 2.11 can be further refined using umbrella sampling calculations. This approach is based on the principle that if the PMF is very close to the exact value, then a flat histogram will be obtained in the space of the collective variable or, in other words, all values will be equally accessible. However, if there is a small difference (a few k_BT) between the calculated PMF and the true value, then the histogram obtained from umbrella sampling can serve as a correction.

To obtain equilibrium data from the ABMD calculations, an HT–REMD (Hamiltonian temperature–REMD) scheme can be implemented using a previously obtained PMF. In these calculations, additional replicas are added to an initially simulated collection of replicas. One of the additional replicas is completely unbiased, while the others are biased based on the PMF of the lowest temperature replica (of the previously simulated replicas) and scaled by a range of factors. This procedure enhances the exchange to the unbiased replica from which equilibrium data can be recorded. The HT-REMD procedure is particularly useful since it combines the efficiency of the ABMD method with the generality of REMD calculations to yield unbiased information.

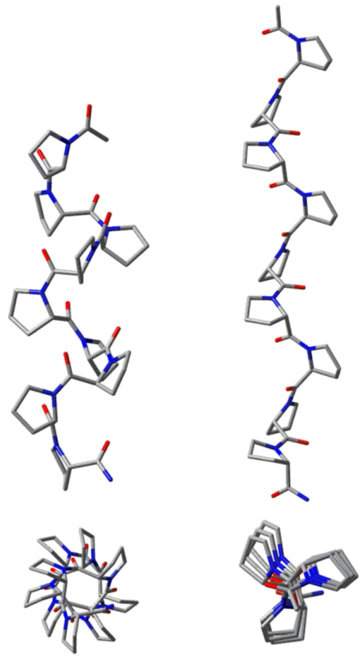

Figure 2.2. Polyproline I (PPI) and polyproline II (PPII) helices of a nonaproline model compound (hydrogen atoms removed for clarity).

2.4. Solvation

Biological processes occur in an aqueous environment and solvation has a profound effect on the structure of biomolecules. For example, hydrophobic interactions have been shown to be a dominant driving force in the protein folding process.[25-26] As a consequence various solvation models, have been developed ranging from implicit models such as Polarizable Continuum Model (PCM) model,[27-28] to explicitly parameterized water models, such as the TIP3P.[29-31] Implicit solvation models typically account for solvation effects using an estimate of the solvation free energy of the solute in a given conformation,[28,32] while the solvent is treated as a continuum. In the QM calculations carried out in this thesis, PCM solvation developed by Tomasi and coworkers is utilized where the solvation free energy (G_{sol}) is:

$$G_{sol} = G_{elec} + G_{dr} + G_{cav} \qquad 2.13$$

and G_{elec}, G_{dr} and G_{cav} are the electrostatic, dispersion-repulsion and cavitation contributions, respectively. The various terms for G_{sol} are calculated using a solvent cavity defined by interlocking van der Waals spheres.

In addition, the generalized Born (GB) implicit solvent model provided in the AMBER software suite is used in the implicit solvent MD simulations reported in this thesis. This model is similar to the PCM model described above and consists of electrostatic and nonelectrostatic contributions. Specifically, the calculations presented within use the GB model developed by Onufriev, Bashford and Case[33-34] to describe the electrostatic solvation energy:

$$G_{elec} = \frac{1}{2}\sum_{ij} \frac{q_i q_j}{f^{GB}(r_{ij}, R_i, R_j)} g(\epsilon_w, f^{GB}) \qquad 2.14$$

where r_{ij} is the distance between atoms having partial charges q_i and q_j, and effective Born radii of R_j and R_j. The dielectric constant is ϵ_w, while the function $f^{GB}(r_{ij}, R_i, R_j)$ is given as $\left[r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2/4R_i R_j\right)\right]^{\frac{1}{2}}$. The non-electrostatic component is approximated by

$$G_{nonelec} = \gamma SASA \qquad 2.15$$

where SASA is the solvent accessible surface area and γ is a surface tension parameter.

The three dimensional reference interaction site model[35-36] (3D-RISM) solvation method was also applied in this thesis to calculate monomer to triple helix conversion energetics for the collagen model peptides. This method is based on solvent radial distribution functions (RDFs), which describe solvent distributions around a solute, and is thus superior to the GB methods in modeling solvation. However, 3D-RISM calculations are computationally more expensive than explicit solvent calculations, and therefore were only used in this thesis to describe solvation of snapshots obtained from MD simulations.

Finally, explicit solvent molecules were used in many of the MD calculations described in this thesis to provide discrete solute-solvent interactions that are not correctly modeled using implicit solvation. In particular, the TIP3P water model was used since it reasonably reproduces the density of water, is computationally not very expensive and has been widely used in biomolecular simulations. The TIP3P model has fixed bonds that are constrained to their starting value using the SHAKE algorithm.[37-38]

While implicit solvation can lead to significant reductions in computational cost, certain solute–solvent interactions, such as specific water–solute interactions hydrogen bonds are not properly described.[32] Thus, while it is desirable to utilize explicit solvation for all the calculations carried out in this thesis, a mixture of implicit and explicit is used in most cases to reduce computational cost. For instance, in Chapter 4, ABMD calculations are first carried out in implicit solvent to obtain a PMF, which is then refined in explicit solvent to describe discrete solute–solvent interactions. Additionally, all first principles calculations utilized implicit solvation since it is computationally very expensive to consider explicit QM solvent molecules. However, the QM calculations used conformations at minima obtained from explicit solvent MD calculations as starting structures. In Chapter 3 some explicit water molecules were also optimized using QM.

2.5. Some Practical Considerations

2.5.1. Periodic Boundary Conditions

As mentioned previously, solvation is important for a correct description of biochemical processes. In particular, it has been shown in the literature that solvation is necessary to model the structural preferences of sugars,[7] which form part of the glycopeptides modeled in this thesis. Thus, starting solute (peptide) structures for most MD simulations reported in this thesis are surrounded by a box of water molecules. However, this box alone is not enough to describe solvation since the water molecules near the solute, in the bulk and at the edge of the box are in different environments. To overcome this problem, the periodic boundary condition (PBC) was implemented. In PBC calculations, the box containing both solvent and solute is replicated in all directions to mimic the solution phase. This approximation also addresses problems with solvent molecules at the edge of the simulation box which would otherwise be in a different environment from solvent molecules in the bulk. Under PBC atoms that diffuse from the central box are replaced by a mirror image from another box. The particle mesh Ewald (PME) summation[39] technique is used to compute long range electrostatic interactions (Equation 2.5) in the periodic box. The PMEMD module of the AMBER software suite, which implements the PME method, was used for this purpose.

In the explicit solvent MD calculations reported in this thesis, a layer of TIP3P[40] water molecules with a thickness of 8 Å surrounded the dipeptide systems (Chapter 3), 10 Å thickness for the oligoprolines (Chapters 4 and 5) and 15 Å thickness for the collagen model peptides (Chapter 6). Since more water molecules around the solute can significantly increase the simulation time, an octahedron box obtained by truncating the ends of the cubic box was used in most cases to reduce the amount of water molecules needed to solvate the solute. An octahedral box can also be replicated in all directions around the central box, which achieves the same objective as a cubic box.

2.5.2. Minimization and Equilibration

As described in the previous subsection, the starting structure must undergo certain preparatory steps before production MD simulations. Structure minimization is usually the first step in this process. In this thesis, the steepest descent (SD) method was used to initiate minimization. SD minimizes downhill in the direction of the local gradient and can quickly bring the structure close to a local minimum. Since SD suffers from instabilities, the conjugate gradient (CG) method, which uses the gradient history to better decide the minimization direction, was used to complete the process. It should be noted that the primary goal of minimization of the solvated structure is not to find the global energy minimum, but to reduce steric clashes, which introduce instabilities in the MD simulation.

Subsequently, velocities must be assigned to the particles from a Boltzmann distribution. Since the velocities are typically randomly assigned to each particle, there is a need to equilibrate the system at a particular temperature. This was done in the present thesis by adjusting the velocities of individual particles over the simulation run until the instantaneous temperature, potential and kinetic energies oscillate with a small variance about a given mean value. This temperature equilibration is typically carried out at constant number of particles (N) and volume (V) (NVT ensemble). Temperature regulation was implemented by the use of a thermostat, which involves the addition of an extra stochastic friction term to the equations of motion (Equation 2.6). In this thesis, the Langevin thermostat was used since it correctly reproduces a canonical (NVT) distribution. In addition, the initial box was created by adding a layer of water molecules around the solute, which generate open spaces in the simulation box that could introduce instabilities. Therefore, to obtain a well defined density, MD calculations were carried out while slowly adjusting the (using a barostat) volume of the box at constant pressure (NPT). This was done until there were small oscillations about a mean density as the simulation progresses. The final structures obtained from this process were then used for production MD calculations.

2.6 Quantum Mechanics Based Calculations

Apart from the MM approximation, the potential, U (Equation 2.1), can be obtained from first principles using quantum mechanics (QM). This approach is based on obtaining an approximate solution to the time-independent Schrödinger equation (SE) for the molecular wavefunction ($\boldsymbol{\Psi}$):

$$\hat{\mathrm{H}}\boldsymbol{\Psi}(\mathbf{r},\mathbf{R}) = \mathrm{E}\boldsymbol{\Psi}(\mathbf{r},\mathbf{R}) \qquad 2.16$$

where the Hamiltonian ($\hat{\mathrm{H}}$) is given by:

$$\hat{\mathrm{H}} = -\frac{\mathrm{h}}{2\pi \mathrm{m}_e}\sum_{\mathrm{i}} \nabla_{\mathrm{i}}^2 - \frac{\mathrm{h}}{2\pi}\sum_{\mathrm{I}} \mathrm{M}_{\mathrm{I}}\nabla_{\mathrm{I}}^2 - \sum_{\mathrm{i,I}} \frac{\mathrm{Z_I e^2}}{\mathbf{r}_{\mathrm{i}}-\mathbf{R}_{\mathrm{I}}} + \frac{1}{2}\sum_{\mathrm{i}\neq \mathrm{j}} \frac{\mathrm{e}^2}{\mathbf{r}_{\mathrm{i}}-\mathbf{r}_{\mathrm{ij}}} + \sum_{\mathrm{I}\neq \mathrm{J}} \frac{\mathrm{Z_I Z_J e^2}}{\mathbf{R}_{\mathrm{I}}-\mathbf{R}_{\mathrm{J}}} \qquad 2.17$$

where h is Planck's constant, e is the electronic charge, $\mathrm{Z_I}$ is the atomic number, $\mathbf{r}_{\mathrm{i}}$ and $\mathbf{R}_{\mathrm{I}}$ are the electronic and nuclear positions, respectively, and m_e and $\mathrm{M_I}$ are the electronic and nuclear masses, respectively.

There are several approximate solutions for the molecular SE based on the above Hamiltonian. However, the density functional theory (DFT) approach is the most popular due to its relatively good balance between accuracy and computational cost. Specifically, DFT casts the electronic structure problem in terms of the density ($\rho(\mathbf{r})$), which is dependent on just three variables instead of the many-body (3N) problem encountered when using the wavefunction ($\boldsymbol{\Psi}$). Within the DFT scheme, the total electronic ground state energy of the system can be written as a functional of the density as:

$$E[n(\mathbf{r})] = T_s[\rho(\mathbf{r})] + E_{Hartree} + \int d\mathbf{r}V_{ext}\,\rho(\mathbf{r}) + E_{xc}[\rho(\mathbf{r})] \qquad 2.18$$

where $T_s[\rho(\mathbf{r})]$ is the kinetic energy functional, $E_{Hartree}$ is the electron-electron interaction, the integral is the electron–nuclear interaction, and $E_{xc}[\rho(\mathbf{r})]$ is the exchange–correlation energy functional. E_{xc}, which is usually divided into the exchange E_x and correlation E_c components, contains all interactions not included in the other terms and is the principal approximation in DFT.

There are several approximations of $E_{xc}[\rho(\mathbf{r})]$ including the local density approximation (LDA), which describes $E_{xc}[\rho(\mathbf{r})]$ in terms of the electron density only:

$$E_{XC}^{LDA}[\rho(\mathbf{r})] = \int \epsilon_{XC}\,(\rho(r))\rho(r)d^3r \qquad 2.19$$

where ϵ_{XC} is the exchange-correlation energy per particle of a uniform electron gas. The LDA accurately describes the structure and energetics of metals, but does not correctly describe the interactions in complex molecular systems, such as biomolecules. Improved performance is achieve by including the gradient of the electron density into the calculation of $E_{xc}[\rho(\mathbf{r})]$, which constitutes the generalized gradient approximation (GGA):

$$E_{XC}^{GGA}[\rho(\mathbf{r})] = \int \epsilon_{XC}\,(\rho(r), \nabla\rho(r))\rho(r)d^3r \qquad 2.20$$

It was further determined that adding a portion of exact Hatree-Fock (HF) exchange to the GGA further improves the $E_{xc}[\rho(\mathbf{r})]$ description of molecular systems:

$$E_{XC}^{Hybrid} = aE_X^{HF} + (1-a)E_{XC}^{GGA} \quad 2.21$$

The hybrid exchange and correlation functional developed by Becke, Lee, Yang and Parr (B3LYP) is used in this thesis since it has been shown in the literature to yield accurate results for systems similar to those considered herein. [17,41-47]

In QM calculations, the orbitals are usually expanded in the basis of a finite number of functions (basis set). The size of the basis set is important in determining the accuracy of the calculation. In addition, basis set superposition errors[48] (BSSEs) can occur when the interaction energies of weakly bound systems are calculated using a small basis set. Specifically, BSSE occurs when the molecular association is artificially stabilized because it is better described by a combination of the basis sets of its components. For instance, in a dimer system consisting of monomers A and B, the dimer can be artificially stabilized if monomer A borrows accessible basis functions from B in the dimer structure and better describes the molecular association as a result.

BSSE can be corrected using the Boys and Bernardi counterpoise (CP)[49] approximation. In this scheme, a correction term (Equation 2.22) is added to the dimer energy or, in other words,

$$\Delta E_{int}^{CP}(AB) = E_{AB}^{AB} + E_A^{AB} + E_B^{AB} \quad 2.22$$

where ΔE_{int}^{CP} is the CP correction, ΔE_{AB}^{AB} is the energy of the dimer using the dimer basis set, ΔE_A^{AB} is the energy of monomer A in the dimer basis set and ΔE_B^{AB} is the energy of monomer B in the dimer basis set. All DFT calculations in this thesis were carried out using the Gaussian 09 software.[50]

2.7. Strengths and Weaknesses of Computational Methodology

The two approaches (MM and DFT) outlined above were used to derive the interaction potential for molecular systems considered in this thesis. The force field method is a basic approximation of the true potential. However, it has been shown in the literature that careful parameterization and application of MM forcefields yields accurate structural and energetic information in comparison to experiments.[4,7-8] The AMBER FF99SB and GLYCAM parameter sets used in this thesis have been extensively applied to study oligoproline and glycan(glycopeptide) systems in the literature and have been shown to correctly describe structure and energetics.[5-8,20,51-56] The MM methodology has the advantage of being computationally inexpensive, which allows for the simulation of large systems of biological interest over relatively long simulation times. This is particularly relevant for the glycopeptide systems studied in this thesis. Specifically, it has been shown in the literature[54,57] that long simulation times of the order of hundreds of nanosecond (ns) are required to adequately sample the rotamers of the exocyclic dihedrals of sugar systems.

As mentioned earlier, in MM calculations, biomolecules are typically built from amino acid residues (or monosaccharides) that have been separately parameterized. Therefore, in general, it is difficult to make *a priori* determination of expected errors, particularly for energetics. Thus, the ultimate test of the applied MM force field is reproducing experimental data. In this light, it is important that relative trends in energertics are reproduced. In particular, for the oligoprolines considered in this thesis, the force field should energetically distinguish between the PPI and PPII conformations, as well as intermediate structures. The AMBER FF99SB parameter set used in this thesis has been shown to accurately distinguish between these conformations in water, where the PPII conformation is more stable than PPI, as well as in propanol, where PPI is more stable than PPII.[6,52]

The DFT method, on the other hand, is in principle more accurate than MM since the electronic structure is used to derive the interaction potential. As a result, it can properly describe effects, such as polarization by nearby groups or molecules. However, the computational cost of DFT calculations increases very steeply with the number of degrees of freedom in the molecular system. As a consequence, DFT calculations in the literature rarely consider more than 100 atoms in the model system. Notably in this thesis, DFT calculations on triple helix models consisting of 27 amino acids for the (Pro–Hyp–Gly)$_3$ and (Pro–Hyp–Gly)–(Pro–Hyp[β-Galactose]–Gly)–(Pro–Hyp–Gly) were composed of 351 and 432 atoms, respectively. To my knowledge, these models are the largest all electron DFT calculations on such collagen model peptides to date.[46-47,58]

The B3LYP functional used in this thesis has been shown in the literature to accurately model the structure of many biomolecular systems. For example, a recent study compared structures and energetics calculated using several DFT methods to those obtained from high-level coupled cluster (CCSD (T)) calculations for a series of furanose, pyranose and open chain monosaccharides.[59] Root mean square errors of 1.43 kJ mol^{-1}, 1.9 kJ mol^{-1} and 2.46 kJ mol^{-1} where obtained for isomerisation of exocyclic hydroxyls, conversion between anomeric forms and ring structure conversions, respectively. The results compared B3LYP/def2-TZVPP calculations to CCSD(T)/CBS (extrapolation to the complete basis set) for a series of eight pyranose sugars including galactose, which forms a glycosidic linkage to Hyp in the glycopeptides studied in this thesis. The furanose and open chain monosaccharides were not as accurately modeled as the pyranose isomers. Another recent study characterized the hydrogen-bonding complexes of DNA fragments and determined that B3LYP has a mean unsigned error (MUE) of 1.4 kJ mol^{-1} even when a small basis set was used.[60] With this knowledge of the deficiencies in a DFT method, trends in calculated energies can be used to rationalize and complement experimental information.

Optimized structures can be compared to experimental crystal structures when such structures exist to help assess the accuracy of the applied DFT methodology. This was done in Chapter 6 of this thesis, where optimized triple helix structures were compared to experimental crystal structures. It is well known that the B3LYP functional does not properly describe dispersion interactions;[61-62] however, the aforementioned study[60] concluded that structure is well reproduced by the method, as long as the model is large enough to include appropriate steric constraints. The DFT models used for the collagen model peptides considered in this thesis (Chapter 6) are large enough to provide such sterics. This is evidenced by how well structural details, such as backbone dihedral angle values for the (Pro–Hyp–Gly)$_3$ peptide, compare to experimental crystal structures.

2.8. Experimental Methods

2.8.1 Determination of Melting Temperatures Using Circular Dichroism Spectroscopy.

As outlined in the previous section, it is important to compare computational results to experimental data. Structural changes, such as conversion between the PPI and PPII conformations for the oligopeptide systems considered in thesis, can be monitored experimentally using circular dichroism (CD) spectroscopy. Subsequently, data from such CD spectra is processed to obtain melting temperatures (T_m), which correlate with to the stability of the PPII conformation. In this section, the CD methodology and how it was used to determine the T_m referenced in this thesis are briefly outlined. CD is a spectroscopic tool that uses the difference in absorption of right and left circularly polarized radiation.[63-65] The absorbance (A_{rcp}) of right circularly polarized light can be written as:

$$A_{rcp} = log_{10}\left(\frac{I^o}{I}\right) = \varepsilon_{rcp}Cl \qquad 2.22$$

where Beer's law has been utilized in the right hand side of the equation. I^o (I) is the intensity of the incident (transmitted) radiation, ε_{rcp} is the molar extinction coefficient of the solution, C is the concentration and l is path traveled by the radiation. Applying a similar analysis for left circular polarized (lcp) radiation and subtracting gives:

$$\Delta A = \varepsilon_{lcp}Cl - \varepsilon_{rcp}Cl = \Delta\varepsilon Cl \qquad 2.23$$

where $\Delta\varepsilon = \varepsilon_{lcp} - \varepsilon_{rcp}$ is the molar extension coefficient. According to the experimentally obtained and theoretically defined relationships for $\Delta\varepsilon$ to rotational strength (how electronic transitions contribute to CD), $\Delta\varepsilon$ can only be non-zero for chiral molecules.[66]

In practice, CD is usually reported in terms of the molar ellipticity, which is related to $\Delta\varepsilon$ as:

$$[\theta] = 3298\ \Delta\varepsilon \qquad 2.24$$

The concentration and mean residue molecular weight of the sample, as well as the wavelength, are required to completely describe the CD spectra.

The CD signature of many secondary structure elements of biomolecules has been determined in the literature.[67-68] The polyproline II (PPII) conformation that is important for the (glyco)peptides considered in this thesis has two distinctive extrema in the molar ellipticity. Specifically, a positive band between 220 to 230 nm and a negative band between 195 and 210 nm in the CD spectra is characteristic of the PPII structure,[68-69] with a change in the intensity of the corresponding minima usually being correlated with changes in the PPII content.[70] Using this correlation between the magnitude in ellipticity and PPII stability, T_m for such oligopeptide systems were estimated by computing the first derivative of [θ] as a function temperature.[71]

Schweizer and coworkers made use of the same technique to determine T_m for the oligopeptides considered in this thesis.[72] Figure 2.3 shows the molar ellipticity of the Ac–[Hyp–(β-Gal)]$_9$–NH_2 peptide, as a function of temperature and the first derivative of this data with respect to temperature, which gives $T_m \approx 70$ °C. Fitting the CD data to a two state model provided the same values for T_m. In this thesis, T_m determined by CD is compared to calculated free energy difference between the PPI and PPII conformations (ΔF(PPII-PPI)) to relate experimental and calculated PPII stability measures in line with other literature studies.[6,73]

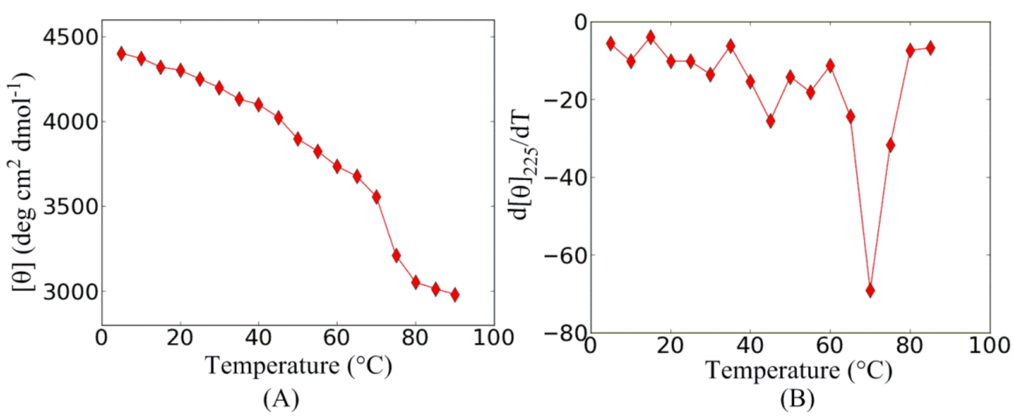

Figure 2.3. (A) Ellipticity at 225 nm of Ac–[Hyp–(β-Gal)]$_9$–NH_2 as a function of temperature in water from 5 to 90°C and (B) first the derivative of the molar ellipticity as a function of temperature.

2.9. References[b]

1. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995,** *16*, 1357–1377.

2. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179–5197.

3. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65*, 712–725.

4. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29*, 622–655.

5. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., A Statistical Analysis of the PPII Propensity of Amino Acid Guests in Proline-Rich Peptides. *Biophys. J.* **2011,** *100*, 1083–1093.

6. Moradi, M.; Babin, V.; Roland, C.; Darden, T. A.; Sagui, C., Conformations and free energy landscapes of polyproline peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 20746–20751.

7. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 10541–10545.

8. Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B., Effects of Glycosylation on Peptide Conformation: A Synergistic Experimental and Computational Study. *J. Am. Chem. Soc.* **2004,** *126*, 8421–8425.

9. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26*, 1668–1688.

10. Verlet, L., Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Physical Review* **1968,** *165*, 201–214.

11. Geyer, C. J., Markov-chain Monte Carlo Maximum-likelihood. *Computing Science and Statistics* **1991**, 156–163.

12. Sindhikara, D.; Meng, Y.; Roitberg, A. E., Exchange frequency in replica exchange molecular dynamics. *J. Chem. Phys.* **2008,** *128*, 024103–024110.

[b] ACS referencing style was implemented throughout this thesis.

13. Wang, L.; Friesner, R. A.; Berne, B. J., Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011,** *115*, 9431–9438.

14. García, A. E.; Sanbonmatsu, K. Y., Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins* **2001,** *42*, 345–354.

15. Shen, T. Y.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S., Conformational Flexibility of Soluble Cellulose Oligomers: Chain Length and Temperature Dependence. *J. Am. Chem. Soc.* **2009,** *131*, 14786–14794.

16. Ellis, C. R.; Maiti, B.; Noid, W. G., Specific and Nonspecific Effects of Glycosylation. *J. Am. Chem. Soc.* **2012,** *134*, 8184–8193.

17. Kang, Y. K.; Choi, H. Y., Cis-trans isomerization and puckering of proline residue. *Biophys. Chem.* **2004,** *111*, 135–142.

18. Kang, Y. K., Conformational preferences of non-prolyl and prolyl residues. *J. Phys. Chem. B* **2006,** *110*, 21338–21348.

19. Krimm, S.; Bandekar, J., Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem.* **1986,** *38*, 181–364.

20. Babin, V.; Karpusenka, V.; Moradi, M.; Roland, C.; Sagui, C., Adaptively Biased Molecular Dynamics: An Umbrella Sampling Method With a Time-Dependent Potential. *Int. J. Quant. Chem.* **2009,** *109*, 3666–3678.

21. Babin, V.; Roland, C.; Sagui, C., Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **2008,** *128*, 134101–134108

22. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., PPII Propensity of Multiple-Guest Amino Acids in a Proline-Rich Environment. *J. Phys. Chem. B* **2011,** *115*, 8645-8656.

23. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., A Statistical Analysis of the PPII Propensity of Amino Acid Guests in Proline-Rich Peptides. *Biophys. J.* **2011,** *100*, 1083–1093.

24. Babin, V.; Sagui, C., Conformational free energies of methyl-alpha-L-iduronic and methyl-beta-D-glucuronic acids in water. *J. Chem. Phys.* **2010,** *132*, 104108–104115

25. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990,** *29*, 7133-7155.

26. Dobson, C. M.; Sali, A.; Karplus, M., Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem. Int. Ed.* **1998,** *37*, 868–893.

27. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105*, 2999–3094.

28. Roux, B.; Simonson, T., Implicit solvent models. *Biophys. Chem.* **1999,** *78*, 1–20.

29. Jorgensen, W. L., Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981,** *103*, 335–340.

30. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926–935.

31. Mahoney, M. W.; Jorgensen, W. L., Quantum, intramolecular flexibility, and polarizability effects on the reproduction of the density anomaly of liquid water by simple potential functions. *J. Chem. Phys.* **2001,** *115*, 10758–10768.

32. Chen, J.; Brooks, C. L., Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. *Phys. Chem. Chem. Phys.* **2008,** *10*, 471–481.

33. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000,** *104*, 3712-3720.

34. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004,** *55*, 383–394.

35. Kovalenko, A.; Hirata, F., Three-dimensional density profiles of water in contact with a solute of arbitrary shape: A RISM approach. *Chem. Phys. Lett.* **1998,** *290*, 237–244.

36. Kovalenko, A.; Hirata, F., Self-consistent description of a metal-water interface by the Kohn-Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999,** *110*, 10095–10112.

37. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23*, 327–341.

38. Miyamoto, S.; Kollman, P. A., Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992,** *13*, 952–962.

39. Toukmaji, A.; Sagui, C.; Board, J.; Darden, T., Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *J. Chem. Phys.* **2000,** *113*, 10913–10927.

40. Jorgensen, W. L.; Jenson, C., Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J. Comput. Chem.* **1998,** *19*, 1179–1186.

41. Momany, F. A.; Willett, J. L.; Schnupf, U., DFT molecular dynamics (DFTMD) simulations of carbohydrates: COSMO solvated alpha-maltose. *Theochem-J. Mol. Struct.* **2010,** *953*, 61–82.

42. Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L., B3LYP/6-311++G** study of monohydrates of alpha- and beta-D-glucopyranose: hydrogen bonding, stress energies, and effect of hydration on internal coordinates. *Carb. Res.* **2004,** *339*, 553–567.

43. Momany, F. A.; Schnupf, U., DFTMD studies of beta-cellobiose: conformational preference using implicit solvent. *Carb. Res.* **2011,** *346*, 619–630.

44. Aliev, A. E.; Courtier-Murias, D.; Bhandal, S.; Zhou, S., A combined NMR/MD/QM approach for structure and dynamics elucidations in the solution state: pilot studies using tetrapeptides. *Chem. Commun.* **2010,** *46*, 695–697.

45. Aliev, A. E.; Courtier-Murias, D., Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly. *J. Phys. Chem. B* **2010,** *114*, 12358–12375.

46. Xu, Y.; Dannenberg, J. J., Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProDAla, and ProProDSer Triads. *J. Am. Chem. Soc.* **2005,** *127*, 14130–14131.

47. Palfi, V. K.; Perczel, A., Stability of the Hydration Layer of Tropocollagen: A QM Study. *J. Comput. Chem.* **2010,** *31*, 764–777.

48. Simon, S.; Duran, M.; Dannenberg, J. J., How does basis set superposition error change the potential surfaces for hydrogen bonded dimers? *J. Chem. Phys.* **1996,** *105*, 11024–11031.

49. Boys, S. F.; Bernardi, F., Calculation of Small Molecular Interactions by Differences of Seperate Total Energies - Some Procedures with Reduced Errors. *Mol. Phys.* **1970,** *19*, 553-566.

50. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, ., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. , Gaussian 09, Revision A.1, Gaussian, Inc.,: 2009.

51. Moradi, M.; Babin, V.; Roland, C.; Sagui, C., A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *J. Chem. Phys.* **2010,** *133*, 124104–125122.

52. Moradi, M.; Lee, J.-G.; Babin, V.; Roland, C.; Sagui, C., Free Energy and Structure of Polyproline Peptides: An Ab Initio and Classical Molecular Dynamics Investigation. *Int. J. Quant. Chem.* **2010,** *110*, 2865–2879.

53. Seo, M.; Castillo, N.; Ganzynkowicz, R.; Daniels, C. R.; Woods, R. J.; Lowary, T. L.; Roy, P. N., Approach for the simulation and modeling of flexible rings: Application to the alpha-D-arabinofuranoside ring, a key constituent of polysaccharides from Mycobacterium tuberculosis. *J. Chem. Theory Comput.* **2008,** *4*, 184–191.

54. Seo, M.; Castillo, N.; Ganzynkowicz, R.; Daniels, C. R.; Woods, R. J.; Lowary, T. L.; Roy, P.-N., Approach for the Simulation and Modeling of Flexible Rings: Application to the α-d-Arabinofuranoside Ring, a Key Constituent of Polysaccharides from Mycobacterium tuberculosis. *J. Chem. Theory Comput.* **2007,** *4*, 184–191.

55. Stortz, C. A.; Johnson, G. P.; French, A. D.; Csonka, G. I., Comparison of different force fields for the study of disaccharides. *Carb. Res.* **2009,** *344*, 2217–2228.

56. Salisburg, A. M.; Deline, A. L.; Lexa, K. W.; Shields, G. C.; Kirschner, K. N., Ramachandran-Type Plots for Glycosidic Linkages: Examples from Molecular Dynamic Simulations Using the Glycam06 Force Field. *J. Comput. Chem.* **2009,** *30*, 910–921.

57. Taha, H. A.; Castillo, N.; Sears, D. N.; Wasylishen, R. E.; Lowary, T. L.; Roy, P.-N., Conformational Analysis of Arabinofuranosides: Prediction of 3JH,H Using MD Simulations with DFT-Derived Spin−Spin Coupling Profiles. *J. Chem. Theory Comput.* **2009,** *6*, 212–222.

58. Pálfi, V. K.; Perczel, A., How stable is a collagen triple helix? An ab initio study on various collagen and β-sheet forming sequences. *J. Comput. Chem.* **2008,** *29*, 1374–1386.

59. Sameera, W. M. C.; Pantazis, D. A., A Hierarchy of Methods for the Energetically Accurate Modeling of Isomerism in Monosaccharides. *J. Chem. Theory Comput.* **2012,** *8*, 2630–2645.

60. Kellie, J. L.; Wetmore, S. D., Selecting DFT methods for use in optimizations of enzyme active sites: applications to ONIOM treatments of DNA glycosylases. *Can. J. Chem.* **2013,** *91*, 559–572.

61. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J., General Performance of Density Functionals. *J.Phys. Chem. A* **2007,** *111*, 10439–10452.

62. Rutledge, L. R.; Wetmore, S. D., The assessment of density functionals for DNA–protein stacked and T-shaped complexes. *Can. J.Chem.* **2010,** *88*, 815–830.

63. Woody, R. W., Circular Dichroism Spectrum of Peptides in the Poly(Pro)II Conformation. *J. Am. Chem. Soc.* **2009,** *131*, 8234–8245.

64. Sreerama, N.; Woody, R. W., Molecular dynamics simulations of polypeptide conformations in water: A comparison of alpha, beta, and poly(Pro)II conformations. *Proteins* **1999,** *36*, 400–406.

65. Sreerama, N.; Woody, R. W., Poly(Pro) II Helices in Globular Proteins - Identification and Circular Dichroic Analysis. *Biochemistry* **1994,** *33*, 10022-10025.

66. Fasman G. D., *Circular Dichroism and the Conformational Analysis of Biomolecules*. 1996.

67. Ronish, E. W.; Krimm, S., Theoretical Calculation of Circular-Dichroism of Unordered Polypeptide Chains. *Biopolymers* **1972,** *11*, 1919.

68. Ronish, E. W.; Krimm, S., Calculated Circular-Dichroism of Polyproline II in Polarizibility Approximation. *Biopolymers* **1974,** *13*, 1635–1651.

69. Pysh, E. S., Randon-phase Calculation of Poly-L-Proline II Circular-Dichroism. *Biopolymers* **1974,** *13*, 1563–1571.

70. Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A. F.; Creamer, T. P., Host-guest study of left-handed polyproline II helix formation. *Biochemistry* **2001,** *40*, 14376–14383.

71. Horng, J.-C.; Raines, R. T., Stereoelectronic effects on polyproline conformation. *Protein Sci.* **2006,** *15*, 74–83.

72. Owens, N. W.; Stetefeld, J.; Lattova, E.; Schweizer, F., Contiguous O-Galactosylation of 4(R)-Hydroxy-L-proline Residues Forms Very Stable Polyproline II Helices. *J. Am. Chem. Soc.* **2010,** *132*, 5036–5042.

73. Kuemin, M.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H., Effects of Terminal Functional Groups on the Stability of the Polyproline II Structure: A Combined Experimental and Theoretical Study. *J. Am. Chem. Soc.* **2009,** *131*, 15474–15482.

Chapter 3. Conformational Study of the Hydroxyproline–O–Glycosidic Linkage: Sugar–Peptide Orientation and Prolyl Amide Isomerization in (α/β)–Galactosylated 4(R/S)–Hydroxyproline[a]

3.1. Introduction

As outlined in the previous chapter, glycosylation is a common post–translational modification of proteins.[1-3] Since it is difficult to crystallize natural glycoproteins, many small-model glycopeptides have been synthesized to study how glycosylation affects peptide structure.[4-5] These studies have concluded that the structure of the glycosidic linkage not only influences the relative orientation of the sugar and peptide, but also influences the peptide backbone conformation and hence peptide properties.[6-9] However, until recently, there was no literature investigation of the configuration of the hydroxyproline O-glycosidic linkage, which is of importance to the structure of plant hydroxyproline rich glycoproteins (HRGPs).

This lack of information was recently alleviated by experimental work on glycosylated monomers of N–acetyl–(2S,4R)–4–hydroxyproline methyl ester (Hyp). $K_{trans/cis}$ determined by NMR spectroscopy at 24.8 °C demonstrated that glycosylation has no measurable influence on amide isomerization in Hyp,[10] but glycosylation affects both the *N*-terminal amide equilibrium and the rate of amide isomerization of N–acetyl-(2S,4S)–4–hydroxyproline methyl ester (hyp).[11] Indeed, both α- and β-anomeric galactosyl linkages stabilize the *trans* amide conformation of hyp by 7% in model amides with an associated 25–50% increase in isomerization rate. The reason for this observation was not clear from analysis of the experimental data. Measurements of ^3J coupling constants for Hyp (hyp) and

[a] Reprinted with permission from E. B. Naziga, F. Schweizer and S. D. Wetmore. Conformational Study of the Hydroxyproline–O–Glycosidic Linkage: Sugar–Peptide Orientation and Prolyl Amide Isomerization in (α/β)–Galactosylated 4(R/S)–Hydroxyproline. *J. Phys. Chem. B*. **2012**, *116*, 860-871.

glycosylated derivatives indicated that glycosylation preserves the puckering preferences of both compounds, which is linked to *trans* stability in proline derivatives. In particular, $J_{\alpha,\beta1}$ and $J_{\alpha,\beta2}$ values of 7.8 to 9.8 Hz and 2.4 to 4.3 Hz were derived for Hyp and its glycosylated variants, respectively, which correspond to the C^{γ}-exo pucker. On the other hand, $J_{\alpha,\beta1}$ ($J_{\alpha,\beta1}$) of 8.2 to 8.3 Hz (8.5 to 8.7 Hz) were obtained for hyp (glycosylated hyp), corresponding to the C^{γ}-endo pucker. Additionally, an approximately 9–10 ppm difference in the $^{13}C^{\gamma}$ chemical shift was obtained between glycosylated and unglycosylated variants of both stereoisomers, showing that inductive effects are not responsible for the differences between the glycosylated versions of Hyp and hyp. It was instead proposed that the loss of intramolecular hydrogen bonding due to glycosylation of hyp could destabilize the *cis* conformation, and hence favour the *trans* isomer. However, the measured $K_{trans/cis}$ for a proline derivative having a 4S methylated C^{γ} atom is very similar to that for unglycosylated hyp, meaning that this hydrogen bond is not heavily populated in aqueous solution. Overall, an explanation of the increase in *trans* population and isomerization rate upon glycosylation of hyp was not provided by the experimental data. However, it could involve differences in the orientation of the sugar with respect to the peptide in Hyp and hyp, since the sugar is projected in different directions by the C^{γ}-*exo* and C^{γ}-*endo* puckers. Therefore, in addition to providing the structure of the Hyp-O-glycosidic linkage as observed in HRGPs, molecular modeling will provide atomic level details that will explain the experimental findings.

The present chapter applies a combined quantum and molecular mechanical approach to examine the conformational properties of the two hydroxyproline stereoisomers: (i) Hyp and (ii) hyp, together with their α/β–galactosylated derivatives (Figure 3.1). A focus is placed on the conformation of the Hyp (hyp)–O–glycosidic linkage, the effects of glycosylation on the structure and hydration of the peptide backbone, and deviations in the *cis*/*trans* isomerization upon glycosylation. Although (hydroxy) proline is fairly rigid, a dynamical methodology must be used since the glycopeptide includes a carbohydrate moiety with inherent flexibility. Additionally, the solvent environment must be carefully modeled since the presence of solvent (water) has been previously shown by MD to be required to correctly sample the rotamers of exocyclic hydroxyls,[12] and the *cis/trans* distribution in related compounds.[13] As discussed in Chapter 1, a two-pronged approach is implemented, which includes MD simulations with explicitly included solvent molecules and DFT based quantum mechanical calculations with implicit solvation effects.

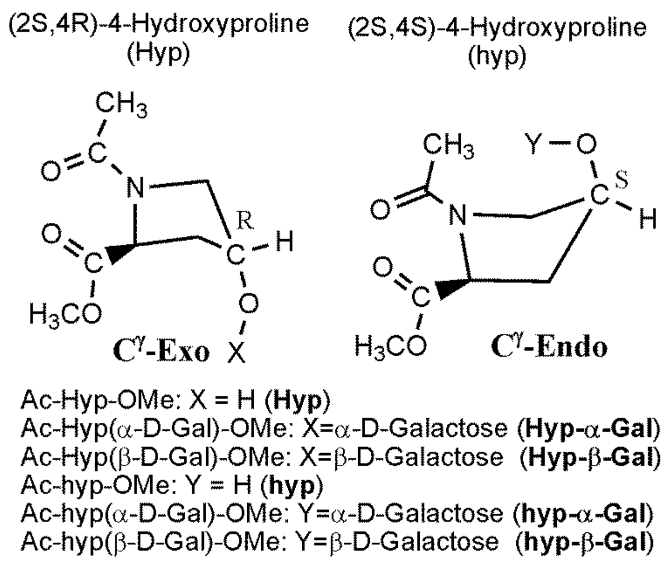

Figure 3.1. Schematic of the hydroxyproline compounds considered in the present chapter. The 4R and 4S stereoisomers at the C^{γ} carbon are considered together with their α and β galactosylated variants.

The calculations reveal details of the Hyp (hyp)-galactose glycosidic linkage that are important for understanding the structure of HRGPs, clarify the implications of glycosylation on the peptide backbone, and provide a possible explanation for the experimentally reported changes in the thermodynamics and kinetics of *cis*/*trans* isomerization of hyp (but not Hyp) upon glycosylation.

3.2. Computational Details

3.2.1. Molecular Dynamics Simulations

Initial structures of all compounds under investigation (Figure 3.1) were created with the XLeap module of the AMBER software suite.[14-15] The ff99SB AMBER force field parameter set[16] was used for the peptide, including an additional dihedral parameter developed by Park et al[17] to account for the "gauche effect" in Hyp due to the OH group and N atom attached to the C^{γ} and C^{δ} atoms, respectively. The GLYCAM06 (version 06c) parameter set[18] was used for the galactose sugar, including the glycosidic-linkage dihedral parameters. Additional RESP derived charges for the methoxyl (OMe) fragment were obtained using standard AMBER protocols.[19-20] To establish the glycosidic linkage between the sugar and peptide, the partial charge on the hydroxyl hydrogen was added to the linking oxygen and subsequently adjusted to yield a net neutral charge in accordance with standard GLYCAM procedure.[18] This approach led to a partial charge of –0.4223 which is similar to that previously used to describe other ester oxygens.[18]

Starting structures were solvated in a periodic octahedron of TIP3P water molecules. Once built, the models were subjected to two rounds of geometry optimizations. In the first round, the solute was held fixed by positional restraints, while the solvent molecules were allowed to relax in 1000 minimization steps, which consist of 500 steepest descent (SD) steps before switching to the conjugate gradient (CG) method. In the second round, the structure output from the first round was further minimized with the same number of SD and CG steps without restraining the solute. Next, the system was heated for 20 ps from a randomly assigned initial temperature of 5 K to the desired temperature of 300 K. During this calculation, the solute was fixed with weak positional restraints. The final result from this simulation was used as a starting point for subsequent calculations in the constant pressure (NPT) ensemble. For all model compounds, a minimum 200 ns MD simulation at 300 K and 1 atm was carried out with the first 5 ns discarded as equilibration. A time step of 1.0 fs was used in the heating equilibration, while constant pressure equilibration/production calculations utilized a 2.0 fs time step. Temperature was controlled using a Langevin thermostat with a 1.0 ps time constant. Long-ranged electrostatic interactions were treated with the particle mesh Ewald (PME) summation, while a cutoff of 8 Å was used for non-bonded interactions. Scale factors of 2.0 and 1.2 were applied to the electrostatic and 1-4 van der Waals interactions, for the AMBER force field[19] respectively. Although GLYCAM parameters are often also scaled using these same AMBER scale factors,[21] this approach is known to lead to improper sampling of some glycosidic dihedral angles.[12,18] and therefore the GLYCAM parameters were not scaled. The SHAKE algorithm[22] was implemented to restrain all bonds involving hydrogen. MD calculations were performed with the PMEMD module of AMBER (version 10).[23]

2.2.2. Density Functional Theory Calculations

Minimum structures obtained from MD simulations were further optimized with B3LYP/6-311++G(d,p) and frequency calculations at the same level of theory were conducted to ensure that minima were obtained. B3LYP was used since this functional has been extensively applied in the literature to study proline peptides and carbohydrates.[24-30] The ϕ_g ($\angle$(O5'–C1'–O^{δ}–C^{γ})) and ψ_g ($\angle$(C^{β}–C^{γ}–O^{δ}– C1')) glycosidic dihedral angles were used to examine the relative orientation of the sugar and peptide moieties (see Figure 3.2). To study the *cis-trans* isomerisation of the peptide backbone, potential energy surface (PES) scans were carried out using the improper dihedral ζ($\angle$(CH_3–O–C^{δ}–C^{α}), Figure 3.2) as the reaction coordinate as suggested by Fischer et al.[31] This dihedral is preferred to the peptide ω dihedral as it captures both *cis* and *trans* conformations of the peptide bond, as well as nitrogen pyramidalization, which is an important conformational feature of the *cis-trans* isomerisation reaction of proline and its derivatives.[32-33] During the scan, ζ in the B3LYP optimized trans isomer was fixed in 10° intervals from 0° to 180°, while the rest of the molecule was allowed to relax. Subsequently, unrestrained optimizations and frequency calculations of stationary points were performed at the same level of theory. All B3LYP calculations employed the IEF–PCM implicit solvation model[34] with ε = 78.39 to represent the water environment and the default parameters in the Gaussian 09 software.[35]

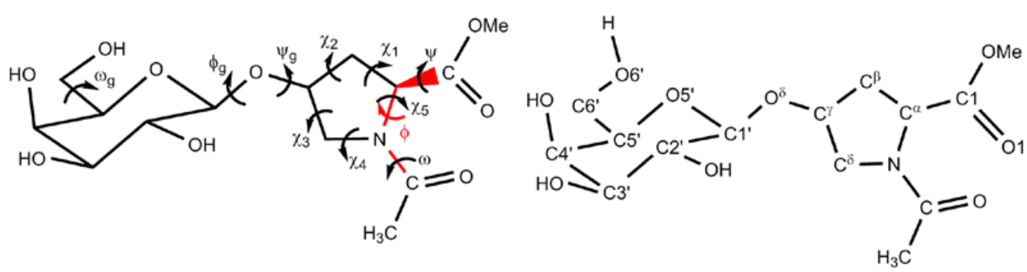

Figure 3.2. Chemical numbering and definitions of various dihedral angles in the galactosylated Hyp compounds.

3.2.3. Structural Analysis

The Westhof-Sundaralingam equations[36] were used to analyze the conformation of the five-membered proline ring according to:

$$P = \tan^{-1}\left(\frac{B}{A}\right) \text{ and } \chi_m = (A^2 + B^2)^{1/2}$$

with

$$A = \frac{2}{5}\sum_{i=1}^{5}\chi_i \cos\left(\frac{4\pi}{5}(i-2)\right) \text{ and } B = -\frac{2}{5}\sum_{i=1}^{5}\chi_i \sin\left(\frac{4\pi}{5}(i-2)\right)$$

where P (pseudorotation phase angle) defines the ring puckering according to the pseudorotation circle and χ_m (pseudorotation amplitude) is the maximum amplitude adopted by any endocyclic torsion angle (χ_i, Figure 3.2). Estimates of the conformational free energy differences between minima from the same MD simulation were obtained using:

$$\Delta G_{12} = -K_b \mathrm{T} \ln\left(\frac{\%\ \text{Population of minima 1}}{\%\ \text{Population of minima 2}}\right)$$

The presence of hydrogen bonds was analyzed based on a interatomic distance of less than 3.0 Å between the donor and acceptor, and a donor oxygen, donor hydrogen and acceptor bond angle greater than 120°.

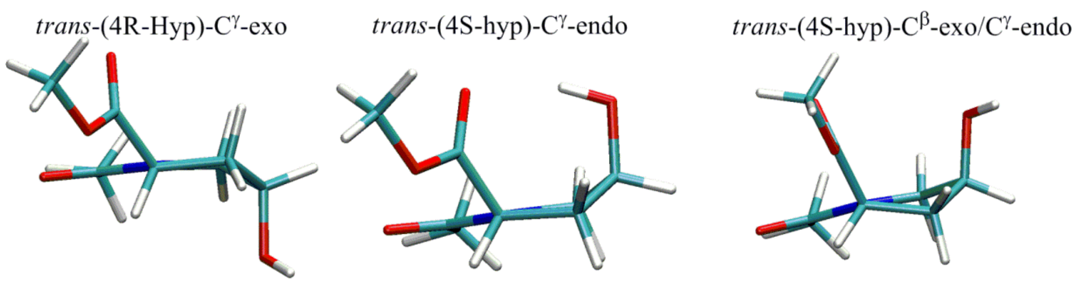

Figure 3.3. Preferred puckering states of (unglycosylated) 4–hydroxyproline compounds.

3.3. Results and Discussion

3.3.1. Proline Puckering

The relative orientation of the pyrrolidine and galactose rings is in part dictated by the direction that the glycan is projected, which is determined by the puckering preference of the hydroxyproline ring. Although different puckering states are predominantly populated by the Hyp and hyp proline derivatives,[37] experimental data has shown that the ring puckering is not substantially affected by glycosylation,[37] and therefore different proline–sugar orientations may prevail for the two stereoisomers. Table 3.1 displays the average and standard deviation (in brackets) of the pseudorotation parameters for minima of all compounds (discussed in detail below) obtained from MD simulations, which were calculated using all snapshots in the trajectory, as well as DFT (PCM-B3LYP/6–311++G(d,p)). In accordance with previous experimental and computational studies,[33,38] both types of calculations predict a C^{γ}–exo ($^{\gamma}E$) envelope-type puckering with a pseudorotation phase angle of P ≈ 10° for Hyp (Figure 3.3, top) and glycosylated Hyp. In contrast, some conformers of the hyp stereoisomer contain a hydrogen bond to the peptide backbone carbonyl oxygen (described in more detail below), which leads to C^{γ}–endo ($_{\gamma}E$) puckering (Figure 3.3, middle) with P ≈ 200° (minimum 2, Table 3.1). However, in explicit water, there is competition between the C-terminal carbonyl and water molecules for hydrogen bonding. Therefore, the average P values obtained from MD indicate that the proline ring predominantly adopts a twisted C^{γ}–endo/C^{β}–exo ($_{\gamma}T^{\beta}$) conformation with P ≈ 180° (Figure 3.3, bottom; minimum 1, Table 3.1). These variations in puckering preference of Hyp and hyp will project a covalently attached glycan in different directions relative to the plane of the pyrrolidine ring. Specifically, the $^{\gamma}E$ pucker directs C^{γ} below the plane defined by C^{δ}, N, C^{α} and C^{β} away from O1 (Figures 3.2 and 3.3), while the $_{\gamma}T^{\beta}$ and $_{\gamma}E$ puckers direct C^{γ} in the opposite direction.

3.3.2. Relative Orientation of Proline and Sugar Rings

In this section, we more carefully examine the arrangement of the sugar moiety with respect to the peptide. Since the sugar solely occupies the preferred $^{4}C_{1}$ conformation for the duration of the MD calculations, this conformation was considered in DFT optimizations.

Additionally, the ω_g (∠(H–O6'–C6'–C5'), Figure 3.2) dihedral angle was set to the most stable gt conformation obtained from MD and the other exocyclic dihedrals in an anticlockwise arrangement. While the general direction of the glycan projection is determined by ring puckering, the preferred values of the ϕ_g and ψ_g glycosidic linkage dihedral angles (Figure 3.3, left) control whether interactions occur between the backbone and sugar moiety. As illustrated by the scatter plots in Figure 3.4, in combination with the histograms in Figure 3.5, all glycosylated compounds occupy a single major minimum with respect to ϕ_g, while two major minima exist with respect to ψ_g. This is true for both the *trans* (black) and *cis* (red) conformations. Preliminary DFT scans of these two dihedral angles (Figure A1, Appendix A) also confirmed the presence of two minima for ψ_g. Furthermore, there is very good agreement in the values of the (ϕ_g and ψ_g) dihedrals predicted by molecular mechanics and DFT for all compounds (Table 3.1). Below, the main (ϕ_g,ψ_g) minima for each glycosylated compound will be individually discussed.

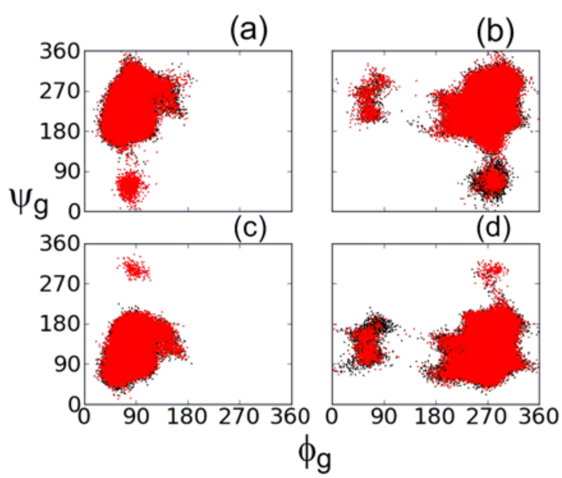

Figure 3.4. Distribution of glycosidic linkage ϕ_g and ψ_g dihedral angles (deg.) obtained from the MD trajectory of the *trans* (black) and *cis* (red) isomers of (a) Hyp–α–Gal, (b) Hyp–β–Gal, (c) hyp–α–Gal and (d) hyp–β–Gal.

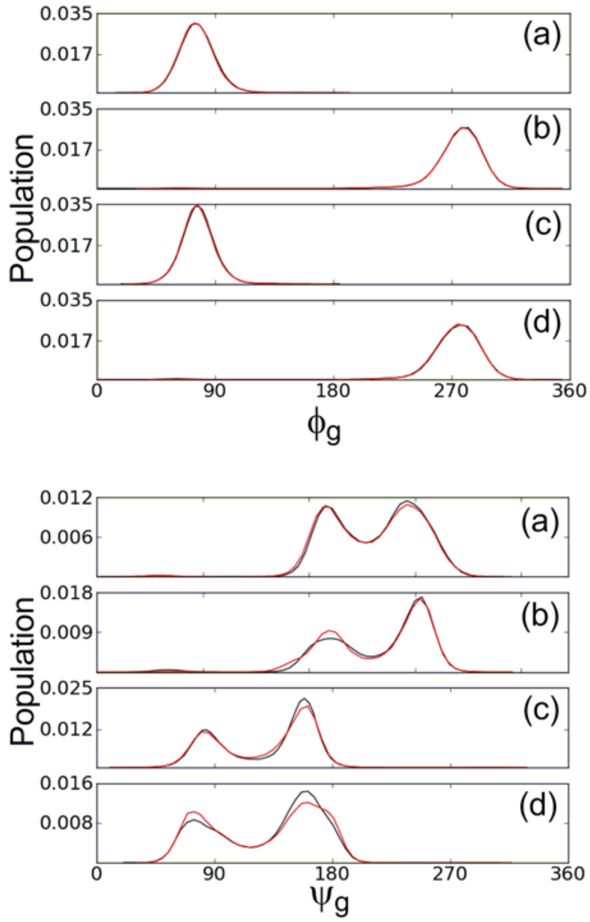

Figure 3.5. Histograms of the ϕ_g (top) and ψ_g (bottom) glycosidic linkage dihedral angles (deg.) obtained from the MD trajectory of the *trans* (black) and *cis* (red) isomers of (a) Hyp–α–Gal, (b) Hyp–β–Gal, (c) hyp–α–Gal and (d) hyp–β–Gal (d).

In minimum 1 of Hyp-α-Gal ((ϕ_g,ψ_g) = (81°,260°)), the sugar and Hyp rings are in a near stacked arrangement (Figure 3.6(a)). Moreover, the hydrophobic face of the sugar (containing three axial non-hydroxylated hydrogen atoms) interacts with the peptide, and is thus sheltered from water, which has been observed in crystal structures of glycans bound in the sugar binding domain of proteins such as Lysozyme.[39-40] In minimum 2 (74°,198°), the peptide and sugar rings are more distally arranged (Figure 3.6(b)). DFT free energy differences predict that minimum 1 is favoured over minimum 2 by approximately 3.4 kJ/mol. Similarly, the estimated free energy from MD simulations suggests minimum 1 is 0.9

kJ/mol more stable than minimum 2. The small free energy differences imply frequent conversion between minima, which can also be seen in the evolution of the ψ_g dihedral angle (Figure A2, Appendix A). In fact, several transitions take place in one nanosecond and hence occur on a timescale of a few 100 ps in all glycosylated peptides.

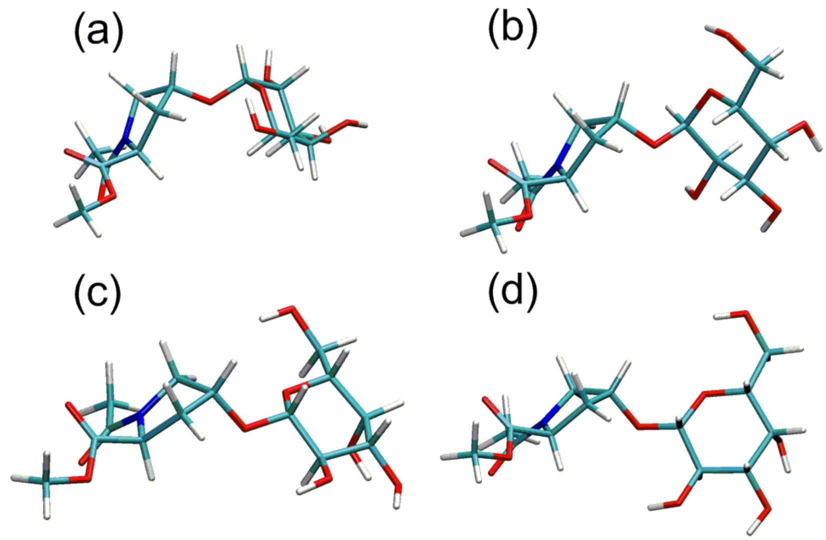

Figure 3.6. Minima 1 (a,c) and 2 (b,d) with respect to glycosidic linkage dihedral angles ϕ_g and ψ_g of Hyp–α–Gal (a,b) and Hyp–β–Gal (c,d) in the *trans* conformation.

In the two major minima of Hyp-β-Gal (Figures 3.7 (c) and (d)), the sugar and peptide rings are more separated than in the corresponding α-galactosylated peptide. MD free energy estimates indicate that minimum 2 (282°,274°) is favoured by approximately 1.3 kJ/mol, while DFT calculations indicate that minimum 1 (284°,212°) is preferred by 1.5 kJ/mol. Nevertheless, these free energy differences are very similar, meaning that both structures are nearly isoenergetic and MD simulations suggest the conversion barrier is small, as observed for the α-glycosylated variant described above (Figure A2, Appendix A).

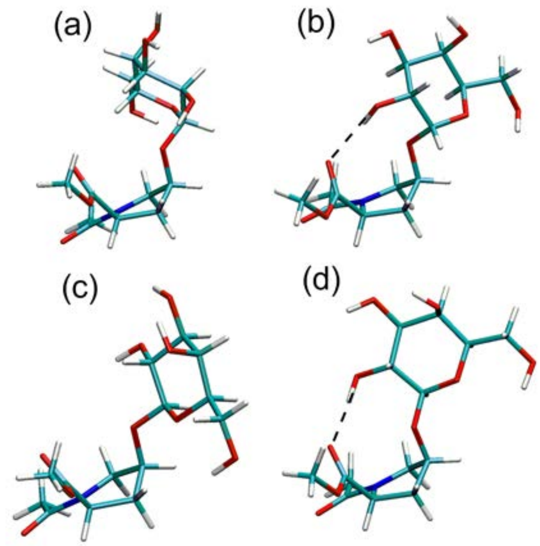

Figure 3.7. Minima 1 (a,c) and 2 (b,d) with respect to glycosidic linkage dihedral angles ϕ_g and ψ_g of hyp–α–Gal (a,b) and hyp–β–Gal (c,d) in the *trans* conformation. Important intramolecular hydrogen bonds shown with dashed lines.

As mentioned above, 4S stereochemistry in hyp leads to C^{γ}–endo/C^{β}–exo ($_{\gamma}T^{\beta}$) puckering, which allows the sugar moiety to be pointed in the opposite direction relative to Hyp. Although the rings are more spatially separated than in Hyp–α–Gal due the protrusion of the C-terminal region of the peptide towards the sugar (a consequence of the proline puckering), a stacked-like arrangement of the sugar and peptide moieties prevails in minimum 1 (83°,155°) of hyp–α–Gal (Figure 3.7(a)). In minimum 2 (75°,83°), an approximate 70° change in ψ_g brings the C2 hydroxyl of the sugar into hydrogen-bonding contact with the backbone C-terminal carbonyl of the peptide (Figure 3.7(b)). DFT calculations predict minimum 2 to be 5.5 kJ/mol more stable due to the intramolecular hydrogen bond, while MD favours minimum 1 by approximately 1.2 kJ/mol. The disparity between DFT and MD may arise due to the lack of explicit water molecules in the QM model or the role of dispersion interactions in the stacked orientation, which may not be fully recovered by B3LYP.[41] However, both types of calculations indicate that the energy

Table 3.1. Pseudorotation parameters and glycosidic dihedral angles (deg.) from MD simulations and PCM–B3LYP/6–311++G(d,p)[a]

Compound	Minimum 1[b]								Minimum 2[b,c]							
	B3LYP				MD[d]				B3LYP				MD[d]			
	P	χ_m	ϕ_g	ψ_g	P	χ_m	ϕ_g	ψ_g	P	χ_m	ϕ_g	ψ_g	P	χ_m	ϕ_g	ψ_g
Hyp	10	37			12(19)	34(7)										
Hyp-α-Gal	12	37	81	260	10(17)	35(7)	78(14)	267(19)	11	37	74	198	10(17)	35(7)	78(14)	202(15)
Hyp-β-Gal	10	37	284	212	12(16)	35(7)	280(20)	196(27)	12	37	282	274	12(16)	35(7)	280(20)	271(17)
hyp	173	34			178(18)	32(6)			192	37						
hyp-α-Gal	178	34	83	155	177(18)	33(7)	78(13)	156(12)	195	34	75	83	177(18)	33(7)	78(13)	88(14)
hyp-β-Gal	173	34	284	98	178(18)	33(7)	277(30)	85(16)	182	37	284	178	178(18)	33(7)	277(30)	162(16)

[a] See Figure 2.2 and the Computational Details for definition of parameters. [b] See Figures 2.6 and 2.7 for structures of minima 1 and 2. [c] For unglycosylated hyp, minimum 2 refers to the conformation with an intramolecular hydrogen bond between the C^{γ}-hydroxyl and the peptide backbone. In glycosylated hyp compounds, minimum 2 has an intramolecular hydrogen bond between the sugar and the peptide. [d] Numbers in brackets refer to the standard deviations.

difference between the two minima are small and therefore predict two nearly isoenergetic main minima (Figure A2, Appendix A).

As discussed for the 4R–stereoisomer, more extended conformations are observed upon β-glycosylation of hyp compared to the α-glycosylated compound. In minimum 1 (284°,98°), the glycan is directed away from the peptide (Figure 3.7(c)), while a hydrogen bond is formed between the sugar C2 hydroxyl group and the C-terminal carbonyl in the peptide backbone in minimum 2 (284°,178°), which creates a 10-membered hydrogen-bonding ring (Figure 3.7(d)). Such ring structures are frequently observed in β-turn structures where they stabilize the peptide backbone conformation.[42-44] These geometric differences lead to a DFT (MD) free energy difference of 0.39 (1.21) kJ/mol in favour of minimum 2, and MD suggests frequent transition between these structures (Figure A2, Appendix A).

The above results suggest that the relative arrangement of the peptide and sugar rings is jointly determined by the puckering of the proline ring and the glycosidic linkage dihedral angles. 4R glycosylation maintains C^{γ}-*exo* puckering on the proline ring, while 4S glycosylation primarily induces a twisted C^{γ}–*endo*/C^{β}–*exo* (${}_{\gamma}T^{\beta}$) ring pucker. These results are fully consistent with previously obtained NMR coupling constants.[11] Specifically, $J_{\alpha,\beta1}$ ($J_{\alpha,\beta2}$) values of 7.8 to 9.8 Hz (2.4 to 4.3 Hz) and 8.2 to 8.3 Hz (8.5 to 8.7 Hz) were obtained experimentally for the R and S compounds respectively which correspond to *exo* and *endo* puckers. The two puckering states project the glycan onto opposite sides of the plane formed by the proline ring and have different consequences for glycan–peptide interactions. In all compounds considered, the ϕ_g dihedral angle is very rigid ($\phi_g \approx 80°$ for α-linked peptides and $\phi_g \approx 280°$ for β-linked peptides, Figure 3.4), where the preferred value is mainly determined by the *exo*-anomeric effect. On the other hand, the ψ_g dihedral appears to be governed by the gauche effect with the η (∠(N–C^{δ}–C^{γ}–O^{δ})) dihedral angle adopting values that allow the C^{γ}–substituent to be in a gauche arrangement relative to nitrogen. Thus, by revealing the previously unknown conformation about the hydroxyproline-galactose linkage, this work determines that a combination of factors leads to the most populated proline–sugar conformations.

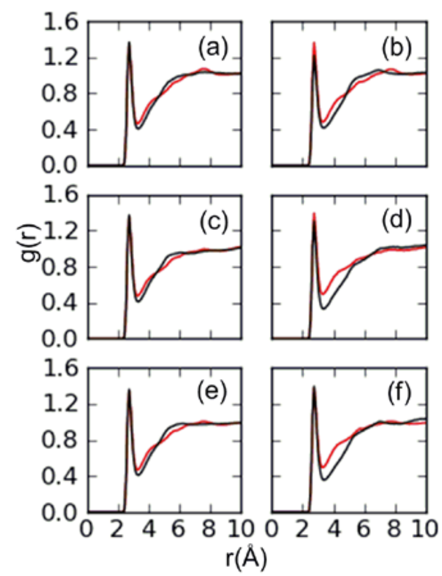

Figure 3.8. RDFs for interactions between a water oxygen (OW) and O (red) or O1 (black) in the peptide backbone of the trans conformation of (a) Hyp, (b) hyp, (c) Hyp–α–Gal, (d) hyp–α–Gal, (e) Hyp–β–Gal, and (f) hyp–β–Gal.

3.3.3. Solvation and Intramolecular Hydrogen Bonding

Carbohydrates are known to be highly hydrophilic, and therefore formation of glycopeptides can change the local structure of water molecules around the original peptide moiety. Additionally, the previous section showed that the differences in the relative position of the proline and sugar rings lead to intramolecular hydrogen–bonding interactions upon glycosylation of Hyp and hyp, which may affect solvation. These questions are explored in the present section by considering the radial distribution functions (RDFs), which describe how particle (water in this case) density changes around the group of interest, as well as permits a more close analysis of intramolecular hydrogen-bonding interactions. Since water can hydrogen bond to the N–terminal (O) and C–terminal (O1) carbonyls, as well as the O^{δ} in unglycosylated Hyp and hyp (Figure 3.8, right), the RDFs for interactions between these sites and a water oxygen (OW) will be discussed (Figures 3.8 – 3.9). The corresponding hydrogen atom (HW) RDFs, as well as all data for the *cis* conformations is provided in Appendix A (Figures A3 – A8).

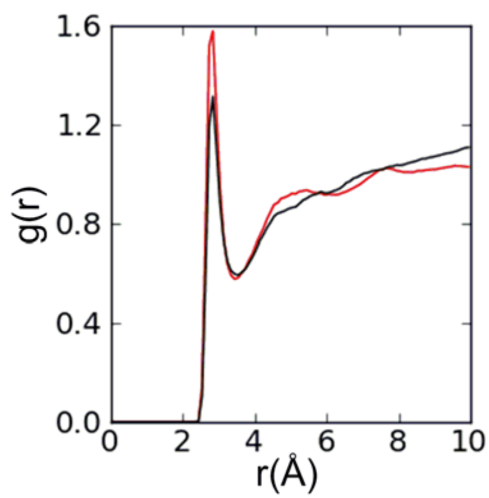

Figure 3.9. RDFs for interactions between a water oxygen (OW) and O^{δ} in the *trans* conformation of unglycosylated Hyp (red) and hyp (black).

For the unglycosylated and glycosylated versions of Hyp, the RDFs involving O (red) and O1 (black) have a peak at r = 2.75 Å (Figure 3.8, left), which confirms the presence of water hydrogen bonding to these sites. The similarity between the RDFs for these two sites for all Hyp derivatives considered here suggests that O and O1 are similarly hydrated, which is also evidenced by the potential of mean force (PMF) between O or O1 and OW (Figure A8,

Appendix A). In unglycosylated Hyp, the O^{δ} site also forms hydrogen-bonding contacts with water (Figure 3.9). No intramolecular hydrogen bond is detected in any of the Hyp compounds using our geometrical criteria (see Computational Details). Overall, the RDFs indicate that glycosylation does not noticeably alter interactions involving backbone carbonyl groups of Hyp.

In the case of unglycosylated and glycosylated hyp, there is a slightly higher peak in the O (red) than O1 (black) RDFs (Figure 3.8, right) compared to Hyp. This suggests that differences exist in the solvation and intramolecular hydrogen-bonding interactions of the N and C-terminal carbonyls, which is also reflected in the PMF plots (Figure A8, Appendix A). In the unglycosylated model, a hydrogen bond exists between O1 and the O^{δ} hydrogen (minimum 2; Figure 3.3, middle) for a significant amount of time (approximately 32% (24%) of the MD trajectory for the *trans* (*cis*) conformation) even though it is frequently interrupted by surrounding water molecules. This unique interaction in hyp is reflected in a significant reduction in the peak of the O^{δ}–OW RDF (Figure 3.9) relative to Hyp. Glycosylation of hyp eliminates the O^{δ}–hydroxyl hydrogen and therefore this hydrogen-bonding interaction. Instead, a hydrogen bond forms between O1 and the C2 hydroxyl group of the sugar. This interaction is present in minimum 2 ((ϕ_g,ψ_g) = (75°,83°)) of the α-glycosylated compound for approximately 12% of the trajectory for the *trans* (8% for *cis*) conformation, and 7% of the trajectory for the *trans* (10% for *cis*) conformation of minimum 2 ((ϕ_g,ψ_g) = (284°,178°)) of β-glycosylated hyp. Thus, hyp and its glycosylated variants exhibit intramolecular hydrogen bonding in solution between the 4S-substituent and the peptide backbone, which affects backbone hydration. This finding mirrors conclusions previously noted in the literature for hyp and related derivatives,[45] but contrasts the lack of intramolecular interactions in Hyp or its galactosylated variants reported in the present work. Additionally, the hydrogen bond can be mediated by a water molecule. Using a simple distance cut-off of 3.5 Å between heavy atoms (atoms other hydrogen), the bridged hydrogen bond exists for 16% (22%) of the simulation in the *trans* conformer of α(β)-glycosylated hyp. This is similar to results in the literature where water bridges have been observed in the MD simulations of model glycopeptides.[46-47] The stark contrast between the lack of intramolecular interaction between the peptide backbone and glycan in glycosylated Hyp compounds and the interactions found in α/β-glycosylated hyp derivatives may explain the experimentally observed difference in amide isomerization of these derivatives.

3.3.4. Peptide Backbone Conformation

Since glycosylation affects the solvation and intramolecular hydrogen-bonding interactions in hyp, it may also have consequences on the backbone structure; therefore, the conformation of the peptide backbone must be carefully examined. Figure 3.10 shows Ramachandran plots of the backbone ϕ and ψ dihedral angles obtained from MD simulations for the *trans* (black) and *cis* (red) conformations of all compounds. There are two main minima on the PES of the peptide in (ϕ,ψ) space, where the ϕ dihedral remains at approximately 300° for the duration of the simulation, while ψ adopts two values. Specifically, the most populated minimum is a polyproline II-like (PPII) conformation ((ϕ,ψ) ≈ (300°,130°)), while the second minimum is a right-handed α–helix-like (α_R) conformation ((ϕ,ψ) ≈ (300°,330°)).

Hyp exists in the PPII conformation for 96% of the simulation, while Hyp–α–Gal and Hyp–β–Gal adopt this conformation for 95% and 94%, respectively (Table 3.2). Furthermore, there is almost no difference in the average values of the backbone dihedral angles in all Hyp compounds, which is supported by DFT optimized structures (Table 3.3). Therefore, glycosylation of Hyp does not change the preferred conformation of the peptide backbone. This is not surprising since glycosylation does not lead to new intramolecular interactions as described in the previous subsection.

In contrast to Hyp, there are significant changes in the peptide backbone upon glycosylation of hyp (Tables 3.2 and 3.3). First, there is an increase in the α_R population, where the PPII population decreases from 94% in hyp to 83% (90%) in the α(β)-glycosylated version (Table 3.2). In DFT optimized α_R conformations (Figure A9, Appendix A), contacts exist between the peptide methoxyl group at C1 and the C2 hydroxyl group of the sugar in both glycosylated versions of hyp, which may provide further stabilization to the α_R structure for these derivatives compared to glycosylated Hyp.

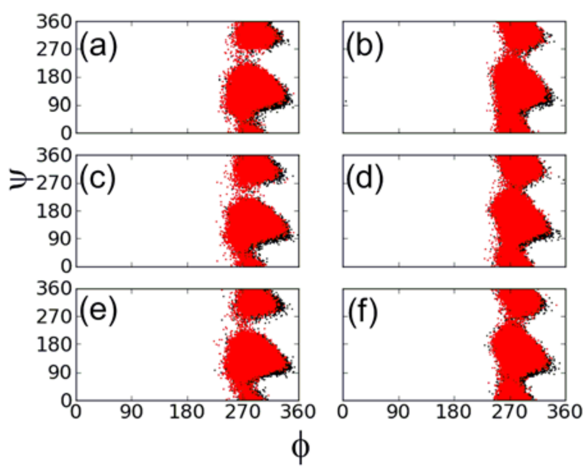

Figure 3.10. Distributions of peptide backbone ϕ and ψ dihedral angles (deg.) in the *trans* (black) and *cis* (red) conformations of (a) Hyp, (b) hyp, (c) Hyp–α–Gal, (d) Hyp–β–Gal, (e) hyp–α–Gal and (f) hyp–β–Gal.

Secondly, there are differences in the average values of important backbone dihedral angles in the PPII conformation of unglycosylated and glycosylated hyp. In particular, ψ from MD simulations averages 135° in hyp, but 140° and 146° in hyp–α–Gal and hyp–β–Gal, respectively. DFT results in this work (Table 3.3), as well as other studies,[38,45] the ψ dihedral angle is close to 180° reveal that in the absence of a hydrogen bond between the C^{γ}–OH group and the C-terminal backbone carbonyl oxygen in hyp (minimum 1; Figure 1, bottom). However, in the presence of this hydrogen bond (minimum 2; Figure 3.3, middle), which provides stabilization to the overall structure, ψ adopts a value of approximately 140°.

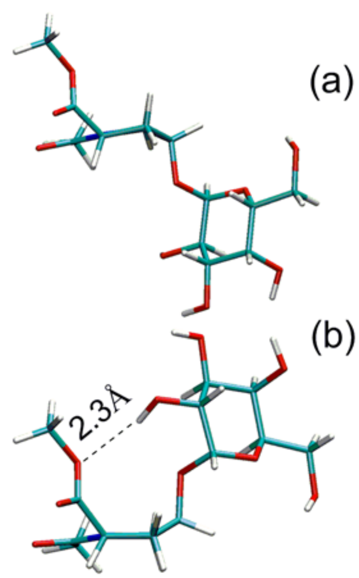

Figure 3.11. The α_R conformation of the ψ dihedral angle in (a) Hyp–α–Gal and (b) hyp–α–Gal, illustrating the weak contact between the sugar C2 hydroxyl group and the ester oxygen of the peptide in the hyp glycopeptide.

Upon glycosylation of hyp, both MD and DFT predict ψ values closer to 180° (minimum 2), while at the same time the structure maintains stability via intramolecular hydrogen bonding. In summary, glycosylation of Hyp does not affect the peptide backbone in accordance with experimental findings.[11,48] However, glycosylation of hyp allows the ψ dihedral angle to relax to larger values, and therefore, induces changes to the backbone of hyp. Since this feature, coupled with the intramolecular hydrogen-bonding interactions discussed previously is the main structural difference between glycosylated Hyp and hyp, this may be the reason for increased *trans* stabilization upon glycosylation of hyp. Hence, the details of the *cis*/*trans* isomerization will be considered in the next section.

Table 3.2. Conformations of the peptide backbone (ψ, deg.) and PPII occupation.

	MD[a]			B3LYP[b]			
				Minimum 1		Minimum 2	
Compound	% PPII	PPII	α_R	PPII	α_R	PPII	α_R
Hyp	96	131(20)	317(17)	147	323		
Hyp-α-Gal	95	130(20)	317(17)	147	323	147	323
Hyp-β-Gal	94	130(20)	316(17)	146	323	147	323
hyp	94	135(29)	328(20)	175	2	139	321
hyp-α-Gal	83	140(38)	331(18)	171	356	150	329
hyp-β-Gal	90	146(31)	331(19)	173	356	164	351

[a] Data shown are averages over the MD simulation with standard deviations provided in brackets.
[b] PCM–B3LYP/6–311++G(d,p).

3.3.5. Prolyl Amide Isomerization

The peptide backbone structure at various stationary points on the PES for *cis*/*trans* isomerization (Table 3.3) is first discussed. For unglycosylated and glycosylated Hyp compounds, ϕ is approximately 300° for the *trans* and *cis* isomers, and decreases by ~ 20° in the transition state (TS). The ψ dihedral angle approximately equals 147° in the *trans* conformation, and shifts to ~ 167° in the TS and 160° in the *cis* conformation. Although the ζ and ω dihedral angles remain at 0° in the *cis* and 180° in the *trans* isomer, ζ and ω become ~ 84° and 116° in the TS, respectively. An interesting known feature of *cis-trans* conversion in proline and its derivatives is the pyramidalization of the nitrogen atom, which leads to a tetrahedral structure in the TS.[33] This feature is well reproduced in our calculations, where ρ (the improper dihedral that measures this pyramidalization) decreases from ~ 355° in both the *trans* and *cis* conformations to 325° in the TS. This pyramidalization changes the pyrrolidine ring conformation (P changes from ~ 10° to 30° and χ_m increases from 37° to over 40°). Overall, these important geometric features are very similar in unglycosylated and glycosylated Hyp regardless of the conformation adopted by the glycosylated derivative, which suggests that glycosylation does not affect the structures of the stationary points involved in the Hyp isomerization reaction.

Figure 3.12 plots the PCM–B3LYP/6–311++G(d,p) energy as a function of the ζ ($\angle(CH_3–O–C^{\delta}–C^{\alpha})$, Figure 3.3) dihedral angle, while thermodynamic parameters for the *cis*–*trans* isomerization of Hyp compounds are given in Table 3.4 and the corresponding activation parameters are in Table A1. All calculated activation enthalpies are approximately 83 kJ/mol, which is in line with other theoretical[24,33] and NMR experimental[33] studies for Hyp. Similarly, computed activation free energies are approximately 84–86 kJ/mol in accordance with previous modeling studies of related compounds.[24] There are no significant differences in the thermodynamic and activation data for the unglycosylated and glycosylated Hyp compounds, which suggests that glycosylation does not affect isomerization. MD simulations indicate that both minima are occupied almost equally for both glycosylated Hyp compounds. Therefore, thermodynamic parameters can be estimated by averaging information from both minima. As an example, the calculated average ΔG° for Hyp-α-Gal is -3.69 kJ mol^{-1} compared to -4.52 ± 1.0 kJ mol^{-1} from experiments. Overall, there is reasonably good agreement between calculated and experimental data (Table 3.4) for Hyp and glycosylated Hyp, which indicates that the applied methodology adequately describes this isomerization process.

Table 3.3. PCM–B3LYP/6-311++G(d,p) structural parameters at stationary points of Hyp and hyp compounds

Compound	Minimum 1[a]								Minimum 2[a,b]							
	ϕ[c]	ψ[c]	ξ[d]	ω[c]	ρ[e]	P	χ_m	η[f]	ϕ[c]	ψ[c]	ξ[d]	ω[c]	ρ[e]	P	χ_m	η[f]
Hyp																
Trans	299	147	179	182	356	10	37	86								
TS	281	167	84	116	328	-33	41									
Cis	292	160	357	1	355	17	37									
Hyp-α-Gal																
Trans	298	147	179	182	356	12	37	83	298	147	179	183	356	11	37	85
TS	286	167	83	117	324	264	45		279	167	85	116	328	-28	41	
Cis	292	160	357	0	356	18	37		291	161	357	2	355	19	37	
Hyp-β-Gal																
Trans	299	146	179	183	356	10	37	83	298	147	179	183	357	12	37	85
TS	279	166	84	116	328	-31	41		281	166	84	116	327	-37	41	
Cis	292	161	357	1	4	18	37		292	161	357	1	356	18	37	
hyp																
Trans	282	175	179	183	359	173	34	266	294	139	183	183	1	192	37	272
TS	268	77	81	114	326	229	40		254	117	115	84		203	39	
Cis	278	184	359	1	357	165	36		281	139	1	5	1	180	35	
hyp-α-Gal																
Trans	283	171	179	183	355	178	34	270	294	150	180	183	356	195	34	273
TS	286	169	83	117	324	263	45		260	84	84	115	327	215	39	
Cis	279	178	359	2	356	169	36		282	166	0	4	355	178	36	
hyp-β-Gal																
Trans	282	173	180	183	356	173	34	266	286	164	180	183	357	182	37	278
TS	285	169	83	117	324	263	45		253	166	85	116	329	201	41	
Cis	278	182	359	2	357	166	36		280	173	0	3	356	174	38	

[a] See Figures 3.6 and 3.7 for structures of minima 1 and 2. [b] For unglycosylated hyp, minimum 2 refers to the conformation with an intramolecular hydrogen bond between the C^{γ}-hydroxyl and the peptide backbone. In glycosylated hyp compounds, minimum 2 has an intramolecular hydrogen bond between the sugar and the peptide. [c] Peptide backbone dihedral angles defined in Figure 3.2. [d] ξ is the $\angle(CH_3\text{–}O\text{–}C^{\delta}\text{–}C^{\alpha})$ improper dihedral angle (see Figure 3.2). [e] ρ is the $\angle(N\text{–}C\text{–}C^{\alpha}\text{–}C^{\delta})$ improper dihedral angle. [f] η is the $\angle(N\text{–}C^{\delta}\text{–}C^{\gamma}\text{–}O^{\delta})$ dihedral angle.

In the absence of a hydrogen bond to the peptide backbone in hyp (minimum 1), the φ dihedral angle assumes a value of approximately 282° in the *trans* isomer (Table 3.2). If a hydrogen bond to the backbone exists (minimum 2), then φ becomes 294° in hyp and hyp-α-Gal, and 286° in hyp-β-Gal. This shows that intramolecular hydrogen bonding affects the φ dihedral angle, where the preferred value becomes more similar to that in Hyp. In the *trans*, TS and *cis* conformations of 4S–hyp, ζ is approximately 180°, 84° and 0°, while ω is approximately 180°, 115° and 0°, respectively, in both minima. In minimum 1 for the *trans* isomer of all hyp compounds, ψ is 171–175°. However, there is a significant change in the ψ dihedral upon hydrogen-bond formation (minimum 2). Specifically, ψ decreases to 139° in hyp, 150° in hyp–α–Gal and 165° in hyp-β-Gal. Hence, hydrogen bonding to the peptide backbone in hyp has a significant impact on the structure of the peptide, which could affect isomerization upon glycosylation.

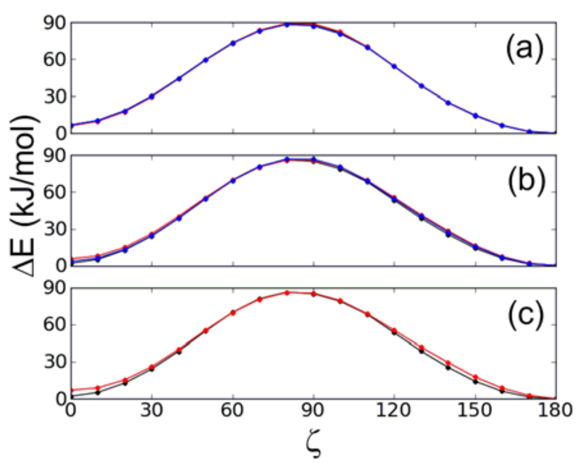

Figure 3.12. PCM–B3LYP/6-311++G(d,p) scans of the energy as a function of the ζ improper dihedral angle (deg.) describing *cis*-*trans* isomerization in (a) Hyp (black), Hyp–α–Gal (red) and Hyp–β–Gal (blue), (b) hyp (black), hyp–α–Gal (red) and hyp–β–Gal (blue), and (c) hyp with (red) and without (black) an intramolecular hydrogen bond.

Table 3.4. Calculated thermodynamic parameters for *cis-trans* isomerization in hyp and Hyp compounds in comparison to experimental data.[a]

Compound	Minimum 1[b]				Minimum 2[b]				Experiment[c]		
	ΔH°	ΔE	ΔS°	ΔG°	ΔH°	ΔE	ΔS°	ΔG°	ΔH°	ΔS°	ΔG°
Hyp	-6.20	-5.93	-5.61	-4.52					-5.98(0.2)	-4.89(0.5)	-4.52(0.39)
Hyp-α-Gal	-6.28	-6.11	-8.28	-3.82	-5.65	-5.40	-7.03	-3.56	-5.94(0.5)	-4.77(1.6)	-4.52(1.0)
Hyp-β-Gal	-5.50	-6.17	-6.35	-3.61	-5.87	-5.23	-12.55	-2.12	-6.06(0.5)	-5.77(1.7)	-4.35(1.0)
hyp	-2.03	-2.29	4.31	-3.31	-6.95	-6.90	-3.81	-5.82	-1.21(0.3)	3.31(1.1)	-2.22(0. 7)
hyp-α-Gal	-2.29	-2.16	-3.56	-1.22	-5.28	-5.06	-0.96	-4.99	-3.92(0.5)	-4.31(1.8)	-2.64(1.1)
hyp-β-Gal	-1.82	-2.01	5.31	-3.40	-3.31	-3.17	-1.63	-2.82	-3.81(0.2)	-4.02(0. 7)	-2.59(0.4)

[a]ΔH°, ΔE and ΔG° are in kJ/mol, while ΔS° is in J/molK. [b]PCM–B3LYP/6-311++G(d,p). [c]From Owens N. W. et al, *Chem. Eur. J.* **2009**, *15*, 10649-10657. The effects of temperature on $K_{trans/cis}$ was determined by fitting a straight line to $\ln(K_{trans/cis}) = (-\Delta H°/R)(1/T) + \Delta S°/R$ using 20°C to 72°C temperature range. Linear van't Hoff plots obtained indicated that the differences in enthalpy and entropy between the *trans* and *cis* isomers are independent of temperature. ΔH° and ΔS° were extrapolated from this plot.

The calculated free energy and enthalpy for the non-hydrogen–bonded conformation of hyp (minimum 1) are closer to the experimental values than those for the hydrogen–bonded conformation (minimum 2). Using the above mentioned distance and angle cut-offs, molecular dynamics calculations indicate that the hydrogen bond exists in unglycosylated hyp for only approximately one third of the simulation time in solution. In contrast, DFT calculations in implicit solvation suggest that the hydrogen–bonded conformation is slightly preferred, but the associated $\Delta S°$ has the opposite sign compared to experiment in this conformation. Thus, it is likely that the lack of explicit water molecules in DFT calculations is responsible for the stability preference of the hydrogen–bonded conformation. Therefore, although hyp can adopt both reported conformations, the non-hydrogen–bonded conformation likely prevails in solution, which is supported by similar $K_{trans/cis}$ measured when the C^{γ}-hydroxyl group in hyp is replaced by fluorine.[49] Despite the lack of complete agreement between the calculated and experimental thermodynamic parameters (Table 3.4), structural information suggests that the observed changes in the *trans* to *cis* ratio and isomerization rate upon glycosylation of hyp are likely due to changes in intramolecular hydrogen bonding. Specifically, the *cis*/*trans* isomerization of proline and derivatives is mostly governed by an $n \rightarrow \pi^*$ interaction between the N–terminal amide carbonyl O and C–terminal carbonyl C bond, which requires $\psi \approx 150°$.[50-52] Changes to ψ upon glycosylation due to changes in intramolecular hydrogen bonding could alter this interaction and thereby lead to a more favourable *trans* conformation. In Hyp compounds, glycosylation has no effect on the isomerization since the C^{γ}-exo puckering orients the sugar away from the peptide backbone, which prevents formation of intramolecular interactions that alter ψ and the $n \rightarrow \pi^*$ interaction. However, several intramolecular hydrogen-bonding interactions involving the peptide backbone occur in hyp and glycosylated compounds, and the discussion below explains how these affect the geometry about ψ, the n → π* interaction and the experimentally measured *cis*/*trans* ratios.

This work emphasizes that there are two main minima of unglycosylated hyp (Table 3.3, Figure 3.3), where the structure with a strong hydrogen bond between 4–OH and the peptide backbone is present for one-third of the MD calculation. Although the effect of the structure with the intramolecular hydrogen bond on the isomerization process in water is not completely understood, experimental studies suggest that it is of little consequence. Specifically, the $K_{trans/cis}$ measured for hyp (2.7) is similar to that for O-methylated hyp (2.0), which does not contain an intramolecular hydrogen bond to the backbone.[53] Similarly, another study reports a $K_{trans/cis}$ of 2.4 for hyp versus 2.5 for 4S-fluoroproline.[50] Furthermore, since the hydrogen-bonded conformation of unglycosylated hyp has a ψ dihedral angle (~140°) appropriate for the n → π* interaction, the dominance of this conformation should lead to an increased *trans* stability, which is in contrast to experimental observations. Thus, when the solvent competes for hydrogen-bonding interactions with 4–OH, the measured $K_{trans/cis}$ suggest that hyp preferentially adopts the non-hydrogen bonded conformation to maximize the solute–solvent interactions. We note that although the hydrogen-bonded conformation is likely not prevalent in water, it may become more important in non-hydrogen-bonding solvents. Indeed, $K_{trans/cis}$ of unglycosylated hyp increases to 5.0 in $CDCl_3$.[53]

The above discussion suggests that the non-hydrogen-bonded conformation of hyp (minimum 2) is dominant in water, and therefore the removal of the 4–OH to backbone hydrogen bond cannot explain the experimentally measured changes in the *cis*/*trans* ratio upon glycosylation. Instead, the creation of new hydrogen-bonding interactions upon glycosylation is responsible for the observed changes. Specifically, our calculations show that in the more prevalent non-hydrogen–bonded conformation of hyp, the ψ dihedral is approximately 180° and the favourable n → π* interaction is largely absent. For comparison, a similar conformation about ψ occurs in the DFT optimized minimum of O-methylated hyp (Figure A11), which cannot form an intramolecular hydrogen bond with the peptide backbone. However, upon α(β)–galactosylation of hyp, the ψ dihedral adopts a value of 150°(164°) in the hydrogen–bonded structure (minimum 2), which strengthens the n → π* interaction and leads to the observed increase in the population of the *trans* isomer. It should be noted that the less populated hydrogen-bonded conformation of hyp also adopts a ψ dihedral that affords a favorable n → π* interaction and therefore would not explain the observed increased stability of the *trans* conformation upon α(β)–galactosylation.

To provide further support for the above analysis, interaction energies were estimated with second-order perturbation theory using natural bond orbital (NBO) calculations. Models of the glycosylated compounds in implicit water, as well as in the gas phase with 10 explicit water molecules that represent the first few solvation shells about O1 (and include the bridging water, Figures A12 and A13, Appendix A), were considered. NBO calculations on the PCM–B3LYP/6–311++G(d,p) optimized structure of hyp–α–Gal indicate that the n → π* interaction stabilizes the *trans* conformation by approximately 3 kJ/mol. However, this favourable interaction is largely absent in the β-glycosylated model. When 10 water molecules were included in the B3LYP/6-31G(d) optimization, NBO calculations reveal that the *trans* conformation in the α(β)-glycosylated hyp is stabilized by approximately 12(4) kJ/mol. For comparison, similar calculations show that essentially zero stabilization is gained in the non-hydrogen-bonded conformation of unglycosylated hyp, as well as 4-O-methyl hyp, in the gas phase or implicit solvent. Therefore, these calculations support the new proposal outlined in this chapter that glycosylation turns on the n → π* interaction that is absent in the dominant conformation of unglycosylated hyp, and leads to the increased *trans* population observed in experiments.[11]

In summary, glycosylation does not significantly change the structure of Hyp and hence does not affect the amide isomerization in this proline derivative. This is further supported by the good agreement between calculated and experimental thermodynamic data. However, intramolecular hydrogen–bonding interactions occur between the backbone of hyp and glycan OH groups. These interactions change the structure of the peptide backbone, which likely leads to the experimentally observed differences in amide isomerization.

3.4. Conclusions

This chapter provides the first molecular level information about the nature of the hydroxyproline-galactose linkage that is prevalent in the hydroxyproline rich glycoproteins of plants, as well as explains the effects of glycosylation on amide isomerization of 4R(S)-hydroxyproline. Extensive MD calculations using the AMBER/GLYCAM force field in explicit solvent and complementary (implicit solvent) PCM-B3LYP/6–311++G(d,p) geometry optimizations were used to examine the conformation of the (glyco)peptides. Both methods show that all glycosylated compounds exhibit two main minima with respect to the glycosidic linkage. Since (unglycosylated) Hyp adopts C^{γ}-exo puckering, the sugar molecule is orientated away from the peptide backbone, and glycosylation has no apparent effect on *cis*–*trans* isomerization. This explains experimental studies that indicate there is no change in the peptide isomerization upon glycosylation. In contrast, a twisted C^{γ}-endo/C^{β}-exo ($_{\gamma}T^{\beta}$) puckering state is dominant in hyp, which brings the sugar and the proline rings in close proximity and leads to hydrogen–bonding interactions between the C2 hydroxyl group of the sugar and the peptide backbone C-terminal carbonyl, which is sometimes mediated by a water molecule. Furthermore, this hydrogen bond affects the structure of the peptide backbone and is likely responsible for the experimentally observed increase in the *trans* conformation population. Therefore, in addition to affording information about the effects of glycosylation on the Hyp backbone, an explanation for the experimentally measured increase in *trans* isomer population upon glycosylation of hyp is provided by the data presented in this Chapter. In addition, this Chapter provides evidence for a 10–membered intramolecular hydrogen-bonding ring between a sugar hydroxyl group and the peptide carbonyl group in hyp–α–Gal and hyp–β–Gal that affects N–terminal prolyl amide *cis*/*trans* isomerization. The preferred orientation of the sugar and proline rings determined in this Chapter indicate that intra (to peptide) and inter (to solvent) interactions could occur in oligopeptides that include the glycosylated Hyp residue. This is examined in the next Chapter with the study of nonapeptides of Pro, Hyp and glycosylated Hyp.

3.5. References[a]

1. Messner, P., Prokaryotic glycoproteins: unexplored but important. *J. Bacteriol.* **2004,** *186*, 2517-2519.

2. Eichler, J.; Adams, M. W. W., Posttranslational protein modification in Archaea. *Microbiol. Mol. Biol. Rev.* **2005,** *69*, 393-425.

3. Weerepana, E.; Imperiali, B., Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems *Glycobiology* **2006,** *16*, 91R-101R.

4. Specker, D.; Wittmann, V., Synthesis and application of glycopeptide and glycoprotein mimetics. *Top. Curr. Chem.* **2007,** *267*, 65-107.

5. Meyer, B.; Moeller, H., Conformation of glycopeptides and glycoproteins. *Top. Curr. Chem.* **2007,** *267*, 187-251.

[a] ACS referencing style was implemented throughout this thesis.

6. Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B., Effects of Glycosylation on Peptide Conformation: A Synergistic Experimental and Computational Study. *J. Am. Chem. Soc.* **2004,** *126*, 8421-8425.

7. O'Connor, S. E.; Imperiali, B., A molecular basis for glycosylation-induced conformational switching. *Chem. Biol.* **1998,** *5*, 427-437.

8. Pao, Y.-L.; Wormarld, M. R.; Dwek, R. A.; Lellouch, A. C., Effect of serine O-glycosylation on cis-trans proline isomerization. *Biochem. Biophys. Res. Commun.* **1996,** *219*, 157-162.

9. Liang, R.; Andreotti, A. H.; Kahne, D., Sensitivity of glycopeptide conformation to carbohydrate chain length. *J. Am. Chem. Soc.* **1995,** *117*, 10395-10396.

10. Owens, N. W.; Braun, C.; O'Neil, J. D.; Marat, K.; Schweizer, F., Effects of Glycosylation of (2S,4R)-4-Hydroxyproline on the Conformation, Kinetics, and Thermodynamics of Prolyl Amide Isomerization. *J. Am. Chem. Soc.* **2007,** *129*, 11670-11671.

11. Owens, N. W.; Lee, A.; Marat, K.; Schweizer, F., The Implications of (2S,4S)-Hydroxyproline 4-O-Glycosylation for Prolyl Amide Isomerization. *Chem.-Eur. J.* **2009,** *15*, 10649-10657.

12. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 10541-10545.

13. Teklebrhan, R. B.; Zhang, K. D.; Schreckenbach, G.; Schweizer, F.; Wetmore, S. D., Intramolecular Hydrogen Bond-Controlled Prolyl Amide Isomerization in Glucosyl 3(S)-Hydroxy-5-hydroxymethylproline Hybrids: A Computational Study. *J. Phys. Chem. B* **2010,** *114*, 11594-11602.

14. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, A Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules *Comput. Phys. Commun.* **1995,** *91*, 1-41.

15. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.

16. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65*, 712-725.

17. Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S., A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **2005,** *26*, 1612-1616.

18. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29*, 622–655.

19. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995,** *16*, 1357-1377.

20. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010,** *12*, 7821-7839.

21. Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J., Involvement of Water in Carbohydrate−Protein Binding. *J. Am. Chem. Soc.* **2001,** *123*, 12238-12247.

22. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

23. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo,; M. Crowley, R. C. W., W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I.; Kolossváry, K. F. W., F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke,; L. Yang, C. T., J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin,; Kollman, a. P. A. *AMBER*, 10; University of California, San Francisco: 2008.

24. Song, I. K.; Kang, Y. K., Conformational Preference and Cis-Trans Isomerization of 4(R)-Substituted Proline Residues. *J. Phys. Chem. B* **2006,** *110*, 1915-1927.

25. Momany, F. A.; Willett, J. L.; Schnupf, U., DFT molecular dynamics (DFTMD) simulations of carbohydrates: COSMO solvated alpha-maltose. *Theochem-J. Mol. Struct.* **2010,** *953*, 61-82.

26. Momany, F. A.; Schnupf, U., DFTMD studies of beta-cellobiose: conformational preference using implicit solvent. *Carb. Res.* **2011,** *346*, 619-630.

27. Momany, F. A.; Appell, M.; Willett, J. L.; Bosma, W. B., B3LYP/6-311++G** geometry-optimization study of pentahydrates of alpha- and beta-d-glucopyranose. *Carb. Res.* **2005,** *340*, 1638-1655.

28. Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L., B3LYP/6-311++G** study of monohydrates of alpha- and beta-D-glucopyranose: hydrogen bonding, stress energies, and effect of hydration on internal coordinates. *Carbohydr. Res.* **2004,** *339*, 553-567.

29. Cheeseman, J. R.; Shaik, M. S.; Popelier, P. L. A.; Blanch, E. W., Calculation of Raman Optical Activity Spectra of Methyl-β-d-Glucose Incorporating a Full Molecular Dynamics Simulation of Hydration Effects. *J. Am. Chem. Soc.* **2011,** *133*, 4991-4997.

30. French, A. D.; Johnson, G. P.; Kelterer, A.-M.; Dowd, M. K.; Cramer, C. J., Quantum Mechanics Studies of the Intrinsic Conformation of Trehalose. *J. Phys. Chem. A* **2002,** *106*, 4988-4997.

31. Fischer, S.; Dunbrack, R. L.; Karplus, M., Cis-Trans Imide Isomerization of the Proline Dipeptide. *J. Am. Chem. Soc.* **1994,** *116*, 11931–11937.

32. Mantz, Y. A.; Gerard, H.; Iftimie, R.; Martyna, G. J., Lsomerization of a peptidic fragment studied theoretically in vacuum and in explicit water solvent at finite temperature. *J. Am. Chem. Soc.* **2004,** *126*, 4080-4081.

33. Aliev, A. E.; Bhandal, S.; Courtier-Murias, D., Quantum Mechanical and NMR Studies of Ring Puckering and cis/trans-Rotameric Interconversion in Prolines and Hydroxyprolines. *J. Phys. Chem. A* **2009,** *113*, 10858-10865.

34. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105*, 2999-3094.

35. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, ., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. , Gaussian 09, Revision A.1, Gaussian, Inc.,: 2009.

36. Westhof, E.; Sundaralingam, M., A Method for the Analysis of Puckering DIsorder in 5-membered Rings - The Relative Mobilities of Furanose and Proline Rings and their Effects on Polynucleotide and Polypeptide Backbone Flexibility. *J. Am. Chem. Soc.* **1983,** *105*, 970-976.

37. Owens, N. W.; Lee, A.; Marat, K.; Schweizer, F., The Implications of (2S,4S)-Hydroxyproline 4-O-Glycosylation for Prolyl Amide Isomerization. *Chem.-Eur. J.* **2009,** *15*, 10649-10657.

38. Improta, R.; Benzi, C.; Barone, V., Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline,

hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* **2001,** *123*, 12568-12577.

39. Quiocho, F. A., Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu. Rev. Biochem.* **1986,** *55*, 287-315.

40. Vyas, N. K.; Vyas, M. N.; Quiocho, F. A., Sugar and signal-transducer binding sites of the Escherichia coli galactose chemoreceptor protein. *Science (Washington, D. C., 1883-)* **1988,** *242*, 1290-1295.

41. Cerny, J.; Hobza, P., Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.* **2007,** *9*, 5291-5303.

42. Richardson, J. S., The Anatomy and Taxonomy of Protein Structure. In *Adv. Protein Chem.*, C.B. Anfinsen, J. T. E.; Frederic, M. R., Eds. Academic Press: 1981; Vol. Volume 34, pp 167-339.

43. Wilmot, C. M.; Thornton, J. M., Analysis and prediction of the different types of [beta]-turn in proteins. *J. Mol. Biol.* **1988,** *203*, 221-232.

44. Rose, G. D.; Glerasch, L. M.; Smith, J. A., Turns in Peptides and Proteins. In *Adv. Protein Chem.*, C.B. Anfinsen, J. T. E.; Frederic, M. R., Eds. Academic Press: 1985; Vol. Volume 37, pp 1-109.

45. Kuemin, M.; Nagel, Y. A.; Schweizer, S.; Monnard, F. W.; Ochsenfeld, C.; Wennemers, H., Tuning the cis/trans Conformer Ratio of Xaa-Pro Amide Bonds by Intramolecular Hydrogen Bonds: The Effect on PPII Helix Stability. *Angew. Chem.-Int. Edit.* **2010,** *49*, 6324–6327.

46. Corzana, F.; Busto, J. H.; Engelsen, S. B.; Jimenez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Avenoza, A., Effect of beta-O-glucosylation on L-Ser and L-Thr diamides: A bias toward alpha-helical conformations. *Chem.-Eur. J.* **2006,** *12*, 7864-7871.

47. Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; Asensio, J. L.; Jimenez-Barbero, J.; Peregrina, J. M.; Avenoza, A., New insights into alpha-GalNAc-Ser motif: Influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **2006,** *128*, 14640-14648.

48. Owens, N. W.; Braun, C.; O'Neil, J. D.; Marat, K.; Schweizer, F., Effects of glycosylation of (2S,4R)-4-hydroxyproline on the conformation, kinetics, and thermodynamics of prolyl amide isomerization. *J. Am. Chem. Soc.* **2007,** *129*, 11670-11671.

49. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T., Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *J. Am. Chem. Soc.* **2001,** *123*, 777-778.

50. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T., Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **2001,** *123*, 777-778.

51. Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N., n ->pi* interactions in proteins. *Nat. Chem. Biol.* **2010,** *6*, 615–620.

52. Jakobsche, C. E.; Choudhary, A.; Miller, S. J.; Raines, R. T., n -> pi* Interaction and n)(pi Pauli Repulsion Are Antagonistic for Protein Stability. *J. Am. Chem. Soc.* **2010,** *132*, 6651–6653.

53. Shoulders, M. D.; Kotch, F. W.; Choudhary, A.; Guzei, I. A.; Raines, R. T., The Aberrance of the 4S Diastereomer of 4-Hydroxyproline. *J. Am. Chem. Soc.* **2010,** *132*, 10857-10865.

Chapter 4. Solvent Interactions Stabilize the Polyproline II Conformation of Glycosylated Oligoprolines[a]

4.1. Introduction

As discussed in Chapter 1, the PPII conformation has become recognised as an important secondary structural motif alongside alpha helices (e.g. the α_R-conformation with (φ, ψ) ≈ (–60°, –50°)) and β-strands with (φ, ψ) ≈ (–120°, 120°).[1] HRGPs contain a significant number of Pro (Hyp) residues and are naturally predisposed to adopt the PPII structure. Additionally, many Hyp residues in HRGPs undergo extensive glycosylation, which has been proposed in the literature to strengthen the PPII helix.[2-3] Therefore, studies that provide insights into the direct effects of Hyp glycosylation are necessary to clarify the carbohydrate–protein structure–function relationship in HRGPs.

Additionally, as outlined in Chapter 1, Schweizer and coworkers recently synthesised Ac–(Pro)$_9$–NH$_2$ ((Pro)$_9$), Ac–(Hyp)$_9$–NH$_2$ ((Hyp)$_9$)and Ac–[Hyp–(β–D–galactose)]$_9$–NH$_2$ ([Hyp–(β–Gal)]$_9$) model compounds (Figure 4.1) to obtain quantitative information about the effects of glycosylation on the PPII helix.[4] Circular dichroism (CD) was used to yield T_m estimates of 22°C, 38°C and 70°C, respectively, which indicates that glycosylation significantly increases the stability of the PPII conformation. Although the increased stability of the *trans* isomer due to hydroxylation explains the T_m difference between (Pro)$_9$ and (Hyp)$_9$, the reason for the increased stability of the PPII conformation in the glycosylated peptide is not clear from the experimental data.

---

[a] Reprinted with permission from E. B. Naziga, F. Schweizer and S. D. Wetmore. Solvent Interactions Stabilize the Polyproline II Conformation of Glycosylated Oligoprolines. *J. Phys. Chem. B*. **2013**, *117*, 2671-2681.

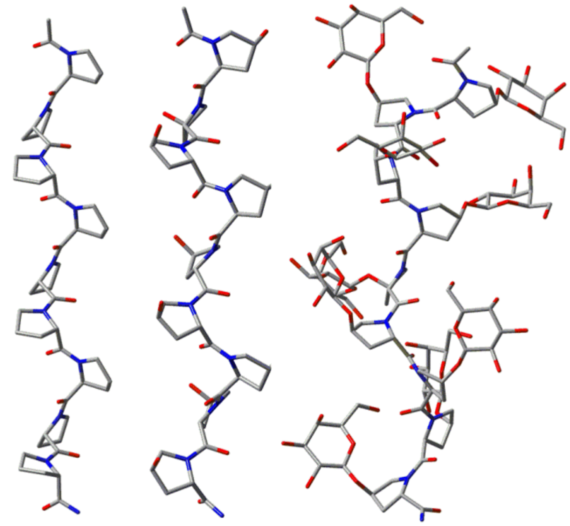

Figure 4.1. PPII structure of the Ac–(Pro)$_9$–NH_2, Ac–(Hyp)$_9$–NH_2 and Ac–[Hyp–(β–D–galactose)]$_9$–NH_2 peptides considered in this study.

Specifically, a positive band between 220 to 230 nm and a negative band between 195 and 210 nm in the CD spectra is a signature of the PPII structure,[5-6] with a change in the intensity of the corresponding minima usually being correlated with changes in the PPII content.[7] In the hydroxylated and glycosylated compounds, a decrease in the intensity of both bands was observed, which suggests a decrease in the PPII content despite the higher melting temperatures. Similar results were obtained by Horng and Raines for the (Flp)$_{10}$ and (Hyp)$_{10}$ peptides compared to (Pro)$_{10}$.[8] However, the calculated relative band strengths are 0.06, 0.21 and 0.29 for the (Pro)$_9$, (Hyp)$_9$ and [Hyp–(β–Gal)]$_9$ peptides, respectively, where a larger relative band strength correlates with a distortion of the PPII helix or a decrease in solvation.[5] Therefore, the loss in CD maxima intensity for [Hyp–(β–Gal)]$_9$ relative to (Hyp)$_9$ may not be due to loss of PPII character. Indeed, preliminary molecular mechanics minimizations in implicit solvent identified several inter–glycan or glycan–peptide hydrogen bonds in the glycosylated peptide, and these interactions could be responsible for the changes in the melting temperatures upon glycosylation.[4] However, other proposals have been put forth in the literature regarding possible modes of action of a covalently-linked sugar, including intramolecular interactions, sterics and direct or indirect solvent–sugar interactions.[9-12]

It is clear from the above discussion that more detailed information is needed about how glycosylation affects the PPII structure of oligoprolines and leads to the observed relative thermal stabilities of the unglycosylated and glycosylated oligomers. Therefore, this chapter uses a variety of sophisticated molecular modeling methods to provide additional insight into the structure of the galactosylated oligopeptide. Specifically, the three model peptides examined by Schweizer and coworkers (Figure 4.1) are considered using MD simulations. Several advanced techniques and extensive sampling procedures, such as REMD, ABMD, and HT–REMD, are used to obtain free energy estimates of the PPII to PPI transition. In contrast to previous proposals,[4] the data obtained in this Chapter reveals for the first time that interactions between the hydroxyl groups in hydroxylated or glycosylated compounds and solvent water molecules act in a complementary fashion with stereoelectronic effects to stabilize the PPII conformation in these substituted oligoproline peptides.

4.2. Computational Details

A thorough conformational analysis of the $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ peptides must investigate all available degrees of freedom. In particular, the puckering transitions of the Pro ring and variations in the backbone dihedral angles (φ, ψ and ω, Figure 1.4) should be considered. There are two major puckering states in Pro (C^{γ}–*exo* and C^{γ}–*endo*), which have been demonstrated by theoretical studies to be well sampled over the course of a standard MD simulation.[13-14] As mentioned in Chapter 1, the φ (∠(C–N–C^{α}–C), Figure 1.4) angle in Pro is restricted by the ring system to occupy a narrow band around –75°. While the ψ (∠(N–C^{α}–C–N), Figure 1.4) dihedral angle is more flexible, it can also be adequately sampled in MD simulations.[15] Thus, the main consideration when designing a sampling protocol for (substituted) oligoprolines is the ω (∠(C^{α}–C–N– C^{α}), Figure 1.4) dihedral angle, which dictates whether a particular residue adopts the *cis* or *trans* conformation. While the *cis* and *trans* structures of Pro are very close in energy, inter–conversion between these states requires more than 10 kcal/mol of energy.[16-18] Therefore, such conformational changes are not readily observed in a standard MD simulation of several hundred nanoseconds for these types of systems. In addition to the challenges associated with sampling ω, solvent effects must be carefully taken into consideration when studying oligoprolines. Unlike other secondary structural motifs, such as α-helices, the PPI and PPII structures of (unsubstituted) oligoprolines do not afford intramolecular hydrogen bonding. Therefore, interactions with solvent is essential for determining structural preferences.

With the above considerations in mind, a series of MD simulations that adequately sample all low vibrational modes of the oligopeptides were designed. All calculations were carried out using AMBER 10[19] or 11[20] with the AMBER FF99SB[21] and GLYCAM 06f[22] force fields describing the peptide and sugar moieties, respectively. Additional parameters for the hydroxyproline residue were obtained from Park *et al.*,[23] and the glycosylated Hyp units were modeled as described in Chapter 3. Implicit solvent calculations used a generalised Born model developed by Onufriev, Bashford and Case,[24-25] while explicit water molecules were represented by the TIP3P model.[26] Further details of each type of calculation implemented are given below.

4.2.1. MD simulations in explicit water

The conformations of all three peptides were sampled using standard MD simulations. Initial models (adopting an ideal PPI or PPII conformation) were created with the xLeap module of AMBER and solvated in a 48, 48, and 52 Å octahedral box of water for $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$, respectively. Two stages of minimization were carried out. In the first step, the solute molecule was held fixed, while the water molecules were relaxed. Next, the entire system was minimized. The resulting structures were slowly heated from 5 to 300 K over 50 ps and allowed to further equilibrate at the final temperature for another 50 ps in an NVT simulation. The equilibrated systems were then subjected to NPT production calculations of 101, 103 and 82 ns for $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$, respectively, in the PPII conformation. A 100 ns calculation was carried out for $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ in the PPI conformation. A timestep of 1.0 fs was used in the NVT equilibration and 2.0 fs in the NPT calculations. Van der Waals interactions were consistently truncated at 10 Å.

4.2.2. REMD simulations in explicit water

Replica exchange MD (REMD)[27] simulations were implemented for the $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ peptides starting from the final structures obtained from the regular MD simulations. A total of 36 replicas were implemented for $(Hyp)_9$ with the temperature ranging from 290.5 to 492.5 K, while 48 replicas and temperatures between 296.0 and 552.6 K were employed for $[Hyp–(β–Gal)]_9$. During these simulations, the ω dihedral angle was constrained to represent the *trans* conformation ($\omega = 180 \pm 30°$) by applying a 50.0 kcal mol^{-1} rad^{-1} restraint force. This was done to comprehensively explore the PPII conformation, as well as to identify any significant deviations from the PPII structure that do not involve *trans* to *cis* isomerization. Exchanges were attempted every 125 MD steps, while all other parameters were the same as discussed for the regular NPT MD simulation. However, REMD calculations were carried out in the NVT ensemble. A 50 and 100 ns simulation per replica was implemented, which led to a total simulation time of 1.8 and 4.8 μs for $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$, respectively.

4.2.3. ABMD simulations

To overcome the barriers associated with peptide *cis* to *trans* isomerization, ABMD[28-29] simulations were implemented. ABMD is effectively an umbrella sampling method that makes use of a time–dependent potential to bias the dynamics of the molecular system. This flattens the potential energy surface (PES) with respect to a selected collective variable that is related to the property under study. As in recent ABMD studies of oligoprolines,[30-32] the following CV was used, which is based on the cosines of the ω dihedral angle:

$$\Omega = \sum_i \cos(\omega_i)$$

Thus, Ω sums to +9 and −9 for PPI (all *cis* or ω = 0°) and PPII (all *trans* or ω = 180°) structures, respectively. In the present chapter, two variations of ABMD were implemented: (1) a replica exchange variant with each replica at a different temperature and biased by different potentials; and (2) a multiple walker strategy with all replicas at one temperature and biased by the same potential. Within this scheme, the following four simulation types were carried out:

a. **ABMD flooding in implicit water**: To obtain a rough estimate of the PMF, implicit solvation was used in the first instance since it is very computationally expensive to derive the biasing potential from explicit solvent calculations. However, once a PMF is developed in implicit solvent, refinement in explicit solvent becomes computationally feasible. Therefore, initially a short 12 ns simulation was conducted for $(Pro)_9$ in implicit water at 1200 K, which employed 24 replicas, a multiple walker approach, τ_F = 25 ps and $4\Delta\xi = 0.2$. A 1.0 fs MD timestep and infinite cut–off for the non–bonded interactions were implemented. Within this period, the collective variable explored most parts of the PES (i.e., Ω evolved from −9 to +9 several times). Using this estimate as input, more refined flooding calculations were performed with $4\Delta\xi$ reduced to 0.1 and τ_F increased in stages from 25 to 200 ps. At this stage, 24 replicas were implemented, each with a different biasing potential and temperatures ranging from 300 to 1200 K, where the wide temperature range (300–1200 K) will aid transition between the PPII and PPI structures. A total of 328 ns per replica simulation was used to refine the PMF for $(Pro)_9$. Using this refined potential, τ_F = 100 and 200 ps, a PMF for $(Hyp)_9$ was obtained from an additional 200 ns per replica simulation. Similarly, using PMFs from $(Hyp)_9$ and a 150 ns per replica simulation, the PMF for $[Hyp–(\beta–Gal)]_9$ was obtained.

b. **Umbrella sampling corrections to the PMF**: The PMFs obtained from the flooding stage described above were used in umbrella sampling calculations. This approach is based on the principle that if the PMF is very close to the exact value, then a flat histogram will be obtained in the space of the collective variable or, in other words, all values will be equally accessible. However, if there is a small difference (a few k_BT) between the calculated PMF and the true value, then the histogram obtained from umbrella sampling can serve as a correction. For each peptide under study, umbrella sampling simulations were conducted for at least 250 ns per replica and used to correct the PMFs. The same number of replicas and temperature range were employed as discussed for the flooding calculations.

c. **HT–REMD calculations**: To obtain equilibrium data from the ABMD calculations, an HT–REMD scheme was implemented using PMFs from the flooding stage. In these calculations, a total of 28 replicas (rather than the 24) were used. At 300 K, one of the 4 additional replicas is completely unbiased, while the other three additional replicas are biased based on the PMF of the 300 K replica and scaled by a range of factors (0.49, 0.76 or 0.90) that have been previously used and shown to work well for these types of systems.[33] This procedure enhances the exchange to the unbiased replica. All other simulation parameters were the same as for the flooding simulations. These calculations were run for at least 250 ns per replica for each peptide.

d. **PMF refinement in explicit water**: For the $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ peptides, calculations were conducted in explicit solvent to determine the discrete effects of water on the calculated PMFs. As done previously, PMFs obtained from the implicit solvent flooding stage were used as the starting point. Due to the large number of replicas required to cover the (300–1200 K) temperature range in an explicit solvent calculation, all 24 replicas were simulated at 300 K using a multiple walker approach. Initial structures of the solute in TIP3P water were taken from the explicit solvent REMD calculations. All other simulation parameters were the same as for REMD and ABMD, with the exception that the non–bonded cut–offs and timestep were reduced to 8.0 Å and 1.0 fs, respectively.

4.3. Results

4.3.1. Intermolecular/Intramolecular Interactions and Structural Information from MD Simulations in Explicit Water

Initially, equilibrium MD simulations were performed on the $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ peptides (Figure 4.1) in explicit water. This was done to observe the structural implications of contiguous glycosylation of Hyp as observed in HRGPs. Additionally, analysis of the structures from these equilibrium calculations could possibly provide clues regarding the significant stabilization of the PPII conformation as indicated by the high melting temperatures measured for $[Hyp–(\beta–Gal)]_9$ compared to the hydroxylated compound. In the analysis of these simulations, a particular focus is placed on the hydration of, as well as intramolecular hydrogen bonding within, the PPII structure. This choice was made since it is well known that interactions between solvent (water) molecules and the peptide backbone are important for maintaining the PPII conformation in unsubstituted polyproline helices since this secondary structure does not contain backbone–to–backbone intramolecular interactions.[34-35] Additionally, the preferential stability of PPII relative to PPI is solvent dependent for unsubstituted polyproline, with PPII being the most stable structure in water and PPI favoured in solvents such as methanol and 1-propanol.[36-38] Nevertheless, glycosylation may change solvent accessibility, as well as introduce new intramolecular hydrogen bonding to the peptide backbone.

To gain insight into backbone hydration, the first shell of water molecules around the peptides was analysed to determine the number of solvent molecules in close proximity and thereby able to form water–backbone interactions. Using a cut–off of 3.4 Å, which coincides with the first minimum in the water–backbone C=O radial distribution function, the number of water molecules in the vicinity of the backbone carbonyl groups was counted. The resulting averages over the duration of the MD simulations for the peptides in the PPII conformation are presented in Table 4.1 for each residue (1 to 9) of the oligoproline peptides with the capping groups ignored. The number of water molecules solvating the carbonyl backbone of $(Pro)_9$ and $(Hyp)_9$ are nearly identical, which indicates that the backbone solvation is not affected by hydroxylation. On average approximately 2.5 water molecules solvate each carbonyl group for the $(Pro)_9$ and $(Hyp)_9$ compounds. This result is reasonable since the OH groups in $(Hyp)_9$ are not sterically large enough to reduce water access to the backbone. Moreover, the OH groups are also involved in hydrogen-bonding interactions with water molecules. However, there is an overall slight (0.5 – 1) decrease in solvation of the peptide backbone upon complete glycosylation of $(Hyp)_9$. This observation is not surprising since the bulky sugar groups can effectively shield the peptide backbone from water. Additionally, as discussed in Chapter 3, there is fast conversion between the two minima of the glycosidic linkage of Hyp, which creates an effective cover for the backbone, reducing solvent access. Furthermore, these results are consistent with a decrease in the CD maxima for the glycosylated peptide compared to $(Hyp)_9$.[4] Indeed, this observation was in part attributed to a change in the solvation environment of the peptide backbone,[4] which is confirmed for the first time in this thesis.

The backbone solvation of the PPI conformation was also examined in both the hydroxylated and glycosylated compounds to determine if differences in the experimentally-observed thermal stability[4] can be attributed to variations in the solvation of PPI and PPII. Table 4.2 compares the same solvent accessibility data from the PPI and PPII conformations in the $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ compounds. On average fewer (Table 4.3) water molecules interact with the backbone in the PPI conformation of both peptides.

This is in line with experimental[36,38] and computational studies,[30] which show that the PPI conformation is not favoured in water due to loss of backbone-water hydrogen-bonding interactions. This finding also correlates with the more compact structure of the PP1 conformation (Figure 1.5).

Table 4.1. Average number of water molecules solvating the backbone carbonyl groups in the $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ peptides in the PPII conformation according to MD simulations in explicit water.[a,b]

Residue (*i*)	1	2	3	4	5	6	7	8	9
$(Pro)_9$	2.73	2.52	2.51	2.50	2.50	2.51	2.49	2.53	3.31
$(Hyp)_9$	2.72	2.45	2.47	2.47	2.47	2.48	2.46	2.52	3.35
$[Hyp–(\beta–Gal)]_9$	1.99	1.76	1.81	1.74	1.62	1.48	2.07	2.60	3.41

[a]Solvation was determined to occur if the water molecule was within 3.4 Å of the backbone, [b]A 0.002 standard error of the mean was obtained in all cases.

Table 4.2. Comparison of the average number of water molecules that solvate the backbone carbonyl groups in the PPII and PPI conformations of $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ peptides according to MD simulations in explicit water.[a,b]

Residue (*i*)		1	2	3	4	5	6	7	8	9
$(Hyp)_9$	PPII	2.72	2.45	2.47	2.47	2.47	2.48	2.46	2.52	3.35
	PPI	2.62	1.99	1.67	1.46	1.44	1.46	1.45	1.42	2.04
$[Hyp–(\beta–Gal)]_9$	PPII	1.99	1.76	1.81	1.74	1.62	1.48	2.07	2.60	3.41
	PPI	1.84	1.66	1.71	0.13	0.58	1.20	0.29	0.82	1.80

[a]Solvation was determined to occur if the water molecule was within 3.4 Å of the backbone, [b]A 0.002 standard error of the mean was obtained in all cases.

As expected, no intramolecular interactions were found in the $(Pro)_9$ and $(Hyp)_9$ peptides in the PPII conformation. However, several peptide–sugar and sugar–sugar interactions are present in $[Hyp–(\beta–Gal)]_9$. Table B1 (Appendix B) provides details (atoms involved and percentage occupancies) of these intramolecular hydrogen-bonding interactions. Overall, these contacts occur for a small portion of the MD calculation, with a maximum occupancy of approximately 13%, since there is competition between the backbone carbonyls and water for interactions with the sugar hydroxyl groups (> 84%, Table B2). The sugar–backbone hydrogen-bonding interactions that occur for at least 5% of the total simulation time include interactions between the O6'-hydroxyl hydrogen (Figure 4.2) at position *i* and the carbonyl oxygen (C=O) in the peptide backbone at positions *i*–2 and *i*−3. Additionally, hydrogen bonding occurs between the O2' hydrogen of residue *i* and O6' of residue *i*+1. Despite the large number of such interactions, their low occupancies indicate that intramolecular hydrogen bonding cannot explain the increased thermal stability of $[Hyp–(\beta–Gal)]_9$ relative to $(Pro)_9$ and $(Hyp)_9$.

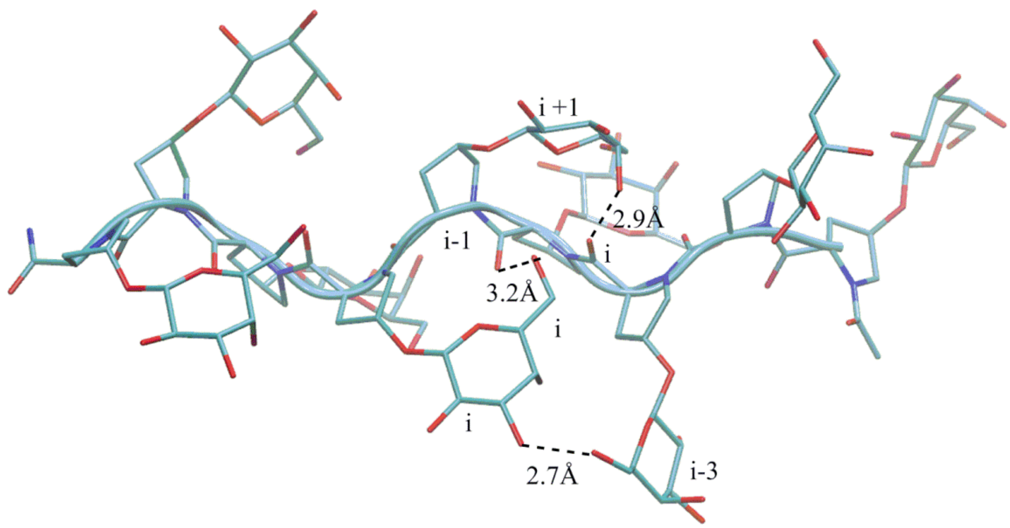

Figure 4.2. A snapshot from explicit solvent MD simulations of the $Ac–[Hyp–(\beta–D–galactose)]_9–NH_2$ peptide showing observed intramolecular hydrogen bonding interactions.

Compared with the PPII conformation, intramolecular hydrogen bonds are observed in the PPI conformation with a higher frequency (Table B3). Indeed, several peptide–sugar and sugar–sugar interactions exist, with 6 of these having occupancies over 50%. The peptide–sugar interactions primarily occur between the O6'-hydroxyl hydrogen at position *i* and the carbonyl oxygen (C=O) in the peptide backbone at position *i*−1. Sugar–sugar interactions mainly take place between the O2' (O4') hydrogen of residue *i* and O6' of residue *i*+3 (*i*+3). Therefore, while solvent–backbone hydrogen bonding is greatly reduced in the PPI conformation of the $[Hyp–(\beta–Gal)]_9$ peptide, strong intramolecular interactions appear.

The backbone (φ,ψ) dihedral angles for the calculations carried out in the PPII conformation adopt average values of (–63°, 155°), (–61°, 153°) and (–65°, 160°) for the $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ peptides, respectively. Furthermore, all ω dihedral angles in the three peptides remain close to 180°. This indicates that all three peptides remain in the PPII conformation for the duration of the calculation as expected based on previous literature.[16,18,30] While the length of these simulations is sufficient to obtain equilibrium information about the PPII conformation, it may not be sufficient to overcome barriers to other possible low energy states on the PES. For instance, other peptides that significantly populate the PPII conformation can also adopt β–strand conformations.[39-46] In addition, it would be interesting to determine whether structures that significantly deviate from the PPII conformation can be adopted without invoking *trans* to *cis* isomerization. This will clarify proposals in the literature that *cis*/*trans* isomerization is the main source of conformational heterogeneity in oligoproline peptides.[47-49] To address these questions, REMD simulations on the $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ peptides will be discussed in the next section. The $(Pro)_9$ peptide was not considered with REMD since the data presented in this section suggests that this peptide and $(Hyp)_9$ have similar solvation properties. Additionally, the most likely explanation for the observed increased PPII stability of $(Hyp)_9$ relative to $(Pro)_9$ is increased *trans* stabilization due to the addition of the electronegative OH groups.[50]

4.3.2. Structural Information from REMD Simulations in Explicit Water

To comprehensively explore the PPII conformation, as well as to identify any significant deviations from the PPII structure that do not involve a *trans* to *cis* isomerization, a restraint force was applied in the REMD simulations in explicit water to restrain the ω dihedral angle to represent the *trans* conformation.

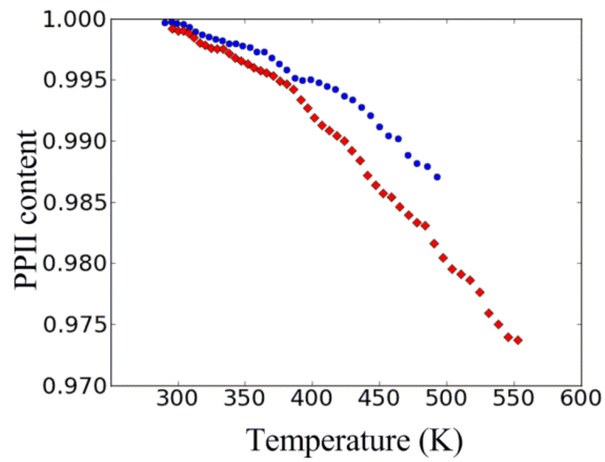

Figure 4.3. Change in PPII content as a function of temperature according to REMD simulations of the Ac–$(Hyp)_9$–NH_2 (circles) and Ac–[Hyp–(β–D–Gal)]$_9$–NH_2 (diamonds) peptides in explicit water.

Table 4.3. Comparison of the average number of water molecules that solvate the backbone carbonyl groups of the $(Hyp)_9$ and [Hyp–(β–Gal)]$_9$ peptides at different temperatures according to REMD simulations in the PPII conformation.[a,b]

Residue (*i*)		1	2	3	4	5	6	7	8	9
$(Hyp)_9$	299.4K	2.72	2.51	2.45	2.48	2.49	2.49	2.47	2.53	3.35
	492.5K	2.79	2.57	2.55	2.58	2.55	2.56	2.54	2.67	3.72
[Hyp–(β–Gal)]$_9$	300.0K	1.98	1.68	1.70	1.69	1.64	1.54	2.06	2.58	3.40
	552.6K	2.29	1.99	1.97	2.00	1.95	1.99	2.12	2.72	3.81

[a]Solvation was determined to occur if the water molecule was within 3.4 Å of the backbone, [b]A 0.002 standard error of the mean was obtained in all cases.

Following the analysis carried out in the previous subsection, Table 4.3 displays the average number of water molecules within 3.4 Å of the peptide backbone carbonyls in the $(Hyp)_9$ and [Hyp–(β–Gal)]$_9$ peptides at room temperature and the highest temperature simulated. Solvation of the backbone carbonyl groups does not significantly deviate with temperature for either peptide. However, the intramolecular hydrogen-bond occupancy decreases with increased temperature for the [Hyp–(β–Gal)]$_9$ peptide (Table B1). Indeed, at the highest temperature considered (552.6 K), all but one peptide–sugar and sugar–sugar intramolecular hydrogen bond is lost in the glycosylated peptide.

The PPII content versus temperature is displayed in Figure 4.3, where a given residue was considered to adopt a PPII conformation if (–110° < φ < –20°), (50° < ψ < 180°) or (–120° < ψ < –180°) and the ω dihedral angle adopts a *trans* configuration. The data indicates that even at very high temperatures, the φ and ψ dihedral angles remain in the PPII region of the Ramachandran plot in both peptides when ω is constrained to adopt the *trans* conformation. Analysis of the data reveals a mostly PPII family of structures with a root mean square deviation (RMSD) of ≈ 1–2 Å from an idealised PPII and end-to-end distances of approximately 22–28 Å (Figure 4.4). However, there are a few excursions from an idealised PPII structure to more globular-like structures (RMSD ≈ 3–4 Å). These globular structures arise in both peptides when the backbone dihedral angle ψ ≈ –50° in the residues in the center region of the peptide (*i* = 4–5), which defines the α_R-configuration. The number of such transitions increases with temperature (Figure 4.4), but still corresponds to a very small portion of the total trajectory. That very few excursions to the α_R conformation occur in these peptides is in agreement with the results obtained in Chapter 3. Specifically, the monomeric Hyp compound and its glycosylated variants were found to occupy the PPII conformation for 94–95% of the MD simulation initiated with the ω backbone dihedral angle in the *trans* conformation.

The above homogenous results indicate that *cis*–*trans* isomerization is likely the most important factor in the observed conformational heterogeneity of oligoprolines and derivatives. Specifically, when the PPII conformation melts, *cis* conformations must be adopted. Additionally, there is no evidence of β-sheet or random coil structures, even at the high temperatures simulated. To determine whether the conformational ensemble changes when the restraint on the ω dihedral angle is removed and gain free energy estimates for conversion between the PPI and PPII conformations, ABMD calculations were conducted as described in the following sections.

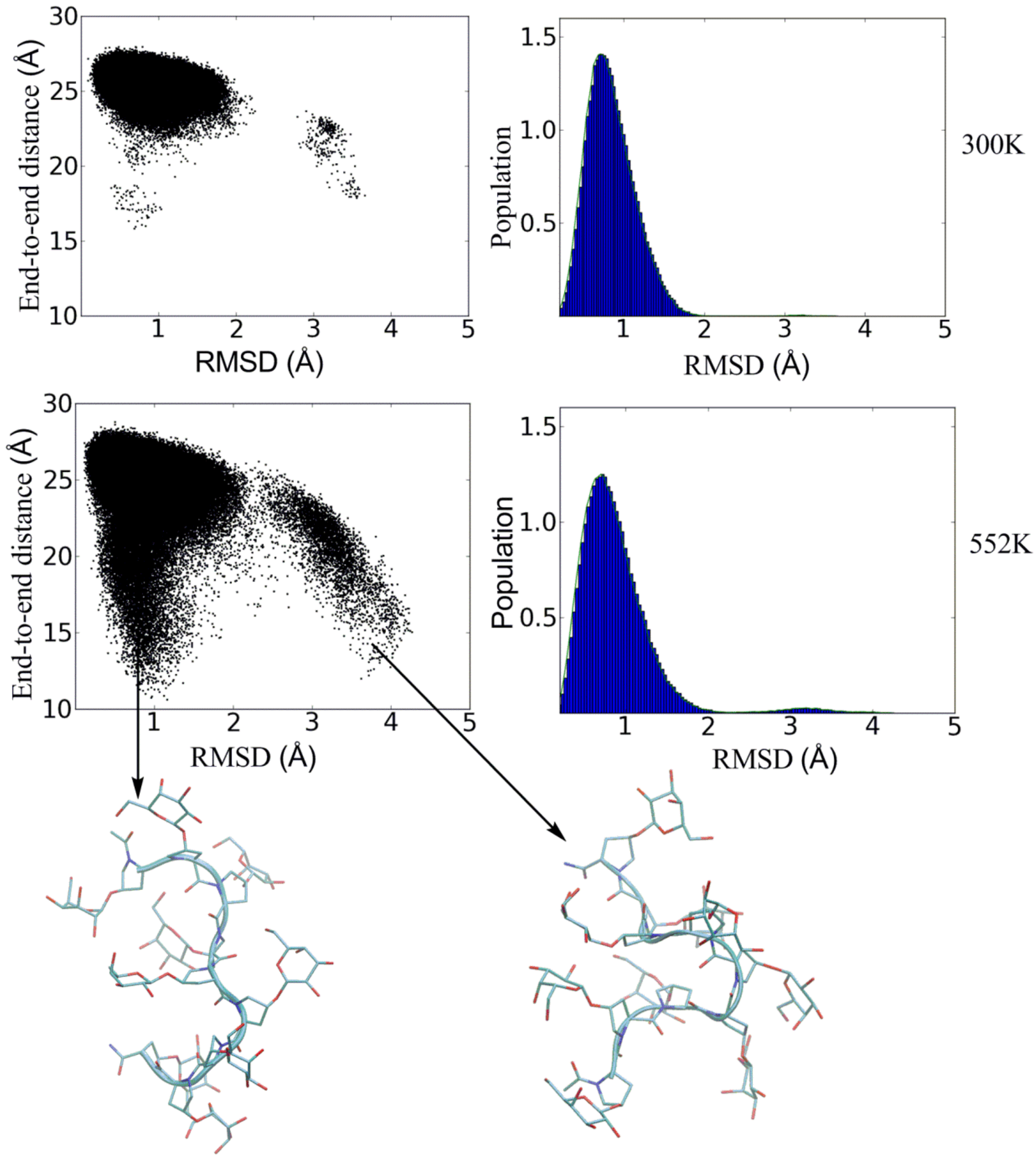

Figure 4.4. End-to-end distance (residue 2 N to residue 10 C) versus root mean squared deviation (RMSd) from an ideal PPII structure (left) and RMSD histograms (right) obtained from explicit solvent REMD calculations on the Ac–[Hyp–(β–D–Gal)]$_9$–NH$_2$ peptide at 300 K (top) and 552 K (bottom). Representative structures are shown for the conformations with the large deviations from an ideal PPII helix.

4.3.3. ABMD Simulations in Implicit Water

ABMD simulations were initially carried out in implicit water to investigate the influence of *cis*–*trans* isomerization on the structural ensemble of all three peptides (Figure 4.1) since this transition does not occur on feasible timescales during regular dynamics. As mentioned in the Computational Details, implicit solvation was used in the first instance since it is very computationally expensive to derive the biasing potential from explicit solvent calculations. Furthermore, a wide temperature range (300–1200 K) was implemented in the implicit solvent calculations to aid transition between the PPII and PPI structures. This combination of temperature and umbrella (ABMD) sampling has been previously shown to adequately characterize the conformational dynamics of polyproline peptides.[30,32] Initial calculations were performed on $(Pro)_9$ for comparison to previous studies, as well as to obtain a PMF that can be extended to the $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ models.

From the PMFs (Figure 4.5), the free energy for transitioning from PPII to PPI (ΔF(PPII–PPI)) is estimated to be –20.5 kJ mol^{-1}, –37.6 kJ mol^{-1} and –37.6 kJ mol^{-1} for $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$, respectively. The more negative ΔF(PPII–PPI) for $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$ relative to $(Pro)_9$ is expected since these PTMs have been shown to stabilize the *trans* conformation.[51] However, a higher value for ΔF(PPII–PPI) is expected for $[Hyp–(\beta–Gal)]_9$ based on T_m data. This lack of agreement in the relative values for ΔF(PPII–PPI) is further explored below. The resulting PMFs were subsequently used in HT–REMD simulations, which included an unbiased replica to obtain equilibrium data. Results from a residue–based analysis of this equilibrium data are shown Table 4.4, which include the percentage of conformations that adopt the true PPII structure (P(PPII)), *trans* ω dihedral angle (P(T)), as well as lie in the F region of the Ramachandran plot (P(F)), which encompasses both PPI and PPII structures.

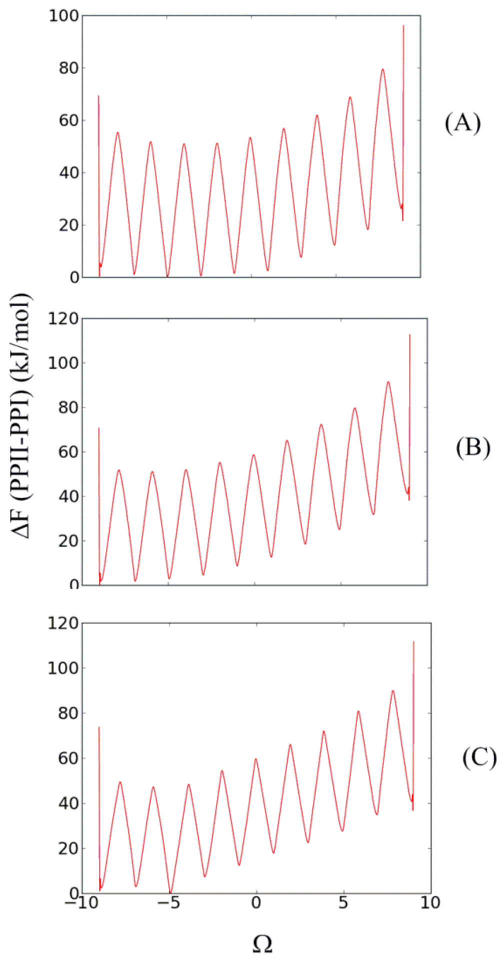

Figure 4.5. Potential of mean force (PMF) for conversion from PPII (Ω = –9.0) to PPI (Ω = +9.0) for the (A) Ac–$(Pro)_9$–NH_2, (B) Ac–$(Hyp)_9$–NH_2 and (C) Ac–[Hyp–(β–D–Gal)]$_9$–NH_2 peptides obtained from implicit solvent ABMD simulations.

The unbiased HT–REMD data indicates a high PPII content in the $(Pro)_9$ peptide, with P(PPII) ranging between 56–86% and on average being 65%. However, P(T) is on average 69%, which implies there is a significant amount of the *cis* conformation about the ω dihedral (31%). Furthermore, there is a decrease in the percentage occupation of the *trans* isomer towards the middle of the peptide. Specifically, the occupation decreases from 74% for the first Pro residue in the peptide (Ac–P_1) to 55% for the fifth residue (P_4–P_5), and then increases to 90% for the final residue (P_8–P_9). The P(F) region is occupied for approximately 100% of the simulation for all residues in the strand except for the last prolyl amide bond. Similar results have been obtained in other computational studies of oligoprolines,[32,52-53] which verifies the approach used in this chapter.

Table 4.4. PPII and *trans* content of each residue in all model compounds obtained from unbiased HT–REMD simulations in implicit water.

	$(Pro)_9$			$(Hyp)_9$			$[Hyp–(\beta–Gal)]_9$		
Bond	P(PPII)[a]	P(T)[b]	P(F)[c]	P(PPII)[a]	P(T)[b]	P(F)[c]	P(PPII)[a]	P(T)[b]	P(F)[c]
Ac–P_1	74	74	100	77	77	99	72	74	100
P_1–P_2	75	76	100	89	89	100	90	92	98
P_2–P_3	64	65	100	87	88	100	96	96	100
P_3–P_4	61	61	100	85	84	100	84	81	100
P_4–P_5	56	55	100	87	87	100	80	81	100
P_5–P_6	56	58	100	82	83	98	88	89	99
P_6–P_7	61	62	100	87	87	100	82	85	99
P_7–P_8	86	87	100	89	91	99	90	94	97
P_8–P_9	56	90	67	67	91	76	64	90	75
Average[d]	65	69		83	86		83	87	

[a]P(PII) refers to the percentage of time that a residue adopts the PPII conformation. [b]P(T) refers to the percentage of time that a residue adopts the *trans* conformation. [c]P(F) denotes the percentage of time spent in the F region (PPI and PPII) of the Ramachandran plot. [d]A standard error of the mean of less than 1% was obtained in all cases for P(PPII) and P(T) for all peptides in this thesis.

There is an increase in the number of *trans* conformers in both modified peptides relative to $(Pro)_9$, which is expected since the PTMs have been shown to clearly enhance the *trans* conformation.[51,54-55] Indeed, the *trans* amide isomer exists for an average of 86% and 87% of the simulation time in $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$, respectively. Furthermore, P(PPII) is on average 83% for both peptides. Similar to $(Pro)_9$, P(F) is nearly 100% in all but the last residue in both modified peptides.

As discussed above, a sequence–based analysis on the equilibrium data can be done to determine whether each residue in the peptide adopts the *cis* or *trans* conformation. Subsequently, the probability that a particular conformational sequence is adopted can be obtained. The results of this analysis are shown in Table 4.5 for the seven most likely sequences. In the $(Pro)_9$ peptide, the all *trans* sequence is the most probable conformation.

Table 4.5. The seven most probable conformers of the $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ peptides in terms of the *cis* (C) or *trans* (T) conformation about the ω dihedral angle for each residue according to unbiased HT–REMD simulations in implicit water.

$(Pro)_9$		$(Hyp)_9$		$[Hyp–(β–Gal)]_9$	
Sequence	Prob. (%)	Sequence	Prob. (%)	Sequence	Prob. (%)
TTTTTTTTT	3.3	TTTTTTTTT	26.5	TTTTTTTTT	27.8
TTTTCTTTT	3.0	CTTTTTTTT	7.9	CTTTTTTTT	9.7
TTTTTCTTT	2.4	TTTTTCTTT	5.3	TTTCTTTTT	6.4
TTTTCCTTT	2.2	TTTCTTTTT	4.9	TTTTCTTTT	6.2
TTTCTTTTT	2.1	TTTTCTTTT	3.9	TTTTTTCTT	5.0
TTTTTTCTT	1.9	TTTTTTCTT	3.9	TTTTTCTTT	3.3
TTTCCTTTT	1.9	TTCTTTTTT	3.8	TTTTTTTTC	3.2

However, this sequence has a low probability (3.3%) and a strand with *cis* interruption in the middle of the strand is very close in probability. This implies that the $(Pro)_9$ oligomer likely adopts an ensemble of structures with consecutive *trans* residues interspaced by *cis* residues even at 300 K. This result is in line with several recent experimental and computational findings.[47,49,52] More importantly, the calculations show a significant increase in the all *trans* sequence in both the $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ peptides (26.5 and 27.8% respectively) relative to $(Pro)_9$ (3.3%). This finding correlates with the experimentally-observed extra stabilization of the *trans* isomer due to these modifications.[51] Additionally, sequences with at least 6 *trans* residues exist for 72% of the simulation time in the ensemble of structures obtained for the unmodified peptide compared to 98% of the simulation time for both the hydroxylated and glycosylated peptides.

Using the HT–REMD data, ΔF(PPII–PPI) is estimated to be –20.5, –42.2and –46.0.0 kJ mol^{-1} for $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$, respectively. These values are very similar to those obtained from the PMFs, which suggests that both sets of calculations are well converged. While this similarity is encouraging in terms of convergence, experimental T_m values show that the glycosylated peptide (T_m = 70°C) is more stable than the hydroxylated peptide (T_m = 38°C), which should translate to a significantly more negative calculated ΔF for the glycosylated peptide relative to $(Hyp)_9$. It is anticipated that the absence of explicit water molecules in the ABMD calculations could be responsible for this discrepancy. To test this hypothesis, explicit solvent refinement of the implicit solvent PMFs was performed for the $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ peptides as discussed below. $(Pro)_9$ is not considered in this analysis since the corresponding PMF is already much lower than that of $(Hyp)_9$, which corresponds well with experiment.[4]

4.3.4. ABMD Simulations in Explicit Water

Multiple walker ABMD simulations at 300 K were used to refine the PMFs obtained in implicit solvent. Figure 4.6 compares the final results obtained from explicit solvent simulations for $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$, respectively. This comparison suggests that explicit water molecules have a profound effect on the PMFs. In $(Hyp)_9$, ΔF(PPII–PPI) is approximately −83.6 kJ mol^{-1} in explicit solvent compared to −37.6 kJ mol^{-1} in implicit solvent (Figures 4.5B and 4.6A). An even larger effect occurs for the $[Hyp–(β–Gal)]_9$ peptide, where the free energy difference decreases to approximately −192.3 kJ mol^{-1} from −37.6 kJ mol^{-1} (Figures 4.5C and 4.6B).

Therefore, the state-of-the-art ABMD calculations in explicit water suggest that the PPII structures are more strongly preferred relative to PPI by 108.7 kJ mol^{-1} in $[Hyp–(β–Gal)]_9$ compared to $(Hyp)_9$. This finding is consistent with the experimentally observed T_m values for the two peptides suggesting that discrete sugar–solvent and peptide–solvent interactions are likely responsible for the increased thermal stability of the glycosylated peptide. A similar effect of sugar–explicit solvent interactions has been previously reported in the literature for an N-glycosylated protein.[10]

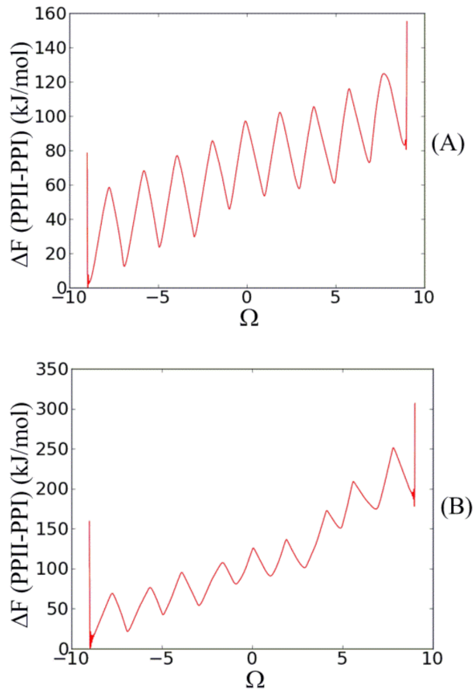

Figure 4.6. Potential of mean force (ΔF(PPII–PPI)) for conversion from PPII (Ω = −9.0) to PPI (Ω = +9.0) for the (A) Ac–$(Hyp)_9$–NH_2 and (B) Ac–[Hyp–(β–D–Gal)]$_9$–NH_2 peptides obtained from explicit solvent ABMD simulations.

4.4. Discussion

Experimental results have shown that modifications to Pro can lead to an increased structural integrity of oligoproline peptides.[4,56] In particular, contiguous galactosylation of a nonaproline oligomer, which is a model of the HRGPs found in plant cell walls, leads to a relatively higher melting temperature when compared to the corresponding hydroxylated model.[4] The stabilizing effects due to contiguous hydroxylation in $(Hyp)_9$ relative to the unglycosylated $(Pro)_9$ peptide can be understood in terms of stereoelectronic ($n \rightarrow \pi^*$) effects that favor the *trans* conformation. 4R-hydroxylation or attachment of electronegative elements such as fluorine to the C^{γ} atom of Pro has been shown to stabilize the *trans* isomer of Pro. In particular $K_{trans/cis}$ of 6.1 and 4.6 was measured for Hyp and Pro in D_2O at 25°C.[50] However, the reason for the increased stability due to glycosylation cannot be fully attributed to such effects since glycosylation and hydroxylation yield similar *trans* preference in a monomeric unit (Chapter 3).[57] To explain this experimental observation, the structure–function relationship of a covalently–linked sugar must be investigated. In this light, the present Chapter outlines detailed conformational sampling of three oligoproline peptides, $(Pro)_9$, $(Hyp)_9$ and $[Hyp–(\beta–Gal)]_9$.

A covalently–linked sugar can stabilize a particular peptide conformation via different modes of action.[10,58-60] First, the glycan could form strong intramolecular hydrogen bonds (see Figure 4.2) to the peptide backbone and/or other glycans. Indeed, this has been proposed in the literature to explain why certain oligomers containing other amino acid residues, such as glutamines (Gln), have a high tendency to form PPII helices.[61] In this case, calculations show that the Gln sidechain forms a hydrogen bond to the preceding main chain carbonyl group, which helps maintain the ψ dihedral of the Gln residue, as well as the φ and ψ angles of the preceding residue, in the PPII conformation. In the case of β–D–galactose addition, the data shows that there are indeed similar intramolecular hydrogen-bonding interactions between neighboring sugars, and between the sugars and the peptide backbone. Specifically, interactions exist between the O6'-hydroxyl hydrogen (Figure 3.2) at position *i* and the carbonyl oxygen (C=O) in the peptide backbone at positions *i*–2 and *i*–3. Additionally, hydrogen bonding occurs between the O2' hydrogen of residue *i* and O6' of residue *i*+1. However, in the presence of explicit solvent molecules, there is a competition between intramolecular interactions and intermolecular interactions with solvent. Indeed, there is a maximum occupancy of only 13% for the sugar–backbone and sugar–sugar interactions in an 80 ns MD simulation. Thus, while the observed intramolecular hydrogen bonding may contribute to the stability of the PPII conformation, it will not lead to the observed 37°C difference in the melting temperature of the $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ oligopeptides. This contradicts a previous proposal (based on molecular mechanics minimizations) that intramolecular hydrogen bonding is responsible for the observed differences in melting temperatures between $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$.[4]

The covalently-linked sugars can also lead to a favoured PPII conformation through a steric mechanism. However, this cannot be the case for the glycosylated peptide studied in this Chapter. Specifically, more energy could be required to transition from a PPII to PPI conformation because of steric clashes during the *cis* to *trans* isomerization. While this may be true for biological HRGP with complex polysaccharides attached,[62] the calculations do not support this proposal for the main source of increased PPII stability in the $[Hyp–(β–Gal)]_9$ model peptide. Indeed, PMFs obtained from implicit solvent ABMD calculations are almost identical for $(Hyp)_9$ and $[Hyp–(β–Gal)]_9$ (Figure 4.5), which leads to very close estimated free energy differences between PPI and PPII. Likewise, the HT–REMD free energy estimates, as well as the estimated proportion of the all *trans* PPII sequence (Table 3.5), are very close for both oligopeptides.

Due to the observations outlined in the previous two paragraphs, the PMFs obtained in implicit solvent were refined in explicit (water) solvent and the subsequent analysis revealed the likely reason for the observed increased stability of the glycosylated peptide. Specifically, the PPII conformation of the Hyp and $[Hyp–(β–Gal)]_9$ peptides are more stabilised relative to PPI in explicit solvent compared to the implicit solvent. However, this effect is greater for the $[Hyp–(β–Gal)]_9$ based on the PMF resulting from explicit solvent calculations. Therefore, the ΔF(PPII–PPI) for $[Hyp–(β–Gal)]_9$ in explicit water is approximately 109 kJ mol^{-1} lower than predicted for the hydroxylated peptide, which agrees with experimental trends in PPII stability as described by the difference in T_m between the peptides.[4] Since implicit solvent does not allow direct hydrogen bonding between the solvent and the backbone or sugar, the only available explanation for the difference in ΔF(PPII–PPI) for $(Hyp)_9$ versus $[Hyp–(β–Gal)]_9$ is that the sugars interact with the water molecules in some fashion that makes the PPII form much more stable. This interaction is likely hydrogen bonding between the sugar hydroxyl groups and water molecules. The equilibrium MD calculations in explicit solvent show that these hydrogen bonds are present most of the time and are only briefly interrupted by intramolecular hydrogen bonding to other sugars or the peptide backbone (Table B2). Conversion of a peptide bond from the *trans* to *cis* configuration would require a disruption of this hydrogen-bonding network and displacement of water molecules around the peptide. Interestingly, several solvent–hydroxyl hydrogen bonds are also observed in the hydroxylated derivatives, which have been highlighted previously in the literature[8] and may work with the stabilizing effects of electronegative hydroxyl substituents to contribute to the greater PPII stability of these oligomers compared to (unsubstituted) oligoprolines.

In summary, the extensive MD calculations reported in this chapter suggest that interactions between sugar moieties covalently attached to an oligoproline peptide and the surrounding water molecules lead to an increased stability of the PPII conformation. Additionally, this explains the recent experimental finding that contiguously galactosylated peptides exhibit higher melting temperatures compared to the corresponding unglycosylated or contiguously hydroxylated peptides.[4] This result highlights a potential mode of action for complex glycans including in the HRGPs of plant cell walls.[2]

4.5. Conclusions

In this chapter, extensive MD simulations were carried out in implicit and explicit water to determine how contiguous glycosylation of Hyp residues as observed in plant HRGPs affects the structure of oligoprolines.. Additionally, it was necessary to rationalize experimental results showing an increase in the stability of the PPII conformation in a contiguously glycosylated nonaproline peptide compared to a hydroxylated variant. A variety of sophisticated molecular modeling techniques were applied including MD in explicit water, REMD in explicit water and ABMD in implicit and explicit water. The results show that glycosylation shields the backbone carbonyls from water in the PPII conformation, which is compensated for by intramolecular hydrogen bonding between the backbone carbonyls and the sugar moiety, as well as sugar–sugar interactions. However, these discrete interactions are not persistent enough in explicit solvent to explain the magnitude of the increase in the conformational stability observed experimentally.[4]

Both hydroxylation and glycosylation of proline increase PPII stability, which can be partly explained by stereoelectronic effects induced by the modifications favouring the *trans* amide isomer. Implicit solvent ABMD calculations indicate that the PPII conformation of both hydroxylated and glycosylated compounds are similar in stability compared to the PPI structure, which does not agree with the significantly higher melting temperatures measured for the glycosylated peptide. Extending the ABMD calculations to include explicit solvent yields the correct stability trend, which highlights the importance of sugar–solvent interactions. Thus, it is concluded that intermolecular interactions between the sugar moieties and the water molecules are the primary reason for the experimentally observed increase in thermal stability of the glycosylated oligoproline compared to the hydroxylated and unmodified peptides. These results provide further support for previous literature proposals[2] that structure stabilization is one of the reasons for the extensive glycosylation of biological HRGPs. The next chapter will consider non–contiguous glycosylation in order to provide structural insight into another biologically–relevant proline glycosylation pattern.

4.6. References[b]

1. Adzhubei, A. A.; Sternberg, M. J. E., Left-Handed Polyproline-II Helices Commonly Occur in Globular Proteins. *J. Mol. Biol.* **1993,** *229*, 472-493.

2. Ferris, P. J.; Woessner, J. P.; Waffenschmidt, S.; Kilz, S.; Drees, J.; Goodenough, U. W., Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. *Biochemistry* **2001,** *40*, 2978-2987.

3. Velasquez, S. M.; Ricardi, M. M.; Dorosz, J. G.; Fernandez, P. V.; Nadra, A. D.; Pol-Fachin, L.; Egelund, J.; Gille, S.; Harholt, J.; Ciancia, M.; Verli, H.; Pauly, M.; Bacic, A.; Olsen, C. E.; Ulvskov, P.; Petersen, B. L.; Somerville, C.; Iusem, N. D.; Estevez, J. M., O-Glycosylated Cell Wall Proteins Are Essential in Root Hair Growth. *Science* **2011,** *332*, 1401–1403.

4. Owens, N. W.; Stetefeld, J.; Lattova, E.; Schweizer, F., Contiguous O-Galactosylation of 4(R)-Hydroxy-L-proline Residues Forms Very Stable Polyproline II Helices. *J. Am. Chem. Soc.* **2010,** *132*, 5036-5042.

5. Pysh, E. S., Randon-phase Calculation of Poly-L-Proline II Circular-Dichroism. *Biopolymers* **1974,** *13*, 1563-1571.

6. Ronish, E. W.; Krimm, S., Calculated Circular-Dichroism of Polyproline II in Polarizibility Approximation. *Biopolymers* **1974,** *13*, 1635-1651.

7. Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A. F.; Creamer, T. P., Host-guest study of left-handed polyproline II helix formation. *Biochemistry* **2001,** *40*, 14376-14383.

8. Horng, J. C.; Raines, R. T., Stereoelectronic effects on polyproline conformation. *Protein Sci.* **2006,** *15*, 74-83.

9. Ellis, C. R.; Maiti, B.; Noid, W. G., Specific and Nonspecific Effects of Glycosylation. *J. Am. Chem. Soc.* **2012,** *134*, 8184-8193.

[b] ACS referencing style was implemented throughout this thesis.

10. Cheng, S. M.; Edwards, S. A.; Jiang, Y. D.; Grater, F., Glycosylation Enhances Peptide Hydrophobic Collapse by Impairing Solvation. *ChemPhysChem* **2010,** *11*, 2367-2374.

11. Riederer, M. A.; Hinnen, A., Removal of N-Glycosylation Sites of the Yeast Acid-Phosphatase Severely Affects Protein Folding. *J. Bacteriol.* **1991,** *173*, 3539-3546.

12. Engelsen, S. B.; Monteiro, C.; de Penhoat, C. H.; Perez, S., The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and NMR spectroscopy. *Biophys. Chem.* **2001,** *93*, 103-127.

13. Aliev, A. E.; Courtier-Murias, D., Conformational analysis of L-prolines in water. *J. Phys. Chem. B* **2007,** *111*, 14034-14042.

14. Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S., A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **2005,** *26*, 1612-1616.

15. Melis, C.; Bussi, G.; Lummis, S. C. R.; Molteni, C., Trans-cis Switching Mechanisms in Proline Analogues and Their Relevance for the Gating of the 5-HT3 Receptor. *J. Phys. Chem. B* **2009,** *113*, 12148-12153.

16. Kang, Y. K.; Choi, H. Y., Cis-trans isomerization and puckering of proline residue. *Biophys. Chem.* **2004,** *111*, 135-142.

17. Kang, Y. K.; Jhon, J. S.; Park, H. S., Conformational preferences of proline oligopeptides. *J. Phys. Chem. B* **2006,** *110*, 17645-17655.

18. Venkatachalam, C. M.; Price, B. J.; Krimm, S., A theoretical estimate of the energy barriers between stable conformations of the proline dimer. *Biopolymers* **1975,** *14*, 1121-1132.

19. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.

20. Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; R.C.Walker; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER*, 10; University of California, San Francisco: 2008.

21. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65*, 712-725.

22. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29*, 622-655.

23. Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S., A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **2005,** *26*, 1612-1616.

24. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000,** *104*, 3712-3720.

25. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004,** *55*, 383-394.

26. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

27. Geyer, C. J., Markov-chain Monte Carlo Maximum-likelihood. *Computing Science and Statistics* **1991**, 156-163.

28. Babin, V.; Karpusenka, V.; Moradi, M.; Roland, C.; Sagui, C., Adaptively Biased Molecular Dynamics: An Umbrella Sampling Method With a Time-Dependent Potential. *Int. J. Quant. Chem.* **2009,** *109*, 3666-3678.

29. Babin, V.; Roland, C.; Sagui, C., Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **2008,** *128,* 134101-134108.

30. Moradi, M.; Babin, V.; Roland, C.; Darden, T. A.; Sagui, C., Conformations and free energy landscapes of polyproline peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 20746-20751.

31. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., PPII Propensity of Multiple-Guest Amino Acids in a Proline-Rich Environment. *J. Phys. Chem. B* **2011,** *115*, 8645-8656.

32. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., A Statistical Analysis of the PPII Propensity of Amino Acid Guests in Proline-Rich Peptides. *Biophys. J.* **2011,** *100*, 1083-1093.

33. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., A Statistical Analysis of the PPII Propensity of Amino Acid Guests in Proline-Rich Peptides. *Biophys. J.* **2011,** *100*, 1083-1093.

34. Cowan, P. M.; McGavin, S., Structure of Poly-L-Proline. *Nature* **1955,** *176*, 501-503.

35. Kakinoki, S.; Hirano, Y.; Oka, M., On the stability of polyproline-I and II structures of proline oligopeptides. *Polym. Bull. (Heidelberg, Ger.)* **2005,** *53*, 109-115.

36. Gornick, F.; Mandelkern, L.; Diorio, A. F.; Roberts, D. E., Evidence for Cooperative Intramolecular Transition in Poly-L-Proline. *J. Am. Chem. Soc.* **1964,** *86*, 2549-2555.

37. Steinberg, I. Z.; Harrington, W. F.; Berger, A.; Sela, M.; Katchalski, E., The Configuartional Changes of Poly-L-Proline in Solution. *J. Am. Chem. Soc.* **1960,** *82*, 5263-5279.

38. Strassma.H; Engel, J.; Zundel, G., Binding of Alcohols to Peptide Co-Group of Poly-L-Proline in I and 2 Conformation .I. Demonstration of Binding by Infrared Spectroscopy and Optical Rotatory Dispersion. *Biopolymers* **1969,** *8*, 237-246.

39. Chen, K.; Liu, Z. G.; Kallenbach, N. R., The polyproline II conformation in short alanine peptides is noncooperative. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 15352-15357.

40. Greenfie.N; Fasman, G. D., Computed Circular Dichroism Spectra for Evaluation of Protein Conformation. *Biochemistry* **1969,** *8*, 4108-4116.

41. Sreerama, N.; Woody, R. W., Molecular dynamics simulations of polypeptide conformations in water: A comparison of alpha, beta, and poly(Pro)II conformations. *Proteins* **1999,** *36*, 400-406.

42. Shi, Z. S.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R., Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99*, 9190-9195.

43. Ding, L.; Chen, K.; Santini, P. A.; Shi, Z. S.; Kallenbach, N. R., The pentapeptide GGAGG has PII conformation. *J. Am. Chem. Soc.* **2003,** *125*, 8092-8093.

44. Avbelj, F.; Baldwin, R. L., Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: Distributions of phi. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100*, 5742-5747.

45. Eker, F.; Griebenow, K.; Schweitzer-Stenner, R., Stable conformations of tripeptides in aqueous solution studied by UV circular dichroism spectroscopy. *J. Am. Chem. Soc.* **2003,** *125*, 8178-8185.

46. Eker, F.; Griebenow, K.; Cao, X. L.; Nafie, L. A.; Schweitzer-Stenner, R., Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis of alanine-based (AXA) tripeptides in aqueous solution. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 10054-10059.

47. Doose, S.; Neuweiler, H.; Barsch, H.; Sauer, M., Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 17400-17405.

48. Best, R. B.; Merchant, K. A.; Gopich, I. V.; Schuler, B.; Bax, A.; Eaton, W. A., Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 18964-18969.

49. Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A., Polyproline II helix conformation in a proline-rich environment: A theoretical study. *Biophys. J.* **2004,** *86*, 731-742.

50. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T., Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **2001,** *123*, 777-778.

51. Owens, N. W.; Braun, C.; O'Neil, J. D.; Marat, K.; Schweizer, F., Effects of glycosylation of (2S,4R)-4-hydroxyproline on the conformation, kinetics, and thermodynamics of prolyl amide isomerization. *J. Am. Chem. Soc.* **2007,** *129*, 11670-11671.

52. Moradi, M.; Babin, V.; Roland, C.; Sagui, C., A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *J. Chem. Phys.* **2010,** *133*.

53. Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A., Polyproline II helix conformation in a proline-rich environment: A theoretical study. *Biophys. J.* **2004,** *86*, 731-742.

54. Panasik, N., Jr.; Eberhardt, E. S.; Edison, A. S.; Powell, D. R.; Raines, R. T., Inductive effects on the structure of proline residues. *Int. J. Pep. Protein Res.* **1994,** *44*, 262-269.

55. Horng, J.-C.; Raines, R. T., Stereoelectronic effects on polyproline conformation. *Protein Sci.* **2006,** *15*, 74–83.

56. Kuemin, M.; Nagel, Y. A.; Schweizer, S.; Monnard, F. W.; Ochsenfeld, C.; Wennemers, H., Tuning the cis/trans Conformer Ratio of Xaa-Pro Amide Bonds by Intramolecular Hydrogen Bonds: The Effect on PPII Helix Stability. *Angew. Chem.-Int. Edit.* **2010,** *49*, 6324-6327.

57. Owens, N. W.; Lee, A.; Marat, K.; Schweizer, F., The Implications of (2S,4S)-Hydroxyproline 4-O-Glycosylation for Prolyl Amide Isomerization. *Chem.-Eur. J.* **2009,** *15*, 10649-10657.

58. Ellis, C. R.; Maiti, B.; Noid, W. G., Specific and Nonspecific Effects of Glycosylation. *J. Am. Chem. Soc.* **2012,** *134*, 8184-8193.

59. Riederer, M. A.; Hinnen, A., Removal of N-glycosylation Sites of the Yeast Acid-Phosphate Severely Affects Protein Folding. *J. Bacteriol.* **1991,** *173*, 3539-3546.

60. Engelsen, S. B.; Monteiro, C.; de Penhoat, C. H.; Perez, S., The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and NMR spectroscopy. *Biophys. Chem.* **2001,** *93*, 103-127.

61. Stapley, B. J.; Creamer, T. P., A survey of left-handed polyproline II helices. *Protein Sci.* **1999,** *8*, 587-595.

62. Tan, L.; Varnai, P.; Lamport, D. T. A.; Yuan, C. H.; Xu, J. F.; Qiu, F.; Kieliszewski, M. J., Plant O-Hydroxyproline Arabinogalactans Are Composed of Repeating Trigalactosyl Subunits with Short Bifurcated Side Chains. *J. Biol. Chem.* **2010,** *285*, 24575-24583.

Chapter 5. Conformational Analysis of Non-contiguously Glycosylated Oligoproline Peptides[a]

5.1. Introduction

It was determined in the previous chapter that contiguous glycosylation of a hydroxyproline oligopeptide significantly increases the stability of the PPII conformation through sugar–solvent intermolecular interactions. This finding explains previous experimental observations that contiguous glycosylation leads to a high T_m value (70°C) compared to the unmodified (T_m = 22°C) or hydroxylated (T_m = 38°C) variants. While this provides useful information about the role of the sugar moiety in contiguously glycosylated hydroxyproline peptides, naturally occurring HRGPs also contain non-contiguous segments of glycosylated Hyp.[1-4] Therefore, the effects of non-contiguous glycosylation must be considered. As a result, compounds with varying degrees of non-contiguous glycosylation (Table 5.1) were synthesized and characterized using CD spectroscopy by Schweizer and coworkers.[5] The results show that all compounds exhibit a minimum at 205 nm and a maximum at 225 nm in their molar ellipticities (θ), which is characteristic of a PPII conformation.[6-8]

Table 5.1. Oligoproline compounds considered in this study**[a]**

Compound	Abbreviation
Ac-(Pro)$_9$-NH_2	(Pro)$_9$
Ac-(Pro-Hyp-Pro)$_3$-NH_2	(POP)$_3$
Ac- (Pro-[Hyp-(α-Galactose)]-Pro)$_3$-NH_2	(PO$^\alpha$P)$_3$
Ac-(Pro-[Hyp-(β- Galactose)]-Pro)$_3$-NH_2	(PO$^\beta$P)$_3$
Ac-(Pro-Hyp-Pro)-(Pro-[Hyp-(α-Galactose)]-Pro)-(Pro-Hyp-Pro)-NH_2	(POP)-(PO$^\alpha$P)-(POP)
Ac-(Pro-Hyp-Pro)-(Pro-[Hyp-(β-Galactose)]-Pro)-(Pro-Hyp-Pro)-NH_2	(POP)-(PO$^\beta$P)-(POP)

[a]O is used to represent the Hyp residue.

[a] E. B. Naziga, S. Bommagani, J. O'Neil, E. Lattova, F. Schweizer and S. D. Wetmore, Conformational Analysis of Non-contiguously Glycosylated Oligoprolines, unpublished work. **2013**.

The experimental data[5] also shows that all non-contiguously hydroxylated and glycosylated peptides register an increase in PPII character based on θ_{max} at 225 nm compared to the unmodified peptide. Unglycosylated $(Pro)_9$ has a $[\theta]_{max}$ of approximately 500 deg cm^2 dmol^{-1} (Table 5.2). In contrast, the triply hydroxylated $(POP)_3$ peptide has a substantially larger value of $[\theta]_{max}$ = 2844 deg cm^2 dmol^{-1} relative to $(Pro)_9$. This result reflects the presence of three Hyp residues, which are more stable than Pro residues in the PPII structure due to the more stable *trans* conformation of Hyp compared to Pro.

Table 5.2. Molar ellipticity at 225 nm for all compounds at 25°C[a].

Compound	$[\theta]_{max}$[b]
$(Pro)_9$	500
$(POP)_3$	2844
$(PO^{\alpha}P)_3$	3268
$(PO^{\beta}P)_3$	2126
(POP)-($PO^{\alpha}P$)-(POP)	3892
(POP)-($PO^{\beta}P$)-(POP)	3433

[a] From reference 5. [b] $[\theta]_{max}$ in units of deg cm^2 dmol^{-1}.

The triply α-glycosylated $(PO^{\alpha}P)_3$ peptide has a $[\theta]_{max}$ of approximately 3268 deg cm^2 dmol^{-1}, which is much larger than the value obtained for the unmodified $(Pro)_9$ compound, but only slightly larger than the triply hydroxylated $(POP)_3$ peptide. This indicates that non-contiguous hydroxylation and glycosylation produce similar effects on the PPII stability. The triply β-glycosylated $(PO^{\beta}P)_3$ peptide, on the other hand, has a smaller ($[\theta]_{max} \approx 2126$ deg cm^2 dmol^{-1}) molar ellipticity than the triply hydroxylated compound, implying that triple β-glycosylation destabilizes the PPII helix compared to triple hydroxylation or α-glycosylation. This contrasts the previous study of contiguously glycosylated peptides that found α and β-glycosylation to yield similar effects on the PPII structure and the reason for this contradiction is not immediately apparent.

Peptides with a single α or β-glycosylation, namely (POP)-(PO$^{\alpha}$P)-(POP) and (POP)-(PO$^{\beta}$P)-(POP), have a maximum molar ellipticity at 225 nm of 3892 deg cm^2 $dmol^{-1}$ and 3433 deg cm^2 $dmol^{-1}$, respectively. These values are larger than for the triply α(β)-glycosylated peptides described above, which suggests that PPII stability may not necessarily increase with a higher degree of glycosylation when (at least) two amino acids are between the glycosylated Hyp residues. Despite these interesting results, understanding the molecular basis for these experimental observations requires detailed atomic level information that can be obtained through computational studies for these complex molecules.

Important structural information was obtained from the experimental study described above, however, CD spectra can be difficult to interpret beyond establishing the presence of the PPII conformation. For instance, while the height of the maximum at 225 nm (θ_{max}) has been correlated to PPII content in the literature,[9-10] T_m data suggests that a $(Flp)_{10}$ oligopeptide with a smaller value of θ_{max} is more stable in the PPII conformation than a $(Hyp)_{10}$ compound with greater θ_{max} at 225 nm.[8,11] A unique T_m could not be determined for the non-contiguously glycosylated peptides since multiple deep minima were obtained from $d\theta_{max}/dT$, suggesting that the PES of these compounds is more complex than the contiguously glycosylated peptide simulated in Chapter 4. Furthermore, as outlined above, the peptides with either a single α or β- glycosylation are the most stable among the non-contiguously glycosylated compounds (Table 5.1) according to θ_{max} (Table 5.2). Therefore, the PPII stability may not necessarily increase with a higher degree of glycosylation for the structures considered in this Chapter, which have a minimum of two amino acids between glycosylated Hyp residues. Understanding the molecular basis for these experimental observations requires detailed atomic level information that can be obtained through molecular modeling studies for these complex molecules.

In the present Chapter, sophisticated molecular modeling methods are used for the first time to study the five non-contiguously modified peptides examined by Schweizer and coworkers (Table 5.1 and Figure 5.1).[5] Data obtained for the $(Pro)_9$ peptide in Chapter 4 is also included in this chapter to allow for comparison with the substituted peptides. As done in Chapter 4, advanced simulation techniques, such as ABMD and HT–REMD are used to comprehensively sample the peptide conformations. This investigation reveals that glycosylation can stabilize the PPII structure in agreement with previous experimental data.[12-13] However, there is a decrease in PPII character as the number of attached sugars increases, which is explained by the formation of new intramolecular interactions in the non-contiguously glycosylated peptides.

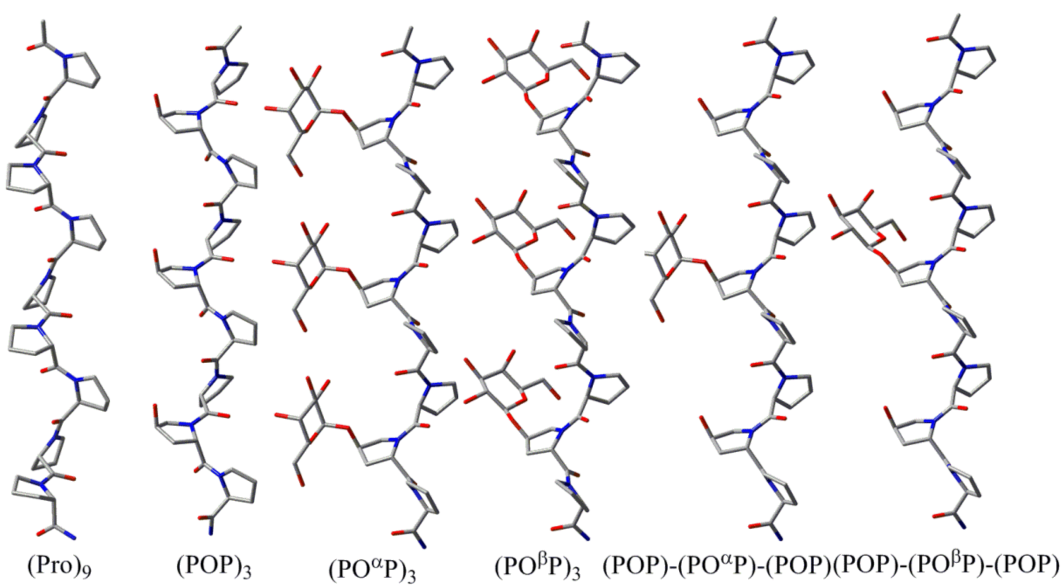

Figure 5.1. Models of the non-contiguously glycosylated oligoproline compounds considered in this chapter.

5.2. Computational Details

To provide structural information on non-contiguous glycosylation of oligoprolines as well as complement and rationalize experimental information regarding the effects of non-contiguous glycosylation on PPII stability, all available degrees of freedom in the five modified peptides considered must be thoroughly sampled, including the puckering transitions of the (substituted) Pro ring and variations in the backbone dihedral angles (φ, ψ and ω). As reported in Chapter 4, several theoretical studies have established that the two major puckering states in Pro (C^{γ}–*exo* and C^{γ}–*endo*), as well as the φ and ψ dihedral angles, are well sampled over the course of a standard MD simulation.[14-17] Consequently, the main consideration for designing a sampling protocol for non-contiguously glycosylated oligoproline peptides is the ω dihedral angle,[18] which dictates whether a particular residue adopts the *cis* or *trans* conformation. Indeed, transitions between unsubstituted Pro isomers require more than 42 kJ mol^{-1},[19-21] and therefore will not readily occur in a standard MD simulation of several hundred nanoseconds.

With the above considerations in mind, a series of MD simulations that adequately sample all low vibrational modes of the oligopeptides was implemented as described in Chapter 4. As per Chapters 3 and 4, all calculations were carried out using AMBER 11[22] with the AMBER FF99SB[23] and GLYCAM 06f[24] force fields describing the peptide and sugar moieties, respectively. Additional parameters for the hydroxyproline residue were obtained from Park *et al*,[25] and the glycosylated Hyp units were prepared as outlined in Chapter 3. Implicit solvent calculations used a generalised Born model,[26-27] while explicit water molecules were represented by the TIP3P model.[28] Further details of each type of calculation are given below.

5.2.1. ABMD Simulations in Implicit Water

ABMD[29-30] simulations were used to overcome the barriers associated with *cis* to *trans* isomerization of the peptide ω dihedral angle. Following recent protocols in the literature for sampling the ω dihedral angle in oligoproline peptides,[17-18,31-32] which was also implemented in Chapter 4, a sum of cosines of ω was adopted as the collective variable (CV). In particular, the following expression was used:

$$\Omega = \sum_i \cos(\omega_i)$$

where Ω sums to +9 or −9 for a PPI (all *cis*) or PPII (all *trans*) structure, respectively. Initially, the ABMD procedure (in a REMD scheme) was implemented to obtain the PMF. The resulting PMFs were subsequently used in HT-REMD calculations to facilitate barrier crossing. Additional details for each type of calculation are outlined below.

a. ABMD PMF Refinement: To facilitate convergence, an initial PMF for the five modified peptides considered in the present chapter was adopted from the $(Hyp)_9$ simulations presented in Chapter 4. Subsequently, refinement calculations were performed with $4\Delta\xi = 0.1$ and τ_F increased in two steps from 100 to 200 ps (see Chapter 3 and literature[33-34] for details about the parameters required in ABMD calculations). At this stage, 24 replicas were implemented, each with a different biasing potential and temperatures ranging from 300 to 1200 K, which has been shown to aid transition between the PPII and PPI helices.[32-33] ABMD simulations of at least 100 ns per replica simulation were used to refine the PMFs for each modified peptide.

b. HT–REMD Calculations: Using PMFs obtained from the ABMD calculations described above, HT–REMD calculations were performed to obtain equilibrium data. In these calculations, a total of 28 replicas were implemented, including 4 additional replicas at 300 K. One of the additional replicas was completely unbiased and was used to collect equilibrium data, while the other three replicas were biased based on the PMF of the 300 K replica and scaled by a range of factors (0.49, 0.76 or 0.90). This procedure enhances the exchange to the unbiased replica.[34] All other simulation parameters were the same as for the refinement simulations. The HT-REMD calculations were run for at least 250 ns per replica for each peptide.

5.2.2. MD Simulations in Explicit Water

The leading conformational sequence in terms of the *cis* and *trans* pattern for each glycosylated compound (see Results section for details) was further simulated in the presence of explicit water. The representative structure from clustering based on the peptide backbone atoms was used as the starting solute (peptide) conformation. Initial models were created with the xLeap module of AMBER, and solvated in an octahedral water box leading to final box sizes of 47, 42, 45, 41 and 42 Å for the $(POP)_3$, $(PO^{\alpha}P)_3$, $(PO^{\beta}P)_3$, (POP)-($PO^{\alpha}P$)-(POP), and (POP)-($PO^{\beta}P$)-(POP) peptides, respectively. Two stages of minimization were carried out. In the first step, the solute molecule was held fixed, while the water molecules were relaxed. Next, the entire system was minimized. The resulting structures were slowly heated from 5 to 300 K over 50 ps and allowed to further equilibrate at the final temperature for another 50 ps in an NVT simulation. The equilibrated systems were then subjected to NPT production calculations of at least 100 ns. A timestep of 1.0 fs was used in the NVT equilibration and 2.0 fs in the NPT calculations. van der Waals interactions were consistently truncated at 10 Å.

Umbrella sampling was implemented in explicit water for the $(POP)_3$ and $(PO^{\beta}P)_3$ peptides to characterize the stability of a specific ω dihedral in the *cis* and *trans* conformation (see Results section for details). The $(POP)_3$ peptide was chosen because it is unglycosylated and has a *cis*/*trans* free energy difference comparable to the Pro residue. On the other hand, the $(PO^{\beta}P)_3$ peptide was chosen to represent the non-contiguously glycosylated peptides that contain intramolecular hydrogen bonds and have a *cis*/*trans* relative free energy that is different from a Pro residue. The −180° to 180° range of the ζ improper dihedral, which was used as the CV (∠(CA–O–CD–CA)),[35] was covered using 36 windows. A force constant of 100 kcal mol^{-1} rad^{-2} was used to restrain the CV around the center of each window considered. A 14 ns MD simulation was carried out for each window to yield a total of a 504 ns MD simulation per peptide. The weighted histogram analysis method[36-37] (WHAM) was implemented to generate the PMF using a computer program developed by Dr. Alan Grossfield.[38]

5.3. Results

5.3.1. Residue-Based Analysis of PPII and Trans Content

Conformations obtained from HT-REMD sampling of $(Pro)_9$, as well as the five non-contiguously hydroxylated/glycosylated peptides (Table 5.1), were analysed for PPII character. A residue was considered to adopt a PPII conformation if (−110° < φ < −20°), (50° < ψ < 180°) or (−120° < ψ < −180°), and the ω dihedral angle is in the *trans* configuration. The PPII conformation is preferred for each residue in the unmodified $(Pro)_9$

peptide (Table 5.3), with the probability of finding a residue in the PPII conformation (P(PPII)) ranging from 56 to 86%, and being on average 65%. Similarly, the percentage of time a particular amino acid adopts the *trans* conformation (P(T), (Table 5.4)) averages 69%, which implies that a significant number of residues (31%) occupy the *cis* conformation. Furthermore, there is a decrease in the frequency of *trans* conformers towards the middle of the peptide. In particular, P(T) decreases from 74% for the first Pro residue in the peptide (Ac–P_1) to 55% for the fifth residue (P_4–P_5), and then increases to 90% for the final residue (P_8–P_9). These results agree with previous computational studies of unmodified oligoprolines of similar length,[32,39] which validates the approach implemented in this thesis.

The average PPII increases for $(POP)_3$ (72%) relative to $(Pro)_9$ (65%). Furthermore, P(T) has a mean value of 75% for $(POP)_3$ (Table 4.4), which is larger than for $(Pro)_9$ (69%). Similar to $(Pro)_9$, P(T) is smallest in the middle of the peptide (52%) for $(POP)_3$, while hydroxylation leads to larger P(T) values for the P_2–P_3, P_5–P_6 and P_8–P_9 ω bonds compared to the unmodified peptide, which is expected since this PTM favours the *trans* conformation.[12,40] Therefore, hydroxylation of three Pro residues stabilizes the PPII conformation compared to $(Pro)_9$, which correlates with the experimentally observed θ_{max} values of 500 and 2844 deg cm^2 dmol^{-1} for $(Pro)_9$ and $(POP)_3$, respectively (Table 5.2).

Table 5.3. PPII content (P(PPII))[a] of each residue in model compounds obtained from unbiased HT–REMD simulations in implicit water.

Compound	Ac–P_1	P_1–P_2	P_2–P_3	P_3–P_4	P_4–P_5	P_5–P_6	P_6–P_7	P_7–P_8	P_8–P_9	Avg.
$(Pro)_9$	74	75	64	61	56	56	61	86	56	65
$(POP)_3$	78	69	85	60	52	82	77	78	65	72
$(PO^{\alpha}P)_3$	82	96	77	30	98	76	29	98	66	72
$(PO^{\beta}P)_3$	68	77	94	28	84	85	31	94	68	70
(POP)-($PO^{\alpha}P$)-(POP)	76	83	76	38	80	79	78	78	64	72
(POP)-($PO^{\beta}P$)-(POP)	77	76	86	35	81	83	75	75	64	72

[a]P(PPII) refers to the percentage of time an ω dihedral angle adopts the PPII conformation.

For the (PO$^{\alpha}$P)$_3$ peptide, the average P(PPII) is the same as for (POP)$_3$ (72%), but different residues have larger (and smaller) contributions to this value than for (POP)$_3$. In particular, P(PPII) for P_1–P_2, P_4–P_5 and P_7–P_8 is 96, 98 and 98% in (PO$^{\alpha}$P)$_3$, respectively, while the same residues have P(PPII) values of 69, 52 and 78% in (POP)$_3$. However, the large P(PPII) for these residues in (PO$^{\alpha}$P)$_3$ is offset by small P(PPII) values for the P_3–P_4 and P_6–P_7 residues (30 and 29%, respectively). In contrast, P(PPII) values in (POP)$_3$ for the P_3–P_4 and P_6–P_7 residues are 60 and 77%, respectively. A similar trend occurs for P(T) (PO$^{\alpha}$P)$_3$, where an average of 75% is obtained, and P(T) equals 94%, 98% and 98% for P_1–P_2, P_4–P_5 and P_7–P_8, but 33 and 31% for P_3–P_4 and P_6–P_7, respectively. Overall, these results imply that glycosylation with three α sugar moieties does not on average enhance the stability of the PPII conformation compared to a triply hydroxylated model.

Table 5.4. *Trans* content (P(T))[a] of each residue in model compounds obtained from unbiased HT–REMD simulations in implicit water.

Compound	Ac–P_1	P_1–P_2	P_2–P_3	P_3–P_4	P_4–P_5	P_5–P_6	P_6–P_7	P_7–P_8	P_8–P_9	Avg.
(Pro)$_9$	74	76	65	61	55	58	62	87	90	69
(POP)$_3$	78	69	85	61	52	82	77	79	94	75
(PO$^{\alpha}$P)$_3$	78	94	76	33	98	75	31	98	97	75
(PO$^{\beta}$P)$_3$	69	76	94	27	85	86	30	95	95	73
(POP)-(PO$^{\alpha}$P)-(POP)	76	83	77	38	79	79	77	79	93	76
(POP)-(PO$^{\beta}$P)-(POP)	77	76	86	35	81	83	75	77	94	76

[a]P(T) refers to the percentage of time that an ω dihedral angle adopts the *trans* conformation.

The β-glycosylated (PO$^{\beta}$P)$_3$ peptide exhibits a similar pattern of high and low P(PPII) and P(T) values across the constituent amino acid residues as observed for (PO$^{\alpha}$P)$_3$. Specifically, the average P(PPII) is 70%, while P(PPII) for P_2–P_3, P_4–P_5 and P_7–P_8 is 94, 84 and 94%, but P(PPII) for P_3–P_4 and P_6–P_7 is 28 and 31%, respectively. Similarly, the average P(T) is 73%, with the P_2–P_3, P_4–P_5 and P_7–P_8 ω bonds having *trans* occupancies of 94, 86 and 95%, respectively. In contrast, the P_3–P_4 and P_6–P_7 ω dihedrals adopt the *trans* conformation for a significantly smaller fraction of the simulation (27 and 30%, respectively). Therefore, glycosylation with a triplet of α or β-galactose sugars produces on average similar effects on the peptide backbone in terms of the P(PPII) and P(T) parameters, albeit slightly smaller values occur for the β-glycosylated variant.

The (POP)-(PO$^{\alpha}$P)-(POP) peptide has an average P(PPII) of 72%. As discussed for the α-glycosylated (PO$^{\alpha}$P)$_3$ peptide, P_1–P_2 and P_4–P_5 have a relatively higher P(PPII) (83 and 80 %, respectively) compared to (POP)$_3$. However, only P_3–P_4 has a small PPII content (38%) in (POP)-(PO$^{\alpha}$P)-(POP). As seen for the other peptides, P(T) exhibits the same trend as P(PPII). Specifically, the average P(T) for (POP)-(PO$^{\alpha}$P)-(POP) is 76%, with 83 and 79% for P_1–P_2 and P_4–P_5, respectively, but 38% for P_3–P_4. Thus, glycosylation with a single α-galactose causes P(PPII) and P(T) to decrease in only one residue, while triple glycosylation causes a decrease in these values on two residues as described in the previous paragraph.

The average P(PPII) for (POP)-(PO$^{\beta}$P)-(POP) is 72%. Additionally, the P_1–P_2 and P_4–P_5 residues have PPII content of 76% and 81%, respectively, while P_3–P_4 adopts the PPII conformation for much less of the total simulation time (35%). The average P(T) is the same as reported for the (POP)-(PO$^{\alpha}$P)-(POP) peptide (76%). Similarly, the P_3–P_4 dihedral angle for (POP)-(PO$^{\beta}$P)-(POP) has a much reduced P(T) (35%) compared to other residues in the same peptide. Thus, a single β-glycosylation reduces the P(PPII) and P(T) of one residue and yields similar average values for P(PPII) and P(T) as a single α-glycosylation.

Some trends emerge from the above analysis of the P(PPII) and P(T) values in the various substituted oligoprolines. In particular, hydroxylation and glycosylation stabilize the *trans* isomer of select residues compared to the unmodified peptide, which is in agreement with several experimental[12-13,41-42] and theoretical[14,35,43-44] studies of hydroxylated Pro and glycosylated Hyp peptides. The PPII and *trans* content is on average larger in all modified compounds compared to (Pro)$_9$, which correlates with the measured relative heights of the ellipticity at 225 nm (Table 5.2). Additionally, glycosylation leads to a low P(PPII) and P(T) in some residues. Since the number of such residues increases with the degree of glycosylation, this phenomena could explain the experimentally observed decrease in θ_{max} for the triply glycosylated peptides compared to the singly glycosylated compounds. However, these results are in contrast to those from implicit HT-REMD calculations on the contiguously glycosylated peptide in Chapter 4, where all residues adopt the *trans* conformation for a majority of the simulation. This implies that the degree and position of glycosylation can modulate the PPII stability of these oligoprolines.

5.3.2. Sequenced-Based Analysis of *Trans* Content.

The previous subsection considered the conformation of individual residues over the course of an unbiased HT-REMD simulation. In the present subsection, a sequence-based analysis is implemented to investigate whether non-contiguous modifications affect the identity and populations of the preferred conformational sequences (Table 5.5).

As reported in Chapter 4, the all *trans* sequence (TTTTTTTTT) is the most prevalent conformational sequence of the unmodified $(Pro)_9$ peptide, with a 3.3% occupancy. However, the next most prevalent sequence has a *cis* interruption in the middle of the peptide and a very similar occupancy (3.0%). Additionally, the next five most occupied sequences have one or two *cis* conformation(s) at various positions (~2% occupancies). This indicates that the unmodified peptide will adopt a variety of sequences in terms of *cis* and *trans* patterns in water at room temperature, which is in accordance with other experimental and theoretical studies of proline peptides.[34,45]

Table 5.5. The seven most probable conformers of the model compounds in terms of the *cis* (C) or *trans* (T) conformation about the ω dihedral angle for each residue according to unbiased HT–REMD simulations in implicit water.

Compound	1st	2nd	3rd	4th	5th	6th	7th
$(Pro)_9$	TTTTTTTTT (3.3)	TTTTCTTTT (3.0)	TTTTTCTTT (2.4)	TTTTCCTTT (2.2)	TTTCTTTTT (2.1)	TTTTTTCTT (1.9)	TTTCCTTTT (1.9)
$(POP)_3$	TTTTTTTTT (6.8)	TTTTCTTTT (6.2)	TTTCTTTTT (4.5)	TTTTCCTTT (4.1)	TTTCTTTTT (2.1)	TCTTTTTTT (3.0)	TCTTCTTTT (2.7)
$(PO^{\alpha}P)_3$	TTTCTTCTT (17.4)	TTTTTTCTT (8.5)	TTTCTTTTT (7.9)	TTTCTCCTT (5.9)	TTCCTCCTT (5.6)	CTTCTTCTT (5.0)	TTTTTTTTT (3.9)
$(PO^{\beta}P)_3$	TTTCTTCTT (16.4)	CTTCTTCTT (7.3)	TTTCTTTTT (7.1)	TTTTTTCTT (6.0)	TCTCTTCTT (5.3)	CTTCTTTTT (3.2)	TTTCCTCTT (3.0)
(POP)-($PO^{\alpha}P$)-(POP)	TTTCTTTTT (10.8)	TTTTTTTTT (6.6)	CTTCTTTTT (3.4)	TTCCTTTTT (3.3)	TTTCTTCTT (3.2)	TTTCCTTTT (2.8)	TTTCTCTTT (2.8)
(POP)-($PO^{\beta}P$)-(POP)	TTTCTTTTT (12.5)	TTTTTTTTT (7.4)	TCTCTTTTT (4.4)	TTTCTTCTT (3.6)	TTTCTTTCT (3.5)	CTTCTTTTT (3.1)	TTTCTCTTT (3.0)

The $(POP)_3$ peptide shows a higher percentage of the all *trans* sequence (6.8%) than $(Pro)_9$ (3.3%). However, similar to the unmodified peptide, the TTTTCTTTT sequence occurs with close probability (6.2%), which implies that the two sequences are nearly isoenergetic. Interestingly, the TTTTTTTTT sequence is the seventh most populated conformational sequence for the $(PO^{\alpha}P)_3$ peptide (3.6% occupancy), and the tenth most frequent sequence for the $(PO^{\beta}P)_3$ peptide (2.6%). Furthermore, the most populated sequence for both triply glycosylated peptides is TTCTTCTT, with occupancies of 17.4 and 16.4% for $(PO^{\alpha}P)_3$ and $(PO^{\beta}P)_3$, respectively. The *cis* interruptions in both $(PO^{\alpha}P)_3$ and $(PO^{\beta}P)_3$ correspond to the P_3–P_4 and P_6–P_7 ω dihedrals, which have the smallest P(PPII) and P(T) values among the ω dihedrals as described in the previous section.

For both the (POP)-$(PO^{\alpha}P)$-(POP) and (POP)-$(PO^{\beta}P)$-(POP) peptides, the TTTCTTTTT sequence is most frequently adopted, with 10.8 and 12.5% occupancies, respectively (Table 5.5). Additionally, the all *trans* TTTTTTTTT sequence is the next most populated for both singly glycosylated peptides, with an occupancy of 6.6%(7.4%) for the α(β)-glycosylated peptide. The *cis* residue in the most populated TTTCTTTTT sequence corresponds to the P_3–P_4 ω bond, which has the lowest *trans* frequency (Table 5.4), and hence PPII conformation, in these peptides (Table 5.3).

In summary, among the peptides considered in this Chapter, only $(Pro)_9$ and $(POP)_3$ most frequently adopt the all *trans* conformational sequence. Furthermore, glycosylation with one (three) galactose sugars introduces one (two) *cis* isomers into the sequence. Thus, the greater the number of non-contiguous sugar moieties, the more *cis* interruptions in the most frequently adopted conformational sequence. A possible explanation for this observation is that the sugar moieties introduce intramolecular interactions that affect the conformation of the peptide backbone, and therefore the PPII stability. This hypothesis is considered in the next section.

5.3.3. Sugar–Peptide Backbone Intramolecular Hydrogen-Bonding Interactions

An important feature of the PPII conformation of unmodified oligoprolines is the absence of intramolecular hydrogen-bonding interactions.[46] However, posttranslational glycosylation can lead to glycan–peptide backbone interactions.[18] Such inter/intramolecular interactions have been used in the literature to explain why certain oligomers containing other amino acid residues, such as glutamines (Gln), have a high tendency to form PPII helices.[46] Specifically, calculations revealed that the Gln sidechain forms a hydrogen bond to the preceding main chain carbonyl group, which helps maintain the ψ dihedral of the Gln residue, as well as the φ and ψ angles of the preceding residue, in the PPII conformation. Therefore, the structures of the non-contiguously glycosylated peptides considered in the present chapter must be analyzed for such contacts in order to understand changes in the PPII content with the degree of glycosylation.

Table 5.6. Intramolecular hydrogen bonding in the PPII conformation of $(PO^{\alpha}P)_3$ observed during HT-REMD simulation at 300 K in implicit water.[a,b]

Acceptor H res@atom	Donor res@atom	% occupied
3@O	9@H3O	52
2@O	6@H3O	22
4@O	6@H2O	19
7@O	9@H2O	17
5@O	9@H3O	13
2@O	6@H2O	11
5@O	9@H2O	7
1@O	3@H2O	7
3@O	9@H4O	5

[a]A hydrogen bond is considered to be present when the donor–acceptor distance is 3.5 Å or less, and the donor–donor hydrogen-bond acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

As expected,[46] no intramolecular hydrogen bonding is observed in the $(Pro)_9$ peptide. Furthermore, apart from Pro residues, the $(POP)_3$ peptide only contains Hyp residues, which does not introduce intramolecular interactions as determined in Chapter 4. However, several sugar–peptide interactions (Table 5.6) are observed in $(PO^{\alpha}P)_3$. The most persistent hydrogen bond (with an approximate 52% occupancy) forms between a glycosylated Hyp (H30, residue 9) and a backbone carbonyl oxygen (O, residue 3). Other hydrogen bonds that are present for approximately 20% of the calculation include one between H3O of residue 6 and the carbonyl O of residue 2, as well as the H2O of residue 6 and the carbonyl O of residue 4.

Similar hydrogen-bonding interactions are observed in the $(PO^{\beta}P)_3$ peptide (Table 5.7, Figure 5.2). Specifically, the most frequent interaction in $(PO^{\beta}P)_3$ occurs between a glycosylated Hyp (H20, residue 9) and a backbone carbonyl oxygen (O, residue 5) for 30% of the trajectory (Table 5.7). The (POP)–(PO$^{\alpha}$P)–(POP) peptide also displays noteworthy intramolecular interactions (Table 5.8), where the longest lasting hydrogen bond (H20, residue 6 and O, residue 3) has an occupancy of 11%. Finally, the (POP)**–**(PO$^{\beta}$P)**–**(POP) peptide also has several hydrogen bonds (Table 5.9), with the most frequent interaction occurring for 15% of the trajectory (H20, residue 6 and O, residue 2). Interestingly, the hydrogen-bonding interaction between the sugar hydroxyl and backbone carbonyl in all glycosylated peptides does not require a direct contact with the residue in the *cis* conformation. For instance, the interaction between H3O of residue 6 and the carbonyl of residue 2 in the (POG)-(PO$^{\alpha}$P)-(POG) peptide is facilitated by a *cis* conformation at residue 4, which appropriately bends the peptide to permit the interaction.

In summary, glycosylation leads to intramolecular sugar–backbone hydrogen-bonding interactions in all glycosylated peptides considered herein, while no corresponding contacts occur in the unglycosylated $(Pro)_9$ and $(POP)_3$ peptides. These hydrogen-bonding interactions are one way that glycosylation can introduce new structural preferences into substituted oligoproline peptides and thereby affect the PPII content in oligoprolines.[46] However, the presence of such intramolecular hydrogen-bonding interactions in contiguously glycosylated peptides is greatly diminished in explicit water (see Chapter 4, Section 4.4), while sugar–solvent interactions become important for dictating the preferred structure of the peptide.[18] Therefore, the presence of the sugar–backbone hydrogen-bonding patterns discussed in this subsection must be validated in the presence of explicit water molecules.

Table 5.7. Intramolecular hydrogen bonding in the PPII conformation of $(PO^{\beta}P)_3$ observed during HT-REMD simulation at 300 K in implicit water.[a,b]

Acceptor H	Donor	
res@atom	res@atom	% occupied
5@O	9@H2O	30
2@O	6@H2O	23
3@O	9@H3O	13
3@O	9@H4O	7
5@O	9@H3O	4
3@O	9@H2O	2
2@O	6@H3O	1
1@O	3@H6O	1
2@O	3@H6O	1
5@O	6@H6O	1

[a]A hydrogen bond is considered to be present when the donor–acceptor distance is 3.5 Å or less, and the donor–donor hydrogen-bond acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

Table 5.8. Intramolecular hydrogen bonding in the PPII conformation of (POP)-$(PO^{\alpha}P)$-(POP) observed during HT-REMD simulation at 300 K in implicit water.[a,b]

Acceptor H	Donor	
res@atom	res@atom	% occupied
4@O	6@H2O	11
2@O	6@H3O	9
2@O	6@H2O	4

[a]A hydrogen bond is considered to be present when the donor–acceptor distance is 3.5 Å or less, and the donor–donor hydrogen-bond acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

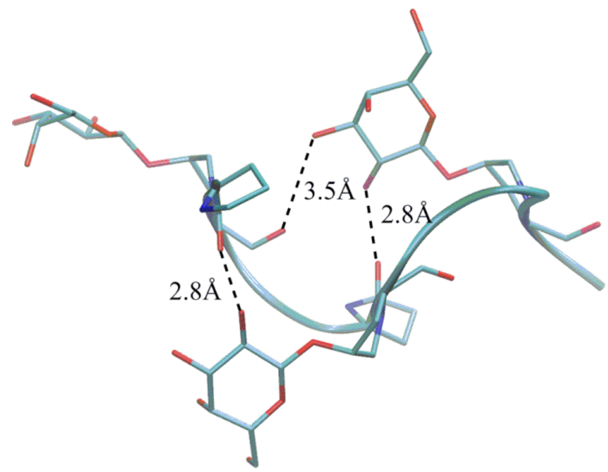

Figure 5.2. Intramolecular hydrogen bonding in the $(PO^{\beta}P)_3$ peptide observed during MD simulation at 300 K in explicit water.

Table 5.9. Intramolecular hydrogen bonding in the PPII conformation of (POP)-$(PO^{\beta}P)$-(POP) observed during HT-REMD simulation at 300 K in implicit water.[a,b]

Acceptor H res@atom	Donor res@atom	% occupied
2@O	6@H2O	15
5@O	6@H6O	1
2@O	6@H3O	1
4@O	6@H6O	1

[a]A hydrogen bond is considered to be present when the donor–acceptor distance is 3.5 Å or less, and the donor–donor hydrogen-bond acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

Table 5.10. Intramolecular hydrogen bonding in the PPII conformation of $(PO^{\alpha}P)_3$ observed during MD simulation at 300 K in explicit water.[a,b]

Acceptor H res@atom	Donor res@atom	% occupied
5@O	9@H3O	34
3@O	9@H3O	26
3@O	9@H4O	20
2@O	6@H3O	19
5@O	9@H2O	5
2@O	6@H2O	4
2@O	6@H6O	3
1@O	3@H2O	3

[a]A hydrogen bond is considered to be present when the donor–acceptor distance is 3.5 Å or less, and the donor–donor hydrogen-bond acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

5.3.4. Conformational Preferences in Explicit Solvent

The persistence of the sugar–backbone intramolecular hydrogen bonds in the non-contiguously glycosylated peptides will be assessed in explicit solvent in the present section. However, the $(Pro)_9$ and $(POP)_3$ peptides are not considered since no intramolecular interactions occur along the corresponding trajectories and the all *trans* TTTTTTTTT conformational sequence is the most frequent for both peptides, even in implicit solvent (Table 5.5).

Table 5.11. Intramolecular hydrogen bonding in the PPII conformation of $(PO^{\beta}P)_3$ observed during MD simulation at 300 K in explicit water.[a,b]

Acceptor H	Donor	
res@atom	res@atom	% occupied
2@O	6@H2O	85
3@O	9@H4O	83
5@O	9@H3O	81
2@O	6@H3O	22
5@O	9@H2O	15
3@O	9@H3O	11
1@O	3@H6O	7
1@O	6@H3O	2
5@O	9@H4O	1

[a]A hydrogen bond is considered to be present when the donor–acceptor distance is 3.5 Å or less, and the donor–donor hydrogen-bond acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

Table 5.12. Intramolecular hydrogen bonding in the PPII conformation of (POP)-($PO^{\alpha}P$)-(POP) observed during MD simulation at 300 K in explicit water.[a,b]

Acceptor H	Donor	
res@atom	res@atom	% occupied
2@O	6@H3O	23
2@O	6@H2O	9
2@O	6@H6O	2

[a]A hydrogen bond is considered to be present when the donor–acceptor distance is 3.5 Å or less, and the donor–donor hydrogen-bond acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

Explicit solvent MD calculations were initiated from the most frequent conformational sequence according to the implicit solvent analysis (subsection 5.3.2) for the glycosylated peptides (Table 5.5), which each contain at least one *cis* ω dihedral angle.

Support for this protocol was obtained from a cluster analysis of the implicit solvent HT-REMD trajectory (based on the CA, N, C and O backbone atoms). For example, clustering the $(PO^{\beta}P)_3$ trajectory into ten groups shows that the first cluster, which represents approximately 56% of the trajectory, corresponds to the same *cis*/*trans* pattern as the most frequent conformational sequence in implicit solvent. Furthermore, hydrogen-bond analysis of the conformation in this dominant first cluster shows that the top four hydrogen-bonding contacts (Table 5.7) for the entire trajectory approximately double for the conformers in this cluster. On the other hand, the same analysis for the next most populated cluster, which represents approximately 13% of the entire HT-REMD trajectory, does not contain any sugar–backbone interactions. This implies that the intramolecular hydrogen-bonding interactions are linked to the conformation in the first cluster, and justifies using the corresponding representative structure as the starting structure in explicit solvent calculations.

Table 5.13. Intramolecular hydrogen bonding in the PPII conformation of (POP)-$(PO^{\beta}P)$-(POP) observed during MD simulation at 300 K in explicit water.[a,b]

Acceptor H	Donor	
res@atom	res@atom	% occupied
2@O	6@O2	64
2@O	6@O3	19
6@O4	6@O6	7
6@O6	6@O4	5
1@O	6@O3	4
2@O	6@O4	1
6@O3	6@O4	1
2@O	6@O6	1

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

The hydrogen-bond analysis of the explicit solvent trajectories is summarised in Tables 5.10 to 5.13. Overall, while the occupancies of the hydrogen-bonding interactions in the $(PO^{\alpha}P)_3$ peptide are altered compared to the implicit solvent calculation (Tables 5.6 and

5.10), the results clearly show that the interactions prevail and, in some cases, are even more prevalent in explicit solvent. For example, the most persistent (52% of the simulation) hydrogen bond in implicit solvent, which occurs between a glycosylated Hyp (H30, residue 9) and a backbone carbonyl oxygen (O, residue 3), is present for 26% of the simulation in explicit solvent. However, the population of the hydrogen bond between H3O of residue 9 and the carbonyl O of residue 5 increases from approximately 13% in implicit solvent to approximately 34% in explicit solvent. Similarly, there is a significant increase in the hydrogen-bonding occupancies of many backbone–sugar interactions in the $(PO^{\beta}P)_3$ peptide upon consideration of solvent (Tables 5.7 and 5.11). For example, the occupancy of the hydrogen bond between H2O of residue 6 and the carbonyl oxygen of residue 2 is approximately 23 and 85% in implicit and explicit solvent, respectively. Interestingly the occupancies indicate that the intramolecular hydrogen bonds are stronger in $(PO^{\beta}P)_3$ compared to $(PO^{\alpha}P)_3$ which implies that the *cis* interruptions will be more stable in the β-glycosylated compound. This correlates with the experimentally observed larger $[\theta]_{max}$ value of 3268 deg cm^{-3} $dmol^{-1}$ for $(PO^{\alpha}P)_3$ compared to 2126 deg cm^{-3} $dmol^{-1}$ for $(PO^{\beta}P)_3$ (Table 5.2).

While there are fewer hydrogen bonds in implicit solvent in the (POG)-$(PO^{\alpha}P)$-(POG) peptide compared to the triply glycosylated compounds, the occupancy of these intramolecular hydrogen bonds also increases in explicit solvent MD calculations (Tables 5.8 and 5.12). Specifically, the most frequent hydrogen bond in implicit solvent, which occurs for approximately 11% of the trajectory between a glycosylated Hyp (H20, residue 6) and a backbone carbonyl oxygen (O, residue 4), disappears in explicit solvent. However, the interaction between H3O of residue 6 and the carbonyl O of residue 2 increases from approximately 9% in implicit solvent to 23% in explicit solvent. Similarly, the hydrogen bond between H2O of residue 6 and the carbonyl O of residue 2 in the (POG)-(PO$^{\beta}$P)-(POG) peptide grows from 15% in implicit solvent to 64% in explicit solvent (Tables 5.9 and 5.13). As observed for the triply glycosylated peptides, the intramolecular hydrogen bonds are stronger in explicit solvent for the β-glycosylated peptide compared to the α-glycosylated variant. The higher hydrogen-bond occupancy for (POG)-(PO$^{\beta}$P)-(POG) correlates with the lower $[\theta]_{max}$ value of 3433 deg cm^{-3} dmol^{-1} compared to the $[\theta]_{max}$ value of 3892 deg cm^{-3} dmol^{-1} determined for (POG)-(PO$^{\alpha}$P)-(POG). The results for both singly and triply glycosylated peptides suggest that non-contiguous glycosylation with a β sugar reduces the PPII content to a greater extent than α glycosylation. This can be explained by the results obtained in the Chapter 3 where both conformations of the β-glycosylated monomer contained more extended conformations. The extended conformations allow for longer range contacts than the stacked-like structures observed in the α-glycosylated variants.

Thus, while the exact percentage occupancies vary, the intramolecular hydrogen-bonding interactions in the non-contiguously glycosylated compounds observed in implicit solvent largely prevail in explicit solvent despite competition with interactions between the sugar and water molecules. In fact, in some cases, the hydrogen-bonding occupancy increases in explicit water. This finding is in contrast to the results for the contiguously glycosylated nonaproline peptide considered in Chapter 4 where a maximum intramolecular hydrogen-bond occupancy of 13% was observed in explicit solvent. This difference is explained by the fact that there are at most three sugars and a minimum of two unglycosylated residues between the glycosylated Hyp residues considered in the present work. Furthermore, intermolecular interactions between the sugar moieties and water molecules will be reduced in the non-contiguously glycosylated peptides due to the persistence of the sugar–backbone intramolecular interactions. In contrast, intramolecular hydrogen bonding is reduced in the contiguously glycosylated peptide due to steric crowding, which permits strong sugar–water interactions. Most importantly, explicit solvent MD calculations support the proposal that sugar–backbone intramolecular hydrogen bonding stabilizes the *cis* conformations of some residues in the non-contiguously glycosylated compounds.

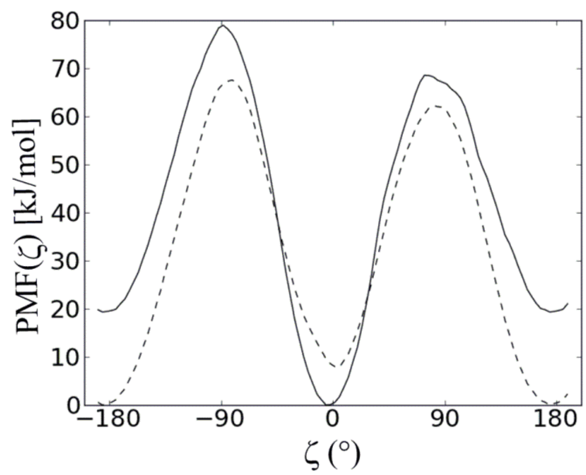

Figure 5.3. Potential of mean force (PMF) for *cis*-*trans* isomerization of the P_3–P_4 ω dihedral angle obtained from explicit solvent umbrella sampling calculations for the $(POP)_3$ (dashed line) and $(PO^{\beta}P)_3$ (solid line) peptides.

The role of intramolecular hydrogen-bonding interactions in dictating the structure of the non-contiguously glycosylated peptides is supported by umbrella sampling calculations that allow comparison of the stability of the *trans* and *cis* conformers of the P_3–P_4 ω dihedral angle in the $(POP)_3$ and $(PO^{\beta}P)_3$ peptides. The $(POP)_3$ peptide was chosen since HT-REMD calculations show that the P_3–P_4 ω dihedral angle exhibits *trans* percentages (61%) that are consistent with an unmodified Pro residue and this peptide does not contain any intramolecular hydrogen bonds. In comparison, the corresponding dihedral angle in the $(PO^{\beta}P)_3$ peptide has an uncharacteristically low *trans* occupation (27%) and this peptide contains the most persistent intramolecular hydrogen bond in explicit solvent (~80% occupancy). Umbrella sampling shows that the *trans* conformation ($\zeta \sim \pm$ 180°) is ~ 8 kJ mol^{-1} more stable than the *cis* conformation ($\zeta \sim$ 0°) for the (POP)$_3$ peptide (Figure 5.3). In contrast, the *cis* isomer is ~ 19 kJ mol^{-1} more stable than the *trans* for the (PO$^{\beta}$P)$_3$ peptide. This result is in accordance with the HT-REMD calculations (Tables 5.3 to 5.5), showing significant population of the *cis* isomer of the P$_3$–P$_4$ ω dihedral for the (PO$^{\beta}$P)$_3$ peptide. Thus, sugar–peptide hydrogen bonding significantly affects the conformational preference of the (PO$^{\beta}$P)$_3$ peptide compared to the corresponding unglycosylated variant. Although these extensive calculations were not conducted for all non-contiguously glycosylated peptides considered within, it is a fair assumption that this conclusion will extend other non-contiguously glycosylated peptides.

5.4. Discussion

Posttranslational glycosylation of Hyp residues is widespread in nature,[3] where the covalently-linked glycan has been proposed to have important biological consequences.[2,47-48] To characterize how glycosylation affects protein structure using controlled models, contiguously glycosylated nonaproline peptides were previously synthesized,[11] and computationally modeled,[18] which revealed that glycosylation enhances the structural integrity of the peptide through sugar–solvent interactions. This finding supports the proposal that enhancing structural stability is one of the functions of the sugar units in biological HRGPs.[47] Non-contiguously glycosylated oligopeptides (Table 5.1), which model naturally occurring HGRPs that contain galactosylated clusters of non-contiguous Hyp,[1] were synthesised in an effort to understand the relative effects of contiguous and non-contiguous glycosylation.[5] However, CD spectroscopy revealed that non-contiguous glycosylation has complex effects on the PPII conformation (Table 5.2), where all glycosylated peptides are more stable than the unmodified peptide, but peptides with three sugars are less stable than those glycosylated with a single sugar (based on the amplitude of the CD maximum at 225 nm). Therefore, molecular modeling were carried out to understand the consequences of non-contiguous glycosylation and provide molecular level information to help rationalize the experimental results.

Sophisticated HT-REMD calculations which enable statistically meaningful sampling of the *cis* and *trans* isomers about the prolyl ω dihedral angles, reveal an overall high *trans* content (69% or more) in all non-contiguously glycosylated peptides considered in the present work. Consequently, a high percentage of residues adopt the PPII conformation. A sequence-based analysis reveals that the all *trans* conformational sequence (TTTTTTTTT) is the most frequent conformer adopted by the $(Pro)_9$ and $(POP)_3$ peptides, although sequences containing one *cis* residue (TTTTCTTTT) have a similar occupancy to the (TTTTTTTTT) conformational sequence. Additionally, no intramolecular hydrogen-bonding interactions occur in either peptide, which is in line with previous literature.[46]

The $(PO^{\alpha}P)_3$, $(PO^{\beta}P)_3$, (POP)-($PO^{\alpha}P$)-(POP) and (POP)-($PO^{\beta}P$)-(POP) peptides yield a very different picture from the corresponding unglycosylated oligoprolines in terms of the most populated conformational sequence and the intramolecular hydrogen-bonding patterns. Specifically, the all *trans* sequence (TTTTTTTTT) is not the most frequently adopted conformer for any of the four non-contiguously glycosylated peptides considered in this work. In fact, (TTTTTTTTT) is the seventh and tenth most populated conformational sequence in the $(PO^{\alpha}P)_3$ and $(PO^{\beta}P)_3$ peptides, respectively. Instead, the most frequent sequence for both triply glycosylated peptides has two *cis* conformers, one at the P_3–P_4 and one at the P_6–P_7 ω dihedral angle. For the singly-glycosylated (POG)-($PO^{\alpha}P$)-(POG) and

(POG)-(PO$^{\beta}$P)-(POG) peptides, the most frequent sequence has one *cis* conformer at P_3–P_4, while the all *trans* sequence is the second most populated conformational sequence.

The presence of residues in the *cis* conformation at certain positions in the non-contiguously glycosylated peptides is associated with the formation of intramolecular hydrogen bonds. Indeed, the *cis* conformation effectively bends the peptide and places the sugar moiety in proximity of the peptide backbone, which make sugar–backbone hydrogen bonding possible. Additionally, the number of such sugar–backbone intramolecular interactions increases with the number of sugar molecules attached to the peptide. For instance, the (PO$^{\alpha}$P)$_3$ peptide (three sugars) has nine intramolecular interactions compared to three in the (POG)-(PO$^{\alpha}$P)-(POG) peptide (one sugar) according to implicit solvent calculations. A similar picture emerges when the most frequent sequence is used in explicit solvent MD simulations. Therefore, non-contiguous glycosylation introduces intramolecular hydrogen bonding that stabilizes the *cis* conformations of some residues in the peptides, and thereby destabilizes the PPII conformation of the glycosylated peptides. The number of intramolecular sugar–backbone hydrogen bonds increases with the number of residues adopting the *cis* conformation, which increases with the number of sugars attached to the non-glycosylated oligoproline peptide. Furthermore, the intramolecular hydrogen bonds in explicit solvent and by extension the *cis* interruptions are more stable in the β-glycosylated compared to the α-glycosylated variants.

The difference in the effects of non-contiguous and contiguous glycosylation noted here for the first time is particular intriguing. Specifically, the present Chapter shows an apparent decrease in PPII stability upon non-contiguous glycosylation with three sugar moieties compared to glycosylation with a single sugar, which correlates with an increase in the number of sugar–peptide backbone interactions. In contrast, the PPII structure is stabilized due to sugar–solvent interactions upon contiguous glycosylation (Ac-[Hyp-(β-Gal)]$_9$-NH$_2$).[11] Previous experimental $K_{trans/cis}$ data[12-13] and computational results[35,41] have shown that hydroxylation and glycosylation lead to comparable stabilization of the *trans* isomer of Pro. Therefore, any further consequences of glycosylation on the structure of oligoprolines are related to specific interactions in the peptide, such as sugar–peptide backbone, sugar–sugar or sugar–solvent interactions.[18] The number and proximity of the sugars dictates which of these interactions will dominate in explicit solvent. In particular, strong sugar–backbone interactions occur more frequently in the non-contiguously glycosylated peptides, which lower the PPII content of some residues, since there are at least two unmodified residues between glycosylated Hyp, which prevents sugar–sugar contacts. In contrast, steric restrictions prevent strong sugar–backbone interactions in fully contiguously glycosylated proline oligomers, and the maximum occupancy of sugar–sugar interactions is too low to explain experimental observations. Instead, sugar–solvent interactions are the main factor stabilizing the dominant backbone conformation. Therefore, the pattern and degree of glycosylation dictates the PPII stability of glycosylated oligoprolines.

The results for non-contiguous glycosylation are similar to previous literature that shows non-covalent interactions can tune *cis/trans* stability in Pro containing peptides and proteins.[49-55] For instance, aromatic amino acid residues can significantly increase the *cis* population of Pro residues contained in Pro-Pro-Xaa peptides[54-55] (Xaa = aromatic residue). Another study examined the structure of the hinge peptide (HP) of the human serum immunoglobulin A1 protein (IgA1), which mainly consists of Pro, Thr and Ser residues, and observed that glycosylation of a Thr or Ser residue preceding a Pro decreases the amount of *cis* Pro residues by 7–8%.[52] Similarly, the data presented in this chapter shows a significant increase in the *cis* population of Pro in non-contiguously glycosylated compounds and therefore such modifications may provide a way to introduce structural bias into oligoproline peptides.

5.5. Conclusions

Detailed MD simulations were carried out on oligoproline peptides in implicit and explicit water to examine the structural consequences of non-contiguous glycosylation that is observed in plant HRGPs. Additionally, experimental maximum molar ellipticity at 225 nm ($[\theta]_{max}$) determined by Schweizer and coworkers indicate that the degree of glycosylation influences the PPII content of oligoprolines. In particular, peptides glycosylated with a single sugar appear to be more stable in terms of $[\theta]_{max}$ than those glycosylated with a triplet of sugars. This implies that the degree of glycosylation does not necessarily correlate with PPII stability in non-contiguously glycosylated oligoprolines, therefore further studies are required. A variety of sophisticated molecular modeling techniques including MD in explicit water, ABMD in implicit water and umbrella sampling in explicit water were applied to study this problem.

The results from all methods applied show strong intramolecular hydrogen bonding between the sugar moieties and the peptide backbone. These interactions stabilize the *cis* conformation of nearby residues and thereby lower the PPII character of the peptide. The number of such hydrogen bonds increase with the number of glycosylations and correlates with the lower $[\theta]_{max}$ measured for the triply glycosylated compounds compared to those glycosylated with a single sugar. Additionally, the intramolecular hydrogen bonding were observed to be stronger in the β-glycosylated (single or triply) peptides compared to their α—glycosylated variants, which also correlates with $[\theta]_{max}$.

In conclusion, while Chapter 4 proposed that contiguous glycosylation stabilizes the *trans* conformation of oligoprolines, the present chapter shows that sugar–backbone interactions stabilize the *cis* conformations of nearby residues in non-contiguously glycosylated oligoprolines. Therefore, the position of the glycan and the degree of glycosylation are important factors in determining the stability of the PPII conformation of oligoproline peptides.

5.6. References[b]

1. Kieliszewski, M. J., The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **2001,** *57*, 319–323.

2. Kieliszewski, M. J.; Lamport, D. T. A., Extensin - repetitive motifs, functional sites, posttranslational codes and phylogeny. *Plant J.* **1994,** *5*, 157–172.

3. Kieliszewski, M. J.; Lamport, D. T. A.; Tan, L.; Cannon, M. C., Hydroxyproline-rich glycoproteins: form and function. *Annu. Plant Rev.* **2011,** *41*, 321–342.

4. Lamport, D. T. A., Structure, biosynthesis and significance of cell wall glycoproteins. *Recent Adv. Phytochem.* **1977,** *11*, 79–115.

5. E. B. Naziga, S. Bommagani, J. O'Neil, E. Lattova, F. Schweizer and S. D. Wetmore, Conformational Analysis of Non-contiguously Glycosylated Oligoprolines, In preparation. **2013**.

6. Pysh, E. S., Random-Phase Calculation of Poly-L-Proline II Circular-Dichroism. *Biopolymers* **1974,** *13*, 1563–1571.

7. Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A. F.; Creamer, T. P., Host-guest study of left-handed polyproline II helix formation. *Biochemistry* **2001,** *40*, 14376–14383.

8. Horng, J.-C.; Raines, R. T., Stereoelectronic effects on polyproline conformation. *Protein Sci.* **2006,** *15*, 74–83.

---

[b] ACS referencing style was implemented throughout this thesis.

9. Rucker, A. L.; Creamer, T. P., Polyproline II helical structure in protein unfolded states: Lysine peptides revisited. *Protein Sci.* **2002,** *11*, 980–985.

10. Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P., Host-guest scale of left-handed polyproline II helix formation. *Proteins* **2003,** *53*, 68–75.

11. Owens, N. W.; Stetefeld, J.; Lattova, E.; Schweizer, F., Contiguous O-Galactosylation of 4(R)-Hydroxy-L-proline Residues Forms Very Stable Polyproline II Helices. *J. Am. Chem. Soc.* **2010,** *132*, 5036–5042.

12. Owens, N. W.; Braun, C.; O'Neil, J. D.; Marat, K.; Schweizer, F., Effects of glycosylation of (2S,4R)-4-hydroxyproline on the conformation, kinetics, and thermodynamics of prolyl amide isomerization. *J. Am. Chem. Soc.* **2007,** *129*, 11670–11671.

13. Owens, N. W.; Lee, A.; Marat, K.; Schweizer, F., The Implications of (2S,4S)-Hydroxyproline 4-O-Glycosylation for Prolyl Amide Isomerization. *Chem.-Eur. J.* **2009,** *15*, 10649–10657.

14. Aliev, A. E.; Courtier-Murias, D., Conformational analysis of L-prolines in water. *J. Phys. Chem. B* **2007,** *111*, 14034–14042.

15. Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S., A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **2005,** *26*, 1612–1616.

16. Aliev, A. E.; Courtier-Murias, D., Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly. *J. Phys. Chem. B* **2010,** *114*, 12358–12375.

17. Moradi, M.; Babin, V.; Roland, C.; Darden, T. A.; Sagui, C., Conformations and free energy landscapes of polyproline peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 20746–20751.

18. Naziga, E. B.; Schweizer, F.; Wetmore, S. D., Solvent Interactions Stabilize the Polyproline II Conformation of Glycosylated Oligoprolines. *J. Phys. Chem. B* **2013,** *117*, 2671–2681.

19. Kang, Y. K.; Choi, H. Y., Cis-trans isomerization and puckering of proline residue. *Biophys. Chem.* **2004,** *111*, 135–142.

20. Kang, Y. K.; Jhon, J. S.; Park, H. S., Conformational preferences of proline oligopeptides. *J. Phys. Chem. B* **2006,** *110*, 17645–17655.

21. Venkatachalam, C. M.; Price, B. J.; Krimm, S., A theoretical estimate of the energy barriers between stable conformations of the proline dimer. *Biopolymers* **1975,** *14*, 1121–1132.

22. Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; R.C.Walker; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER*, 10; University of California, San Francisco: 2008.

23. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins.* **2006,** *65*, 712–725.

24. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29*, 622–655.

25. Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S., A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **2005,** *26*, 1612–1616.

26. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000,** *104*, 3712–3720.

27. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins.* **2004,** *55*, 383–394.

28. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926–935.

29. Babin, V.; Karpusenka, V.; Moradi, M.; Roland, C.; Sagui, C., Adaptively Biased Molecular Dynamics: An Umbrella Sampling Method With a Time-Dependent Potential. *Int. J. Quantum Chem.* **2009,** *109*, 3666–3678.

30. Babin, V.; Roland, C.; Sagui, C., Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **2008,** *128*,134101–134109

31. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., PPII Propensity of Multiple-Guest Amino Acids in a Proline-Rich Environment. *J. Phys. Chem. B* **2011,** *115*, 8645–8656.

32. Moradi, M.; Babin, V.; Sagui, C.; Roland, C., A Statistical Analysis of the PPII Propensity of Amino Acid Guests in Proline-Rich Peptides. *Biophys. J.* **2011,** *100*, 1083–1093.

33. Moradi, M.; Lee, J.-G.; Babin, V.; Roland, C.; Sagui, C., Free Energy and Structure of Polyproline Peptides: An Ab Initio and Classical Molecular Dynamics Investigation. *Int. J. Quantum Chem.* **2010,** *110*, 2865–2879.

34. Moradi, M.; Babin, V.; Roland, C.; Sagui, C., A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *J. Chem. Phys.* **2010,** *133*,125104–125113

35. Naziga, E. B.; Schweizer, F.; Wetmore, S. D., Conformational Study of the Hydroxyproline-O-Glycosidic Linkage: Sugar-Peptide Orientation and Prolyl Amide lsomerization in (alpha/beta) Galactosylated 4(R/S)-Hydroxyproline. *J. Phys. Chem. B* **2012,** *116*, 860–871.

36. Roux, B., The Calculation of the Potential of Mean Force using Computer-Simulations. *Comput. Phys. Commun.* **1995,** *91*, 275–282.

37. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules 1. The Method. *J. Comput. Chem.* **1992,** *13*, 1011–1021.

38. Grossfield, Alan, WHAM: the weighted histogram analysis method, version 2.0.6, http://membrane.urmc.rochester.edu/content/wham

39. Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A., Polyproline II helix conformation in a proline-rich environment: A theoretical study. *Biophys. J.* **2004,** *86*, 731–742.

40. Shoulders, M. D.; Raines, R. T., Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009,** *78*, 929–958.

41. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T., Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **2001,** *123*, 777–778.

42. DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L., Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations. *J. Am. Chem. Soc.* **2002,** *124*, 2497–2505.

43. Song, I. K.; Kang, Y. K., Conformational Preference and Cis-Trans Isomerization of 4(R)-Substituted Proline Residues. *J. Phys. Chem. B* **2006,** *110*, 1915–1927.

44. Improta, R.; Benzi, C.; Barone, V., Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* **2001,** *123*, 12568–12577.

45. Doose, S.; Neuweiler, H.; Barsch, H.; Sauer, M., Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 17400–17405.

46. Stapley, B. J.; Creamer, T. P., A survey of left-handed polyproline II helices. *Protein Sci.* **1999,** *8*, 587–595.

47. Velasquez, S. M.; Ricardi, M. M.; Dorosz, J. G.; Fernandez, P. V.; Nadra, A. D.; Pol-Fachin, L.; Egelund, J.; Gille, S.; Harholt, J.; Ciancia, M.; Verli, H.; Pauly, M.; Bacic, A.; Olsen, C. E.; Ulvskov, P.; Petersen, B. L.; Somerville, C.; Iusem, N. D.; Estevez, J. M., O-Glycosylated Cell Wall Proteins Are Essential in Root Hair Growth. *Science* **2011,** *332*, 1401–1403.

48. Cannon, M. C.; Terneus, K.; Hall, Q.; Tan, L.; Wang, Y. M.; Wegenhart, B. L.; Chen, L. W.; Lamport, D. T. A.; Chen, Y. N.; Kieliszewski, M. J., Self-assembly of the plant cell wall requires an extensin scaffold. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 2226–2231.

49. Zhang, K.; Teklebrhan, R. B.; Schreckenbach, G.; Wetmore, S.; Schweizer, F., Intramolecular Hydrogen Bond-Controlled Prolyl Amide Isomerization in Glucosyl 3'(S)-Hydroxy-5'-hydroxymethylproline Hybrids: Influence of a C-5'-Hydroxymethyl Substituent on the Thermodynamics and Kinetics of Prolyl Amide Cis/Trans Isomerization. *J. Org. Chem.* **2009,** *74*, 3735–3743.

50. Teklebrhan, R. B.; Zhang, K. D.; Schreckenbach, G.; Schweizer, F.; Wetmore, S. D., Intramolecular Hydrogen Bond-Controlled Prolyl Amide Isomerization in Glucosyl 3(S)-Hydroxy-5-hydroxymethylproline Hybrids: A Computational Study. *J. Phys. Chem. B* **2010,** *114*, 11594–11602.

51. Kuemin, M.; Nagel, Y. A.; Schweizer, S.; Monnard, F. W.; Ochsenfeld, C.; Wennemers, H., Tuning the cis/trans Conformer Ratio of Xaa-Pro Amide Bonds by Intramolecular Hydrogen Bonds: The Effect on PPII Helix Stability. *Angew. Chem.-Int. Edit.* **2010,** *49*, 6324–6327.

52. Narimatsu, Y.; Kubota, T.; Furukawa, S.; Morii, H.; Narimatsu, H.; Yamasaki, K., Effect of Glycosylation on Cis/Trans Isomerization of Prolines in IgA1-Hinge Peptide. *J. Am. Chem. Soc.* **2010,** *132*, 5548–5549.

53. Wu, W.-g.; Pasternack, L.; Huang, D.-H.; Koeller, K. M.; Lin, C.-C.; Seitz, O.; Wong, C.-H., Structural Study on O-Glycopeptides: Glycosylation-Induced Conformational Changes of O-GlcNAc, O-LacNAc, O-Sialyl-LacNAc, and O-Sialyl-Lewis-X Peptides of the Mucin Domain of MAdCAM-1. *J. Am. Chem. Soc.* **1999,** *121*, 2409–2417.

54. Ganguly, H. K.; Majumder, B.; Chattopadhyay, S.; Chakrabarti, P.; Basu, G., Direct Evidence for CH⋯π Interaction Mediated Stabilization of Pro-cisPro Bond in Peptides with Pro-Pro-Aromatic motifs. *J. Am. Chem. Soc.* **2012,** *134*, 4661–4669.

55. Zondlo, N. J., Aromatic–Proline Interactions: Electronically Tunable CH/π Interactions. *Acc. Chem. Res.* **2012,** *46*, 1039–1049.

Chapter 6. Effects of Hydroxyproline Glycosylation on Collagen Stability[a]

6.1. Introduction

As described in Chapter 1, collagen is the most abundant protein in the animal kingdom. Its presence in a variety of tissues and organs, including skin, bone, muscles and membranes,[1-2] demonstrates the versatility and importance of this class of proteins. Consequently, collagen has been implicated in many disease conditions, such as various cancers and osteogenesis imperfecta.[3-4] Since structure and function are closely related for most proteins, understanding the factors that determine collagen structural integrity is of great importance. Collagen consists of three protein (peptide) strands that are wound around each other to form a triple helix (Figure 6.1). Each strand contains a distinctive X–Y–Gly amino acid sequence repeat adopting the PPII conformation.[5-8] Pro and its modified derivative Hyp occur frequently in collagen,[9] where experimental[10-14] and computational[15-16] evidence show that Hyp, particularly in the Y position, significantly stabilizes collagen structures due to stereoelectronic effects. This is because the C^{γ}-exo pucker induced by electronegative substituents such as the OH in Hyp stabilizes the *trans* isomer required for the PPII conformation. Furthermore, the pucker orders the backbone ϕ and ψ dihedral angles to values that are optimal for the Y position in triple helix.[1]

In contrast to the HRGPs described in the previous chapters, Hyp glycosylation has not been observed in collagen. Since HRGPs are the plant analog of collagen due to their similar structural roles, it is interesting to determine the reason for the natural lack of glycosylation in animals. Given the experimental difficulties associated with crystallizing biological collagen, synthetic CMPs are usually employed to study collagen

---

[a] Reprinted in part with permission from S. Bommagani, E.B. Naziga, N. W. Owens, E. Lattová, J. D. O'Neil, S. D. Wetmore and F. Schweizer, The Effects of (2*S*,4*R*)-4-Hydroxyproline Glycosylation on the Stability of the Collagen Triple Helix, unpublished work 2013.

structure–function relationships.[17-20] Therefore, to provide a comprehensive picture of the effects of Hyp glycosylation on collagen triple helix stability, CMPs with varying degrees of glycosylation were synthesized (see Table 6.1) by Schweizer and co–workers.[21] The melting temperatures (T_m) were determined using circular dichroism (CD) spectroscopy by monitoring the molar ellipticity at 225 nm over a temperature range and fitting the data to a two state model or taking the first derivative with respect to temperature.

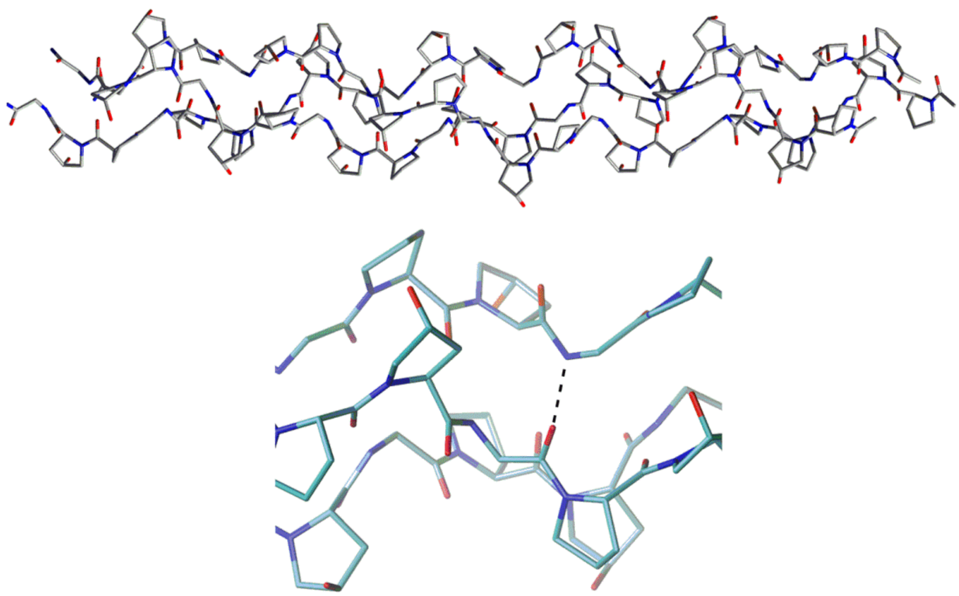

Figure 6.1. (Top) A model of the collagen triple helix of a $(POG)_7$ peptide and a closed up view (bottom) showing an interstrand backbone–backbone (C=O···H–N) hydrogen bond (dashed lines).

The molar experimentally determined ellipticity at 225 nm was monitored over a 5 to 90°C temperature range for the peptides (Table 6.1) in aqueous media.[21] The unglycosylated $(POG)_7$ peptide displayed an unfolding transition, and a melting temperature of 39.2±1°C was determined. The measured T_m is similar to a previous value of 36±1°C obtained for an uncapped $(POG)_7$ triple helix.[12] The experimental results also indicated that a single β–glycosylation per strand ($(POG)_3$–$(PO^{\beta}G)$–$(POG)_3$) decreases the melting temperature of the $(POG)_7$ CMP from 39.2±1°C to 37.1±1°C, while the triple helix no longer forms after three or more β–glycosylations per strand ($[(POG)–(PO^{β}G)]_3–(POG)$ and $(PO^{β}G)_7$). Furthermore, a single α–glycosylation per strand ($(POG)_3–(PO^{α}G)–(POG)_3$) prevents helix formation. Unfortunately, the reason for the loss of the triple helical structure upon glycosylation is not clear from this study. However, steric repulsion due to the large sugar groups was proposed to be the most likely factor.[17] To explain the above experimental findings, and by extension provide a rationalization for the lack of Hyp glycosylation in naturally occurring collagen, molecular modeling studies, including MD simulations and DFT optimizations were carried out in this Chapter. The modeling data provides an explanation for the experimental observation that glycosylation destabilizes the triple helix and rationalize the natural absence of glycosylation in animal collagen.

Table 6.1. Abbreviations for the collagen model compounds considered in this chapter.[a]

Compound	Abbreviation
$Ac–(POG)_7–NH_2$	$(POG)_7$
$Ac–(POG)_3–(PO^{α}G)–(POG)_3–NH_2$	$(POG)_3–(PO^{α}G)–(POG)_3$
$Ac–[(POG)–(PO^{β}G)]_3–(POG)–NH_2$	$[(POG)–(PO^{β}G)]_3–(POG)$
$Ac–(PO^{β}G)_7–NH_2$	$(PO^{β}G)_7$

[a]O is used to represent Hyp residues.

6.2. Computational Details

6.2.1. Molecular Dynamics Simulations

MD simulations were carried out using the AMBER suite of programs[22-23] on the following CMPs: $(POG)_7$, $(POG)_3–(PO^{α}G)–(POG)_3$, $[(POG)–(PO^{β}G)]_3–(POG)$ and $(PO^{β}G)_7$ (Figure 6.2). Initial heavy (non-hydrogen) atom conformations in the $(POG)_7$ triple helical model were taken from a crystal structure of CMP (protein data bank (PDB) entry 1CAG).[5] Hydrogen atoms were added to the experimental structure using the xLeap module of AMBER. Subsequently, three POG triplets were removed from the initial structure, the alanine substitution in the middle of the peptide was mutated to a glycine, and the C and N–terminals were capped with acetyl and amide groups, respectively. In all triple helices, strands 1, 2 and 3 therefore contain residues 1 to 23, 24 to 46 and 47 to 69, respectively.

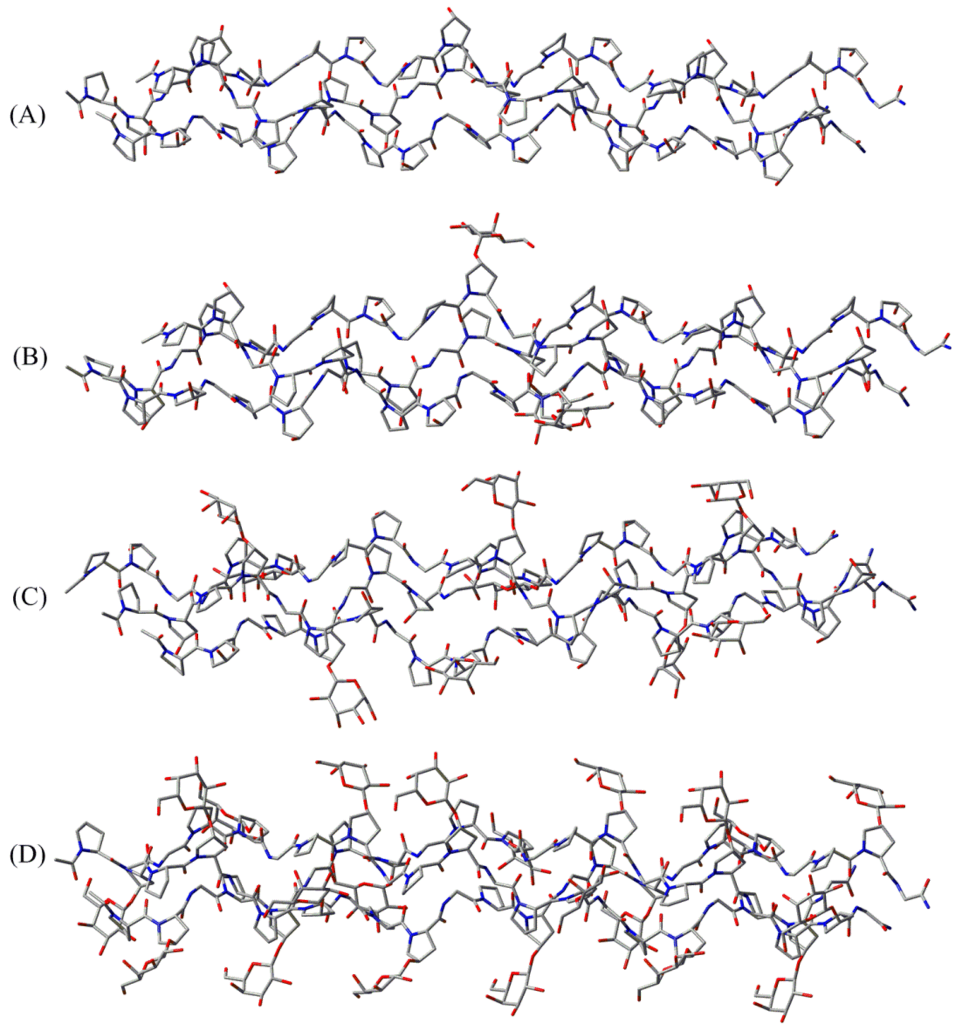

Figure 6.2. Molecular models of the triple helices considered in this chapter (A) Ac–(POG)$_7$–NH$_2$ (B) Ac–(POG)$_3$–(PO$^{\alpha}$G)–(POG)$_3$–NH$_2$ (C) Ac–[POG–(PO$^{\beta}$G)]$_3$–(POG)–NH$_2$ and (D) Ac–(PO$^{\beta}$G)$_7$–NH$_2$. (hydrogen atoms removed for clarity).

Hyp residues were mutated to glycosylated Hyp as required. The AMBER FF99SB parameter set[24] was used for the amino acid residues, while the sugar moieties were modeled with the GLYCAM06c parameters.[25] Using the xLeap module of AMBER, the initial triple helices were solvated in a rectangular box of TIP3P water[26-27] such that the minimum distance between the solute and the edge of the box was 15.0 Å. The resulting structures were subjected to two rounds of minimization. In the first round, the solute was held fixed, while the water molecules were relaxed. In the second round, ten thousand minimization steps were implemented with all backbone heavy atoms held fixed using a restraint force of 50 kcal (mol Å)$^{-1}$. The final structure from this process was slowly heated to 300 K over a 100 ps time frame, while holding the solute fixed using a weaker 10 kcal (mol Å)$^{-1}$ restraint. A second NPT equilibration step of 100 ps was then carried out to stabilize the box dimensions, and a box size of approximately 100 × 50 × 50 Å^3 was obtained in all cases. The equilibrated structures were used for 10 ns production calculations at 300 K. However, in the MD simulations of the fully glycosylated peptide ((PO$^{\beta}$G)$_7$), the long axis of the peptide diffused and approached the short end of the box at approximately 5 ns. Therefore, to ensure that the solute does not interact with its image under the periodic boundary conditions, an octahedral box of size 78 × 78 × 78 Å^3 was used to simulate the system for 19 ns.

MD simulations were also carried out on monomer strands of the CMPs with initial structures taken from one strand of the corresponding crystal structure triple helix. A previous study has shown that the monomeric form of a (POG)$_{10}$ peptide does not remain in an extended PPII conformation in solution, but rather collapses into more globular structures due to the formation of turn segments.[28] Therefore, to save computational cost, initial 100 ns simulations of the single strands were run in implicit solvent to obtain folded structures. The resulting trajectories were clustered into four groups based on the backbone conformation and a representative member of each cluster was subjected to a

Low-MODe (LMOD) minimization.[29-30] LMOD minimizations, which has been applied in protein loop optimizations,[29] utilize the low vibrational modes of a macromolecule to explore the potential energy surface and locate local minima.

The resulting monomer structures were then subjected to the same sequence of minimization and equilibration steps as implemented for the triple helices, which led to cubic boxes with sides of approximately 45 Å in all cases. Production calculations of at least 300 ns were then implemented for each peptide. For the $(POG)_7$ peptide, an additional calculation was started from a fully–extended conformation to allow for comparison to the ensemble of structures obtained from the MD simulation initiated from an implicit solvent/LMOD (or globular) minimized structure. This simulation was equilibrated as described above, which led to an octahedral box size of 81 Å on each side. A production calculation of 350 ns was subsequently conducted. All equilibration calculations employed a timestep of 1.0 fs, while a 2.0 fs timestep was used in the production stages. A non–bonded cutoff of 10 Å was used for all calculations.

The Molecular Mechanics Reference Interaction Site Model (MM–RISM) methodology[31-34] was used with the MMPBSA.py[35] script in the AMBER suite of programs to estimate single strand (M) to triple helix (TH) molecular association energetics as:

$$\Delta G = \frac{\Delta G_{TH}}{3} - \Delta G_M$$

The free energy of the triple helical state was divided by three to average the energies of the three strands in the triple helix and subsequently the energies of the monomer state was subtracted. Averages were computed using 400 and 800 snapshots for the triple helices and monomeric strands, respectively.

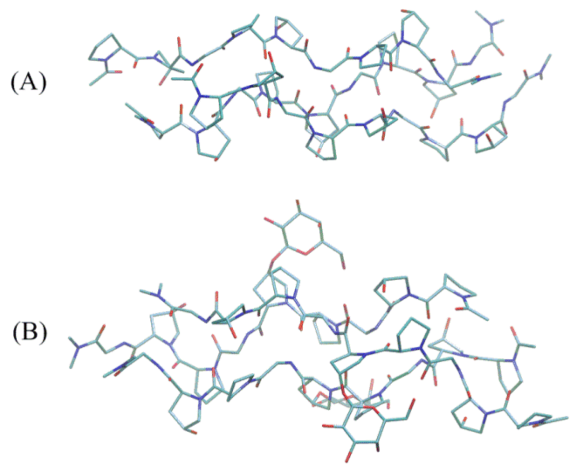

Figure 6.3. PCM–B3LYP/6–31G(d) optimized structure of (A) $(POG)_3$ and (B) (POG)–($PO^{\beta}G$)–(POG) (hydrogen atoms removed for clarity).

6.2.2. DFT Calculations

DFT calculations on triple helix models consisting of 27 amino acids were carried out using Gaussian 09.[36] Specifically, the $(POG)_3$ and (POG)–($PO^{\beta}G$)–(POG) models (Figure 6.3) were composed of 351 and 432 atoms, respectively. The ends of each monomer strand were capped with methyl groups to reduce unphysical hydrogen–bond formation between the end groups. The B3LYP density functional was used with the 6–31G(d) basis set for the geometry optimizations. Solvent (water) was included in the optimization step using the (implicit) IEF–PCM[37] methodology (ε = 78.3). Initial structures for the DFT calculations were taken from arbitrary MD snapshots. Given the size of the basis set implemented and the close packing of residues observed in the triple helix, the solvent–phase interaction energies were corrected for basis set superposition error (BSSE) using a counterpoise methodology.[38-40] Specifically, a solvent correction, $\delta_{sol} = \Delta E_{PCM} - \Delta E_{gas}$, was added to the gas–phase counterpoise corrected interaction energy (ΔE_{cp_gas}) to yield the solvent–phase interaction energy (ΔE_{sol}): $\Delta E_{sol} = \Delta E_{cp\_gas} + \delta_{sol}$, where ΔE_{PCM} and ΔE_{gas} are the PCM and gas–

phase interaction energies, respectively. This methodology has been previously applied in the literature to estimate solvent–phase counterpoise corrected interaction energies.[41]

6.3. Results

6.3.1. Structure of the Triple Helical Collagen Model Peptides

A collagen triple helix is composed of three polyproline II strands coiled about each other and held together by intermolecular C=O···H–N hydrogen bonding, as well as other non–covalent interactions.[6,8] To investigate the effect of glycosylation on the structure of model collagen peptides, structural information from MD simulations in explicit solvent, as well as implicit solvent B3LYP geometry optimization, was obtained.

First, the values of the backbone ϕ and ψ dihedral angles were examined, which are vital for establishing the PPII structure of the individual strands and the intermolecular hydrogen bonds that are necessary for triple helix stability. To remove bias due to the flexibility in the end residues, only the middle five triplets of each strand will be analyzed. The average values of the ϕ and ψ dihedral angles (Figures 6.4 and 6.5) cluster around the PPII region of the Ramachandran plot in all CMPs, which is characteristic of the collagen triple helix.[5] This finding indicates that the backbone conformations of the monomers in the triple helix are not severely altered upon glycosylation, even with seven sugars per strand. This is particularly interesting since experimental data shows multiple glycosylation is not conducive to triple helix formation.[21]

Further structural information was obtained using electronic structure (B3LYP) method. While it has been shown that over five triplets per strand are required for a stable triple helix,[42-44] experimental trends in stability have been reproduced using triple helix models composed of smaller oligomers, such as (Pro–Hyp–Gly)$_2$, (Pro–Pro–Gly)$_2$ and (Pro–Pro–Ala)(Pro–Pro–Gly).[40,45] Therefore, the (POG)$_3$ and (POG)–(PO$^{\beta}$G)–(POG) models consisting of three triplets per strand (a total of 27 residues per model) were used in this study. To my knowledge, these models are larger than those implemented in similar DFT studies in the literature which contained 18 amino acid residues.[40,45-46]

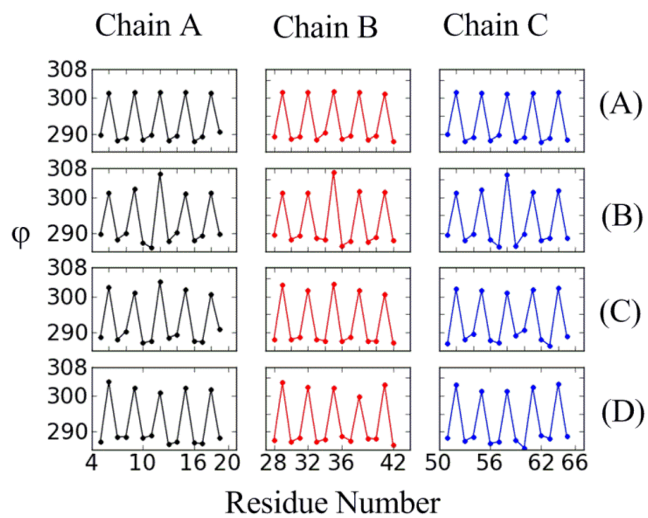

Figure 6.4. Average values for the φ dihedral angles obtained from MD simulations for the (A) (POG)$_7$, (B) (POG)$_3$–(PO$^{\alpha}$G)–(POG)$_3$, (C) [(POG)$_3$–(PO$^{\beta}$G)]$_3$–(POG) and (D) (PO$^{\beta}$G)$_7$ triple helices.

Comparison of the average B3LYP backbone (φ and ψ) dihedral angles (Table 6.2) for the (POG)$_3$ triple helix to available experimental structural data suggests that B3LYP accurately models the structure of this CMP. Specifically, a maximum deviation of approximately 5° is observed in the average φ dihedral angle of the Hyp residues of (POG)$_3$ relative to the experimental average.[5] The B3LYP averages (Table 6.2) are also in line with those from MD (Figures 6.4 and 6.5). Although an experimental crystal structure is not available for the (POG)–(PO$^{\beta}$G)–(POG) triple helix, average φ and ψ dihedral angles of this glycosylated variant do not significantly deviate from the values for the (POG)$_3$ model, which is in agreement with the conclusion from MD. Thus, there is overall agreement regarding the structure of the CMPs between experimental, force field and DFT data, which means that molecular modeling (MM and DFT) results can be used to explain the experimentally–observed properties of these model collagen compounds.

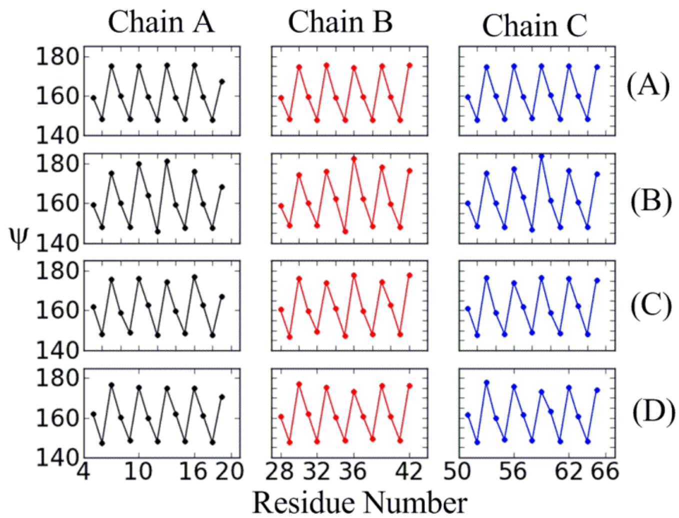

Figure 6.5. Average values for the ψ dihedral angles obtained from MD simulations for the (A) POG_7, (B) $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, (C) $[(POG)_3$–$(PO^{\beta}G)]_3$–(POG) and (D) $(PO^{\beta}G)_7$ triple helices.

Table 6.2. Average backbone dihedral angles in the $(POG)_3$ and the $(PO^{\beta}G)_3$ model triple helices with various conformations of the β–glycosylated Hyp residue (O^{β}).[a]

		Pro				Hyp				Hyp–β–Galactose				Gly			
		φ		ψ		φ		ψ		φ		ψ		φ		ψ	
		Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
POG	Exp.[b]	287.4	1.6	163.7	1.4	302.6	1.7	151.8	1.4					287.5	2.1	174.2	1.5
	Calc.	284.7	4.3	162.1	7.0	297.9	2.3	149.6	8.0					290.6	2.0	170.2	5.3
	Diff.[c]	−2.7		−1.6		−4.7		−2.2						3.1		−4.0	
$PO^{\beta}G_{180}$[d]	Calc.	284.1	2.0	167.5	1.0	299.7	2.3	152.6	2.8	299.9	1.4	153.0	1.0	290.4	0.7	173.4	0.9
$PO^{\beta}G_{210}$[e]	Calc.	283.4	2.8	164.7	2.7	299.6	3.5	153.0	4.1	296.3	0.5	153.5	0.8	290.1	1.3	171.0	5.3
$PO^{\beta}G_{210-h}$[f]	Calc.	283.2	3.3	164.9	3.2	299.0	2.4	152.8	2.9	299.1	3.4	153.0	0.7	290.7	1.1	169.7	5.0
$PO^{\beta}G_{275}$[g]	Calc.	285.3	2.3	165.9	0.9	299.6	2.7	153.0	4.1	303.1	0.7	151.3	3.1	299.0	1.3	175.8	3.4
$PO^{\beta}G_{\varphi g50}$[h]	Calc.	282.2	2.0	167.3	0.5	299.9	3.0	153.4	3.5	293.6	1.2	156.5	0.8	289.6	1.0	173.3	4.0

[a]Geometry optimizations were carried out at the PCM–B3LYP/6–31G(d) level of theory. All dihedral angles are reported in degrees (°). [b]Experimental structure from Bella et al.[5] [c]Difference between experimental and calculated averages. [d]$PO^{\beta}G_{180}$ refers to the conformation with a glycosidic linkage $\psi_g = 180°$. [e]$PO^{\beta}G_{210}$ refers to the conformation with a glycosidic linkage $\psi_g = 210°$. [f]$PO^{\beta}G_{210-h}$ refers to the conformation with a glycosidic linkage $\psi_g = 210°$ and an intramolecular hydrogen bond between sugar and backbone. [g]$PO^{\beta}G_{275}$ refers to the conformation with a glycosidic linkage $\psi_g = 275°$. [h]$PO^{\beta}G_{\varphi g50}$ refers to the conformation with a glycosidic linkage $\varphi_g = 50°$.

Since interstrand hydrogen bonding is the primary attractive interaction responsible for the formation of the collagen triple helix,[5,8] it is intriguing to determine whether glycosylation alters these hydrogen bonds. A list of the backbone C=O⋯H–N hydrogen bonds present in the MD simulations on $(POG)_7$, $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, [(POG)–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ (Tables 6.3 – 6.6) show an occupancy rate greater than 80% in all cases, which indicates that these interactions are not adversely affected by glycosylation. However, new intrastrand hydrogen–bonding interactions are introduced upon inclusion of the sugar moiety. Specifically, for the ([(POG)–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$) compounds, new intrastrand interactions mainly arise between the O6 sugar hydrogen of the glycosylated Hyp residue (H6O) and the backbone carbonyl oxygen of Gly (O) two residues down on the same strand (i+2) that is not involved in interstrand hydrogen bonding. Nevertheless, no new interstrand hydrogen–bonding interactions arise due to the sugar moiety in the β-glycosylated models. Interestingly, additional interstrand hydrogen bonding is observed in $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$ between the sugar O6 hydroxyl of a glycosylated Hyp residue and either another glycosylated Hyp or an unglycosylated Hyp residue. Indeed, three such hydrogen bonds occur with occupancies of approximately 25%, 35% and 45% in a 10 ns MD calculation.

Analysis of the conformational space available to the covalently–linked sugar moiety was considered by monitoring the glycosidic-linkage dihedral angles (φ_g and ψ_g) which revealed additional, significant differences between the α and all β–glycosylated peptides (Figure 6.6). Specifically, β–glycosylation yields greater conformational freedom about ψ_g, with minima adopted at approximately 180°, 210° and 270°. In contrast, there is considerable restriction in the conformational space of the sugar in $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, where ψ_g shows a single minimum at approximately 250°. This is not unexpected since the preferred ψ_g in α–glycosylated Ac–[Hyp–(α–Gal)]–OMe is a stacked orientation of the sugar and peptide (Chapter 3), which would lead to severe steric clashes in the triple helix. Furthermore, comparing the isolated Ac–[Hyp–(α/β–Gal)]–OMe compounds to those in the triple helix shows that an additional minimum with respect to ψ_g is introduced in the β-glycosylated peptides, while there is a reduction from two to one minimum in the α-glycosylated peptide upon triple helix formation. These differences could only arise from steric constrains in the CMPs compared to the isolated Ac–[Hyp–(α/β–Gal)]–OMe compounds.

Table 6.3. Interstrand hydrogen-bonding interactions observed in the triple helix of Ac–$(POG)_7$–NH_2 according to MD simulations in explicit water.[a]

Accepter	Donor	% occupancy
:11@O	:59@N	98
:51@O	:30@N	98
:34@O	:10@N	98
:5@O	:53@N	98
:63@O	:42@N	98
:54@O	:33@N	98
:14@O	:62@N	98
:57@O	:36@N	98
:60@O	:39@N	98
:40@O	:16@N	98
:37@O	:13@N	98
:8@O	:56@N	98
:31@O	:7@N	97
:43@O	:19@N	97
:48@O	:27@N	97
:2@O	:50@N	97
:28@O	:4@N	97
:17@O	:65@N	96
:66@O	:45@N	96
:20@O	:68@N	89
:68@O	:22@N	19

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater.

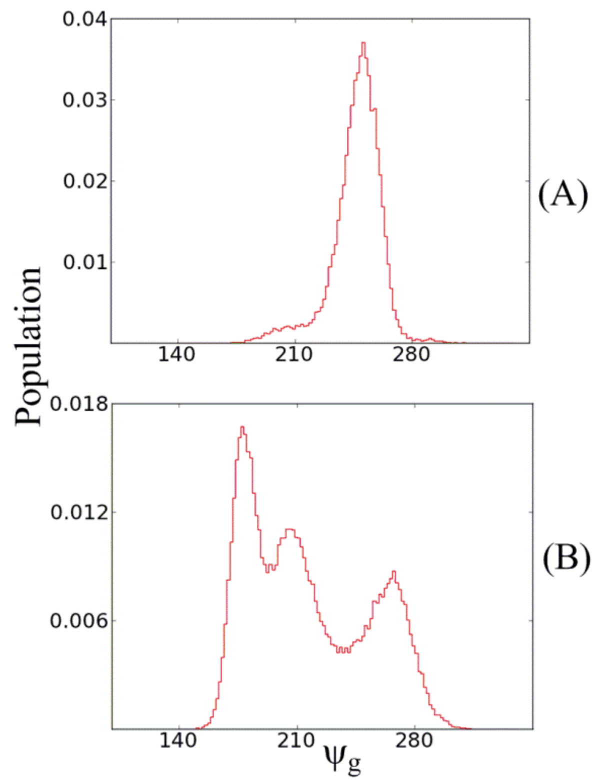

Figure 6.6. Histograms of the ψ_g dihedral angle for the glycosylated Hyp obtained from MD simulations (A) $(POG)_3$–$(PO^{\alpha}G)$ –$(POG)_3$ and (B) $(PO^{\beta}G)_7$ triple helices.

Table 6.4. Interstrand and intrastrand (sugar backbone) hydrogen-bonding interactions observed in the triple helix of Ac–$(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$–NH_2 according to MD simulations in explicit water.[a]

Acceptor	Donor	% occupancy
:31@O	:7@N	98
:5@O	:53@N	98
:8@O	:56@N	98
:37@O	:13@N	98
:63@O	:42@N	98
:34@O	:10@N	98
:51@O	:30@N	97
:14@O	:62@N	97
:40@O	:16@N	97
:28@O	:4@N	97
:60@O	:39@N	97
:11@O	:59@N	97
:43@O	:19@N	96
:54@O	:33@N	96
:57@O	:36@N	96
:20@O	:68@N	96

:17@O	:65@N	96
:48@O	:27@N	95
:2@O	:50@N	94
:66@O	:45@N	94
:4@O	:6@O6	68
:42@O	:44@O6	55
:30@O	:32@O6	50
:13@O	:15@O6	45
:19@O	:21@O6	44
:65@O	:67@O6	44
:62@O	:64@O6	37
:33@O	:35@O6	34
:53@O	:55@O6	32
:36@O	:38@O6	30
:50@O	:52@O6	27
:56@O	:58@O6	26
:68@O	:22@N	25
:16@O	:18@O6	25
:39@O	:41@O6	22
:27@O	:29@O6	22
:1@O	:3@O6	20
:61@O	:38@O2	20
:7@O	:9@O6	18
:10@O	:12@O6	16
:24@O	:26@O6	14
:47@O	:49@O6	14
:64@O6	:38@O3	14
:49@O	:26@O2	13
:18@O6	:18@O4	13
:38@O	:12@O6	11
:12@O	:58@O2	10
:49@O	:26@O6	10
:9@O	:55@O6	9
:52@O	:29@O2	7
:50@N	:27@N	6
:30@N	:4@N	6
:62@N	:39@N	6
:29@O	:3@O2	6
:15@O	:61@O2	6
:39@N	:13@N	5
:36@N	:10@N	5

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater.

Table 6.5. Interstrand and intrastrand (sugar backbone) hydrogen-bonding interactions observed in the triple helix of Ac–[(POG)–(PO$^{\beta}$G)]$_3$-(POG)–NH$_2$ according to MD simulations in explicit water.[a]

Acceptor	Donor	% occupancy
:8@O	:56@N	98
:60@O	:39@N	98
:37@O	:13@N	98
:11@O	:59@N	98
:14@O	:62@N	98
:5@O	:53@N	98
:57@O	:36@N	98
:20@O	:68@N	98
:43@O	:19@N	97
:51@O	:30@N	97
:40@O	:16@N	97
:17@O	:65@N	97
:48@O	:27@N	97
:31@O	:7@N	97
:54@O	:33@N	97
:63@O	:42@N	97
:28@O	:4@N	97
:34@O	:10@N	97
:2@O	:50@N	96
:66@O	:45@N	92
:56@O	:58@O6	63
:50@O	:52@O6	52
:39@O	:41@O6	38
:62@O	:64@O6	27
:4@O	:6@O6	21
:16@O	:18@O6	21
:68@O	:45@N	17
:27@O	:29@O6	12
:18@O4	:18@O6	5
:52@O	:29@O2	5

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater.

Table 6.6. Interstrand and intrastrand (sugar backbone) hydrogen-bonding interactions observed in the triple helix of Ac–(PO$^{\beta}$G)$_7$–NH$_2$ according to MD simulations in explicit water.[a]

Acceptor	Donor	% occupancy
:60@O	:39@N	98
:31@O	:7@N	98
:34@O	:10@N	98
:40@O	:16@N	98
:14@O	:62@N	98
:8@O	:56@N	98
:54@O	:33@N	97

:5@O	:53@N	97
:28@O	:4@N	97
:43@O	:19@N	97
:57@O	:36@N	97
:63@O	:42@N	97
:17@O	:65@N	97
:48@O	:27@N	97
:51@O	:30@N	97
:2@O	:50@N	97
:37@O	:13@N	97
:11@O	:59@N	96
:66@O	:45@N	96
:20@O	:68@N	96
:58@O	:35@O6	37
:12@O	:58@O6	30
:38@O	:12@O6	20
:68@O	:22@N	13
:13@N	:59@N	10
:12@O4	:12@O6	8
:39@N	:13@N	8
:58@O4	:58@O6	7
:59@N	:36@N	7
:12@O6	:12@O4	6
:36@N	:10@N	5

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater.

In summary, the results from MD simulations of the triple helix suggests that glycosylation of the triple helix may not be precluded since the triple helical structure and associated intermolecular (N–H···O=C) hydrogen bonding is preserved in the $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, $[(POG)–(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ CMPs for the duration of the MD simulations. Additionally, new hydrogen–bonding interactions arise between the sugar moieties and the peptide backbone in all glycosylated compounds that have significant (>20%) occupancies. Nevertheless, these observations do not explain the experimentally observed destabilization effects of glycosylation on the triple helix, which suggests that the individual CMP strands must be considered.

6.3.2. Structure of the Monomeric Collagen Model Compounds

Since the MD data for the triple helices does not provide an obvious explanation for the experimentally observed triple helix destabilization upon glycosylation, the conformations of the monomeric strands are studied in this subsection. Indeed, previous investigations[47-50] have reported that N(O)-glycosylation can alter the conformational ensemble available to model glycopeptides and often acts as a conformational switch.[51-53] Therefore, MD simulations on the $(POG)_7$, $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, $[(POG)$–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ monomer strands were carried out.

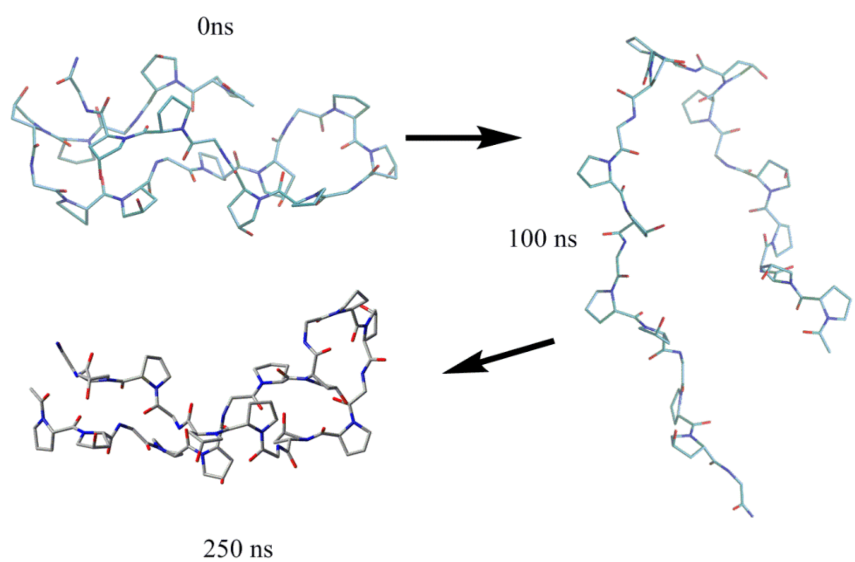

Figure 6.7. Structure of the $(POG)_7$ peptide at various points during the MD simulation in explicit solvent.

The radius of gyration (R_g) based on the backbone C_{α} atoms was used to characterize the structural ensemble of the peptides since this variable adequately describes the level of compactness or 'folding'. R_g adopts a value of approximately 7.8 Å in the initial globular structure of $(POG)_7$ (Figure 6.7 and 6.8), which increases to approximately 12.0 Å during the first 100 ns of the simulation, and corresponds to a partially unwound structure. However, R_g assumes a value of approximately 6.3 Å after approximately 120 ns and remains close to this value for the reminder of the 337 ns simulation. The ensemble of structures obtained during this stable period (120 - 337 ns) exhibit a turn segment from Pro8 to Pro14, with residues 1 to 7 and 16 to 23 exhibiting a parallel conformation (Figure 6.7).

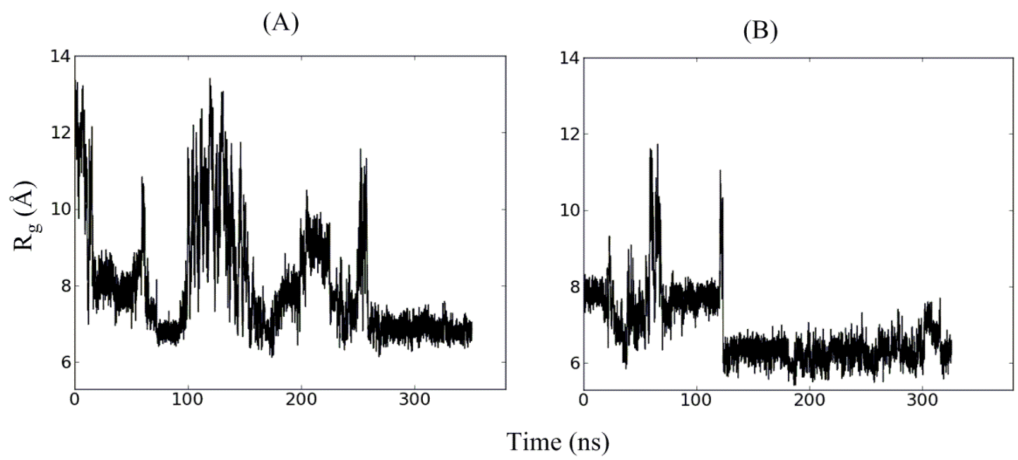

Figure 6.8. Radius of gyrations (R_g) from MD simulations on a monomeric strand of the unglycosylated $(POG)_7$ CMP obtained starting from (A) an extended conformation and (B) a globular structure.

To determine the effect of the starting structure, MD simulations on the $(POG)_7$ peptide starting from an extended conformation were also carried out as described in the Computational Details. During the first 250 ns of the 350 ns simulation, the peptide unfolds several times (Figure 6.8). R_g decreases from over 13 Å, which represents an extended conformation to a more compact structure with $R_g \sim 7$ Å after approximately 100 ns. R_g subsequently oscillates between 7 and 13 Å over the next 60 ns before finding a more compact conformation. The peptide then partially unfolds ($R_g < 10$ Å) over the next 100 ns. However, in the last 100 ns of the MD simulation, R_g stabilizes to approximately 6.8 Å, which is similar to the value obtained from the simulation using the globular starting structure.

This suggests that comparable equilibrated structural ensembles are obtained regardless of whether a globular or extended starting point is used.

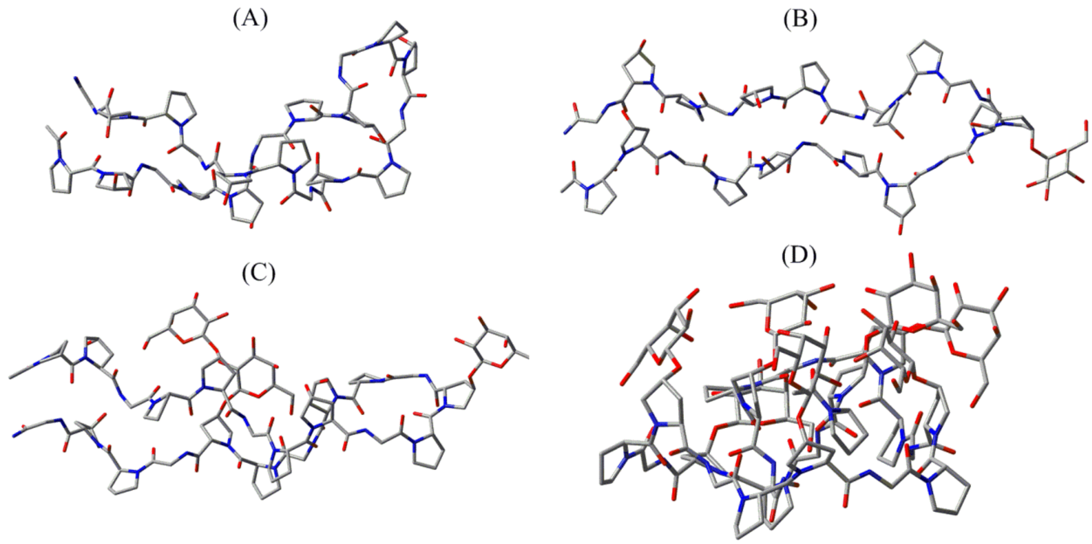

Figure 6.9. Representative structures obtained from MD simulations on the (A) POG_7, (B) [(POG)–(PO$^{\alpha}$G)]$_3$–(POG), (C) [(POG)–(PO$^{\beta}$G)]$_3$–(POG) and (D) (PO$^{\beta}$G)$_7$ monomers.

Additional support that both starting structures lead to similar conformations is obtained from analysis of the intramolecular hydrogen bonding. Specifically, while the occupancies vary, most interactions are present in both simulations. For instance, a hydrogen bond between the carbonyl oxygen of residue 5 and the NH group of residue 19 occurs for 63% of the simulation starting from the extended structure, and 53% of the simulation starting from the globular structure. Similarly, an interaction between the NH group of residue 7 and the carbonyl of residue 17 takes place for 45% and 25% of the calculation initiated from extended and globular structures, respectively. This implies that a similar structural ensemble is sampled irrespective of the starting structure. Thus, the MD structures obtained represent the actual conformational ensemble of the peptide to within the accuracy of the applied force field. Furthermore, this analysis suggests that globular structures obtained in implicit solvent can be used as input for explicit solvent MD for glycosylated monomer strands.

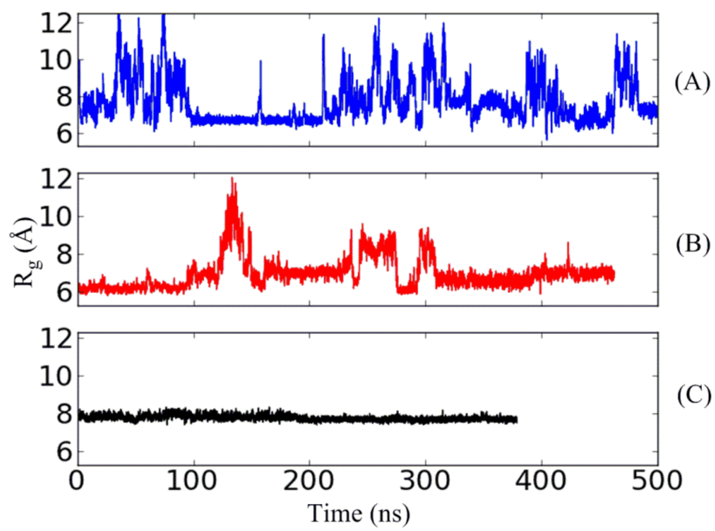

Figure 6.10. Radius of gyrations (R_g) for the monomeric strand of (A) $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, (B) $[(POG)$–$(PO^{\beta}G)]_3$–(POG) (red) and (C) $(PO^{\beta}G)_7$ monomers obtained from MD simulations.

R_g for the $[(POG)$–$(PO^{\beta}G)]_3$–(POG) monomer strand remains close to the starting value of 6–7 Å for approximately 120 ns (Figures 6.9 and 6.10). Subsequently, the peptide undergoes a rearrangement, but quickly refolds to a structure with R_g near the initial value. The structures occurring in the simulation from approximately 170 to 230 ns, 405 to 418 ns and 425 to 465 ns exhibit a turn fragment from Pro11 to Pro14, have several intramolecular hydrogen bonds (Table 6.7), and represent the lowest energy conformers for this peptide (Figure 6.9C). Interestingly, one glycosylated Hyp residue is embedded in the turn segment and directed away from the peptide into solution, which reduces unfavourable steric interactions. Additionally, the two other glycosylated residues are positioned to interact with each other via favorable hydrogen–bonding contacts. The set of structures found between approximately 310 and 380 ns have R_g similar to the aforementioned structures, but do not contain any intramolecular interactions and appear to be an intermediate for the formation of the aforementioned structures. Minimization of the representative structure from this set gives an approximately 13 kcal mol^{-1} higher conformation energy compared to the minimized structure of the representative from the 425 to 465 ns range. Most importantly, the observed intramolecular hydrogen bonding could stabilize the monomeric strand of the glycosylated collagen. Indeed, several such interactions exist for up to 70% of the MD simulation time (Table 6.7).

Table 6.7. Hydrogen-bonding interactions observed in the monomer strand of compound Ac–[(POG)–(PO$^{\beta}$G)]$_3$-(POG)–NH$_2$ according to MD simulations in explicit water.[a]

Acceptor	Donor	% occupancy
:18@O2	:13@N	71
:11@O	:19@N	50
:10@O	:12@O6	45
:8@O	:22@N	40
:8@O	:21@OD1	26
:9@O	:22@N	19
:18@O6	:10@N	18
:21@OD1	:10@N	15
:12@O4	:18@O6	15
:4@O	:6@O6	14
:9@O	:12@O6	10
:10@O	:18@O6	9
:13@O	:18@O3	6
:3@O	:6@O6	5
:14@O	:18@O2	5
:6@O6	:6@O4	5
:18@O6	:12@O3	5
:10@O	:12@O4	5

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater.

Table 6.8. Hydrogen-bonding interactions observed in the monomer strand of compound Ac–(PO$^{\beta}$G)$_7$–NH$_2$ according to MD simulations in explicit water.[a]

Acceptor	Donor	% occupancy
:21@O2	:13@N	86
:16@O	:3@O6	85
:13@O	:21@O3	74
:20@O	:4@N	68
:21@O	:10@N	68
:10@O	:12@O6	65
:21@O6	:12@O2	49
:1@O	:22@N	41
:7@O	:9@O6	37
:2@O	:22@N	34
:15@O2	:21@O4	32
:3@O2	:19@N	30
:1@O	:21@O2	29
:15@O2	:3@O4	24
:21@O3	:15@O2	22
:3@O6	:18@O2	19
:21@O4	:3@O3	16
:1@O	:13@N	14
:21@O4	:3@O4	14
:19@O	:3@O2	13
:3@O4	:15@O2	13
:13@O	:21@O2	11
:14@O	:21@O2	11
:9@O2	:12@O4	11
:3@O6	:16@N	10
:16@O	:18@O6	9
:9@O4	:6@O2	9
:12@O6	:10@N	9
:9@O6	:7@N	9
:6@O2	:9@O4	8
:19@O	:4@N	8
:14@O	:21@O3	7
:15@O3	:21@O4	7
:13@O	:15@O6	6
:6@O6	:6@O4	5
:4@O	:6@O6	5
:21@O4	:15@O2	5
:5@O	:23@N	5
:17@O	:19@N	4
:18@O	:6@O4	4
:12@O2	:21@O6	4

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater.

The $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$ monomer strand is more dynamical in R_g than the β–glycosylated compounds. Starting at approximately 7 Å, R_g rises to over 10 Å several times in the first 70 ns of the simulation (Figure 6.9). After 100 ns, an ensemble of structures with R_g equal to approximately 7 Å is observed. These structures show the formation of a turn fragment from Pro11 to Pro14, as observed for the other monomer strands, but the parallel sheet–like conformation is not present. Furthermore, the strand subsequently unravels and forms the aforementioned parallel conformation in the last 50 ns of the simulation. As observed in the $[(POG)$–$(PO^{\beta}G)]_3$–(POG) monomer, the sugar moiety is embedded in the turn segment and directed into the solvent.

The $(PO^{\beta}G)_7$ monomer strand with seven glycosylations shows remarkable stability in the folded conformation. Specifically, R_g oscillates around the starting (folded) value of approximately 8.0 Å for the duration of a 400 ns MD simulation (Figures 6.9 and 6.10). Hydrogen–bonding analysis (Table 6.8) identifies many hydrogen bonds that exist for over half of the simulation time in the folded structure with occupancies of up to 86%. Furthermore, all sugar moieties are directed towards the solvent, which leads to maximum solvation of the sugar hydroxyl groups. Therefore, a combination of intramolecular and intermolecular interactions, including sugar–solvent contacts, yields a very stable fold in this glycosylated peptide.

In summary, the analysis in this section suggests that glycosylation alters the conformation of the monomeric units of the CMPs; stabilizing folded structures. It further appears that the conformational stability of the folded monomers increases with the number of attached glycans. These findings are in agree with several computational and experimental studies showing that glycosylation modulates the structural stability of peptides and proteins.[47-48,54]

5.3.3. Energetics of Triple Helix Association

To determine how stabilization of the monomer strands affects the trimerization thermodynamics, the energetics of monomer to triple helix formation were evaluated based on PCM–B3LYP/6–31G(d) single–point calculations (Table 6.9) and a molecular mechanics MM–RISM analysis (Table 5.10). PCM–B3LYP/6–31G(d) interaction energies for the truncated $(POG)_3$ model and various conformations about the glycosidic linkage ψ_g dihedral angle in the (POG)–($PO^{\beta}G$)–(POG) peptide (Figure 6.6) suggest that a triplet of sugars destabilizes the unsubstituted triple helix. Specifically, the gas–phase BSSE–corrected interaction energies for triple–helix formation with the β–glycosylated compound are weaker by up to −11 kJ mol^{-1} for all conformations relative to the unglycosylated model. The one exception is the conformation with a glycosidic ψ_g dihedral angle of approximately 50°, which leads to a more stable triple helix than the unglycosylated compound. The fact that the conformation with a glycosidic $\psi_g \approx$ 50° is more stable than the unmodified peptide in the gas phase is interesting since previous work (Chapter 3) on Ac–[Hyp–(α/β–Gal)]–OMe[55] shows that this conformation is higher in energy than the other minimum ($\psi_g \approx$ 275°). However, when $\psi_g \approx 50°$ in the CMPs considered in this Chapter, the sugar points away from the peptide and hence reduces unfavourable steric interactions in the triple helix. This indicates that steric repulsion due to the bulky sugar moiety in part destabilizes the triple helix.

When implicit solvent is taken into account in the B3LYP calculations, the unglycosylated compound becomes more stable than all conformations of the glycosylated compound, which indicates that solvation also plays an important role in triple helix stability. Furthermore, the trend in solvent–phase binding energies correlates with experimental findings that show a drop in melting temperature from 39.2±1°C to 37.1±1°C upon glycosylation with one sugar per strand.[21]

Table 6.8. PCM–B3LYP/6–31G(d) energies for triple helix formation.[a]

Peptide/ Conformation	ΔE_{cp_gas}[b]	ΔE_{sol}[c]
POG	–35.9	–62.7
PO$^{\beta}$G$_{180}$[d]	–30.9	–40.1
PO$^{\beta}$G$_{210}$[e]	–35.5	–47.2
PO$^{\beta}$G$_{210\text{–}h}$[f]	–31.2	–46.0
PO$^{\beta}$G$_{275}$[g]	–24.7	–43.4
PO$^{\beta}$G$_{\varphi g50}$[h]	–42.2	–51.4

[a]Energies are in units of kJ mol^{-1}. [b]Counterpoise–corrected gas–phase interaction energy. [c]Counterpoise–corrected solvent–phase interaction energy. [d]PO$^{\beta}$G $_{180}$ refers to the O$^{\beta}$ conformation with a glycosidic linkage ψ_g = 180°. [e]PO$^{\beta}$G $_{210}$ refers to the O$^{\beta}$ conformation with a glycosidic linkage ψ_g = 210°. [f]PO$^{\beta}$G $_{210\text{–}h}$ refers to the O$^{\beta}$ conformation with a glycosidic linkage ψ_g = 210° and an intramolecular hydrogen bond between sugar and backbone. [g]PO$^{\beta}$G $_{275}$ refers to the O$^{\beta}$ conformation with a glycosidic linkage ψ_g = 275°. [h]PO$^{\beta}$G $_{\varphi g50}$ refers to the conformation with a glycosidic linkage ψ_g = 50°.

To further explore the origins of the preferential stability of the monomer strand in larger models, an MM–RISM analysis was applied to the $(POG)_7$, $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, $[(POG)$–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ CMPs (Table 6.9). In the unglycosylated peptide, ΔE_{elec} is very favourable since backbone intermolecular hydrogen bonding leads to a large electrostatic energy for self association. In addition, ΔE_{vdw} is negative as van der Waals interactions are enhanced by the close packing of the collagen triple helix. In contrast, ΔE_{sol} is positive and offsets the former two components. This is expected given the hydrophobic nature of many of the residues in the peptide. Addition of the vibrational entropy change between the monomeric and triple helical states yields a total free energy of –45.1 kJ mol^{-1}, which compares favourably with the value of –51.8 kJ mol^{-1} obtained experimentally[56] and provides support for applying of this approach to the glycosylated variants.

In the [(POG)–(PO$^{\beta}$G)]$_3$–(POG) CMP, the ΔE_{elec} component of the interaction energy is more positive (–130.4 kcal mol^{-1}) relative to (POG)$_7$ (–173.1 kcal mol^{-1}), which suggests unfavorable steric interactions in the triple helix or more favourable intramolecular interactions in the monomer. In contrast, ΔE_{vdw} and ΔE_{sol} are similar in both compounds. Addition of the vibrational entropy change leads to a positive total free energy of triple helix formation for [(POG)–(PO$^{\beta}$G)]$_3$–(POG), implying that a triple helix is not viable upon glycosylations with three sugars per collagen strand. This result agrees with experimental findings where triple helix formation does not occur for this peptide.

Table 6.10. Differences in the energy components of the triple helical and monomeric states for the (POG)$_7$, (POG)$_3$–(PO$^{\alpha}$G)–(POG)$_3$, [(POG)–(PO$^{\beta}$G)]$_3$–(POG) and (PO$^{\beta}$G)$_7$ CMPs.[a]

Component	POG$_7$	[(POG)–(PO$^{\beta}$G)]$_3$–(POG)	(PO$^{\beta}$G)$_7$	(POG)$_3$–(PO$^{\alpha}$G)–(POG)$_3$
ΔE_{elec}[b]	–173.2	–130.5	28.0	–210.5
ΔE_{vdw}[c]	–166.5	–172.0	–31.8	–200.8
ΔE_{int}[d]	–31.8	–22.2	–41.8	–29.7
ΔE_{pol}[e]	201.7	89.5	–3.3	295.8
ΔE_{apol}[f]	76.1	178.2	21.8	31.4
ΔE_{tot}[g.h]	–119.2	–57.3	–26.4	–113.4
–TΔS	74.1	77.8	40.6	88.7
ΔG	–45.2	21.8	14.2	–24.7

[a]All energies are obtained from the MM–RISM approach and are in units of kJ mol^{-1}. [b]ΔE_{elec} is the force field electrostatic energy. [c]ΔE_{vdw} is the force field van der Waals energy. [d]ΔE_{int} is the force field internal energy (bonds, angles and dihedrals). [e]ΔE_{pol} is the polar solvation energy calculated using 3D–RISM. [f]ΔE_{apol} is the non-polar solvation energy calculated using 3D–RISM. [g]ΔE_{tot} = (ΔE_{elec}+ΔE_{vdw}+ΔE_{int}+ΔE_{pol}+ΔE_{apol}). [h]A maximum 7.5 kJ mol^{-1} standard error of the mean was obtained for the calculation of ΔE_{tot} for all peptides.

The (POG)$_3$–(PO$^{\alpha}$G)–(POG)$_3$ monomer is not as stable as the unglycosylated monomer in terms of the electrostatic and van der Waals energies, which leads to a lower ΔE_{elec} and ΔE_{vdw} in this compound compared to (POG)$_7$. However, both compounds have similar total solvation energies. Furthermore, adding the vibrational entropy correction to the total energy yields a triple helix formation free energy of –24.7 kJ mol^{-1}, which is 20.5kJ mol^{-1} higher than for the unglycosylated peptide. This indicates that the inability of the α-glycosylated monomer to form a triple helix may be due to entropic effects. Entropy can be partitioned into rotational, translational, vibrational and configurational components in single molecule calculations. Following this approximation of separability, TΔS in Table 6.10 contains a sum of the rotational, translational and vibrational components of the peptide entropy. The translational and rotational components were calculated using statistical mechanical estimates for an ideal gas, while the vibrational component was estimated with the harmonic approximation using normal mode calculations. The configurational entropy that is associated with the reorganisation of the internal degrees of freedom of the molecule is not considered in the entropy calculations in Table 6.10. A previous literature study revealed that the configurational entropy can be substantial in situations where the peptide backbone is very dynamic,[28] as observed for the $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$ peptide in the present Chapter (Figure 6.10) and destabilize the triple helix.

The fully β-glycosylated $(PO^{\beta}G)_7$ monomer shows a marked reduction in ΔE_{elec} (28.0 kJ mol^{-1}) and ΔE_{vdw} (–31.8 kJ mol^{-1}), and an associated large decrease in the solvation energy compared to $(POG)_7$. This reduction in ΔE_{elec} and ΔE_{vdw} are explained by the structural information given in the previous section. In particular, the folded structure of the fully glycosylated monomer contains many intramolecular hydrogen bonds, and the sugar moieties are pointed towards the solvent (Figure 6.9), which reduces unfavourable steric interactions. Additionally, the more compact nature of the fold allows for more favourable van der Waals interactions. Most importantly, the total free energy for triple helix formation in $(PO^{\beta}G)_7$ is 14.2 kJ mol^{-1}, which suggests that triple helices are not viable, as proposed by experimental findings.[21] In summary, the MM–RISM analysis suggests that glycosylation increases the stability of the monomer strands through electrostatics, noncovalent interactions and solvation, which in turn disfavours triple helix formation.

6.4. Discussion

Average MD ϕ and ψ backbone dihedral angles (Figures 6.4 and 6.5) show that $(POG)_7$ adopts a triple helix structure, as per experiment. In contrast, while experiments show that $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, $[(POG)$–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ do not form triple helices, average MD and DFT backbone dihedral angles, the PPII conformation and interstrand (N–H···O=C) hydrogen bonds are preseved upon glycosylation (Tables 6.2 to 6.5). This suggests that glycosylation may be permited if a triple helix is in place. Regardless, the thermodynamics of monomer to triple helix formation is not favourable.

Analysis of the simulation data on the monomers reveals a collapse from the extended conformations adopted in the triple helix to more globular forms containing many intramolecular hydrogen bonds. The monomer of the unglycosylated $(POG)_7$ peptide contains a turn segment that effectively bends the peptide and brings the two end segments into a parallel arrangement (Figure 6.9). The $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$ and $[(POG)$–$(PO^{\beta}G)]_3$–(POG) peptides have a similar turn segment which contains a glycosylated Hyp residue (Figure 6.9). The two other sugar moieties in the $[(POG)$–$(PO^{\beta}G)]_3$–(POG) peptide interact via stabilizing hydrogen bonds. For $(PO^{\beta}G)_7$, the monomer adopts a folded structure that is very stable according to the calculations. The alteration of peptide structure by glycosylation has been observed in several literature studies where glycosylation has been deemed to act as a 'conformational switch'.[47-48,52-53] Therefore, the likely reason for triple helix destabilization with an increasing level of glycosylation is the additional conformational stability the sugar moieties provide to the monomer strands.

The structural proposal described above is supported by the calculated trends in monomer to triple helix conversion energetics (Table 6.9). In particular, (force field) electrostatic energies become more unfavourable with increase in degree of glycosylation for the β-glycosylated compounds. ΔE_{elec}, −171.8, −130.4, and 20.8 kJ mol^{-1} for $(POG)_7$, $[(POG)–(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$, respectively, which suggests that the sugar is less sterically hindered in the monomer compared to the triple helix. Additionally, the most stable conformation according to PCM–B3LYP/6–31G(d) calculations for the (POG)–$(PO^{\beta}G)$–(POG) model occurs when $\varphi_g \approx 50°$, which maximally directs the sugar away from the rest of the peptide (Table 6.6). These findings support previous literature proposals that unfavourable steric interactions due to bulky groups can destabilize the triple helix.[17,57] For example, steric repulsions due to 30 acetate groups were used in the literature to explain the 11 °C drop in T_m observed for a contiguously *O*–acylated $[Pro–Hyp(C(O)CH_3)Gly]_{10}$ triple helix compared to the unmodified $[Pro–Hyp–Gly]_{10}$ compound.[57] In addition, sterically demanding groups such as (α/β)–galactosylated triazole moieties produced a 7 °C T_m decrease when attached to an azidoproline (Azp) residue in $(Pro–Hyp–Gly)_3$–[Pro–Azp–Gly]–$(Pro–Hyp–Gly)_3$ triple helices.[17] The stability of the monomeric α-glycosylated $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$ strand may be entropically controlled as $-T\Delta S$ (rotational + translational + vibrational) is the largest (~ 89 kJ mol^{-1}) for this peptide. Additionally, conformational entropy could be significant given the very dynamic backbone.

The finding that monomeric states are favoured over the triple helix in the glycosylated compounds is also buttressed by the electrostatic energies, which point to stabilizing intramolecular hydrogen–bonding contacts as explicitly discussed for the $[(POG)–(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ peptides. Additionally, solvation is important in order to explain the experimental findings. Specifically, calculated solvation energy estimates show that solute–solvent interactions become increasing more favourable upon glycosylation. The solvation energy ($\Delta E_{pol}+\Delta E_{apol}$) is 277.6, 267.5 and 18.4 and kJ mol^{-1} for $(POG)_7$, $[(POG)–(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ respectively which underscores the need to include solvent in the optimization process, which was done in this chapter, but is currently not common in the literature.[40,45] Indeed, the solvent phase PCM-B3LYP calculations (Table 6.8) results agree with the observed drop in melting temperature upon glycosylation with a triplet of sugars. Thus, the monomers are stabilized to a greater extent by solvent than the triple helix. These results are in contrast to previous studies on $(Pro–Thr–Gly)_{10}$ peptides, which require Thr glycosylation for triple helix formation.[58-59] Specifically, the glycosylated $(Pro–Thr–Gly)_{10}$ CMPs that were synthesised based on the collagen observed in the cuticle of deep sea warms notably showed a lack of Hyp residues, which has been shown to correlate with triple helix stability.[60-61] However, very slow proton exchange by the NH groups involved in triple helix hydrogen bonding was observed in NMR experiments, leading the authors to propose that the attached sugar moieties shield the backbone from water. This in turn leads to stronger C=O···H–N hydrogen bonds, and consequently a more stable triple helix. The data obtained in this chapter shows a high occupancy of the C=O···H–N (>95%, Tables 6.2-6.4) hydrogen bonds for all glycosylated compounds, indicating that shielding of the backbone will likely not provide further stabilization.

Thus, a combination of enhanced electrostatic (stabilizing intramolecular hydrogen bonding) and reduced steric clashes, as well as other non–bonding contacts, stabilize the glycosylated collagen monomers compared to the corresponding triple helix. The stability of the monomer strands increases proportionally to the number of attached glycans and explains the experimentally-observed decrease in triple helix stability upon glycosylation.[21]

6.5. Conclusion

As discussed in Chapter 1, PTMs occur frequently in nature and influence the structure and properties of proteins. Collagen (the most abundant protein in animals), undergoes several modifications, including glycosylation. However, hydroxyproline glycosylation has not been observed in collagen even though it occurs in HRGPs (the corresponding plant structural protein).

In this Chapter, the structures and energetics of glycosylated collagen model peptides were studied using extensive MD simulations in explicit solvent, as well as DFT optimizations in implicit solvent. The aim was to explain experimental observations that Hyp glycosylation destabilizes the collagen triple helix,[21] and thereby rationalize the lack of Hyp glycosylation in animal collagen. The results show that the PPII backbone structure, as well as interstrand hydrogen bonding, is preserved in the glycosylated triple helical models. Instead, the conformations of monomer strands are greatly affected by glycosylation through intramolecular hydrogen bonds, sterics and solvation. Indeed, analysis of the energetics of triple helix formation (using both PCM–B3LYP/6–31G(d) and molecular mechanics-based MM–RISM) reveals that electrostatics (intramolecular hydrogen bonding), noncovalent interactions (van der Waals), and solvation (sugar–solvent) destabilize the triple helix upon glycosylation. In contrast, the more compact structures of the monomers upon glycosylation enhance the van der Waals contacts in the monomer phase. Thus, a combination of unfavourable interactions in the triple helix, and favourable interactions in the monomers, contribute to favour the monomeric state of the $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, $[(POG)$–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ peptides over the triple helix as observed experimentally.

The results outline above may explain why Hyp residues in collagen are not glycosylated in nature. This is in contrast to lysine glycosylation which occurs naturally in vertebrates. Additionally, the cuticle of deep sea worms has been found to contain glycosylated collagen. Interestingly, the collagen of these deep sea worms replaces Hyp with glycosylated threonine in the Y position of the collagen X-Y-Gly triplet repeat.

In conclusion, Hyp glycosylation favours the monomer state of collagen model peptides over the triple helical conformation and rationalizes why this post-translational modification is absent in vertebrates.

6.6. References[b]

1. Shoulders, M. D.; Raines, R. T., Collagen Structure and Stability. In *Annu. Rev. Biochem.*, Annual Reviews: Palo Alto, 2009; Vol. 78, pp 929–958.

2. Fields, G. B., The collagen triple-helix: Correlation of conformation with biological activities. *Connect. Tissue Res.* **1995,** *31*, 235–243.

3. Byers, P. H.; Steiner, R. D., Osteogenesis Imperfecta. *Annu. Rev. Med.* **1992,** *43*, 269–282.

4. Bogaert, R.; Tiller, G. E.; Weis, M. A.; Gruber, H. E.; Rimoin, D. L.; Cohn, D. H.; Eyre, D. R., An amino-acis substitution (Gly853-Glu) in the Collagen Alpha-1(II) CHain Produces Hypochondrogenesis. *J. Biol. Chem.* **1992,** *267*, 22522–22526.

5. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M., Crystal structure and molecular structure of a collagen-like peptide at 1.9 Angstrom resolution. *Science* **1994,** *266*, 75–81.

6. Ramachandran, G. N.; Kartha, G., Structure of Collagen. *Nature* **1954,** *174*, 269–270.

7. Ramachandran, G. N.; Kartha, G., Structure of Collagen. *Nature* **1955,** *176*, 593–595.

8. Rich, A.; Crick, F. H. C., Structure of Collagen. *Nature* **1955,** *176*, 915–916.

9. Ramshaw, J. A. M.; Shah, N. K.; Brodsky, B., Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J. Struct. Biol.* **1998,** *122*, 86–91.

[b] ACS referencing style was implemented throughout this thesis.

10. Eberhardt, E. S.; Loh, S. N.; Raines, R. T., Thermodynamic origin of prolyl peptide bond isomers. *Tetrahedron Lett.* **1993,** *34*, 3055–3056.

11. Panasik, N., Jr.; Eberhardt, E. S.; Edison, A. S.; Powell, D. R.; Raines, R. T., Inductive effects on the structure of proline residues. *Int. J. of Pept. Protein Res.***1994,** *44*, 262–269.

12. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T., Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **2001,** *123*, 777–778.

13. Shoulders, M. D.; Hodges, J. A.; Raines, R. T., Reciprocity of Steric and Stereoelectronic Effects in the Collagen Triple Helix. *J. Am. Chem. Soc.* **2006,** *128*, 8112–8113.

14. Shoulders, M. D.; Kotch, F. W.; Choudhary, A.; Guzei, I. A.; Raines, R. T., The Aberrance of the 4S Diastereomer of 4-Hydroxyproline. *J. Am. Chem. Soc.* **2010,** *132*, 10857–10865.

15. Improta, R.; Benzi, C.; Barone, V., Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* **2001,** *123*, 12568–12577.

16. Improta, R.; Mele, F.; Crescenzi, O.; Benzi, C.; Barone, V., Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 2. A Quantum Mechanical/Molecular Mechanical Study of (Proline-Proline-Glycine)n Polypeptides. *J. Am. Chem. Soc.* **2002,** *124*, 7857–7865.

17. Erdmann, R. S.; Wennemers, H., Functionalizable Collagen Model Peptides. *J. Am. Chem. Soc.* **2010,** *132*, 13957–13959.

18. Myllyharju, J.; Kivirikko, K. I., Collagens and collagen-related diseases. *Ann. Med.* **2001,** *33*, 7–21.

19. Fields, G. B., Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.* **2010,** *8*, 1237–1258.

20. Fields, C. G.; Mickelson, D. J.; Drake, S. L.; McCarthy, J. B.; Fields, G. B., Melanoma cell adhesion and spreading activities of a synthetic 124-residue triple helical mini-collagen. *J. Biol. Chem.* **1993,** *268*, 14153–14160.

21. S. Bommagani, E.B. Naziga, N. W. Owens, E. Lattová, J. D. O'Neil, S. D. Wetmore and F. Schweizer, The Effects of (2*S*,4*R*)-4-Hydroxyproline Glycosylation on the Stability of the Collagen Triple Helix, Unpublished work. Copyright 2013 Wiley VCH Verlag

22. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, A Package of Computer-Programs for

Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995,** *91*, 1–41.

23. Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; R.C.Walker; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER*, 10; University of California, San Francisco: 2008.

24. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65*, 712–725.

25. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29*, 622–655.

26. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phy.* **1983,** *79*, 926–935.

27. Jorgensen, W. L.; Jenson, C., Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J. Comput. Chem.* **1998,** *19*, 1179–1186.

28. Suarez, E.; Diaz, N.; Suarez, D., Entropic Control of the Relative Stability of Triple-helical Collagen Peptide Models. *J. Phys. Chem. B* **2008,** *112*, 15248–15255.

29. Kolossvary, I.; Guida, W. C., Low mode search. An efficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.* **1996,** *118*, 5011–5019.

30. Kolossvary, I.; Keseru, G. M., Hessian-free low-mode conformational search for large-scale protein loop optimization: Application to c-jun N-terminal kinase JNK3. *J. Comput. Chem.* **2001,** *22*, 21–30.

31. Kovalenko, A.; Hirata, F., Three-dimensional density profiles of water in contact with a solute of arbitrary shape: A RISM approach. *Chem. Phys. Lett.***1998,** *290*, 237–244.

32. Kovalenko, A.; Hirata, F., Self-consistent description of a metal-water interface by the Kohn-Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999,** *110*, 10095–10112.

33. Kovalenko, A.; Hirata, F., Potential of mean force between two molecular ions in a polar molecular solvent: A study by the three-dimensional reference interaction site model. *J. Phys. Chem. B* **1999,** *103*, 7942–7957.

34. Gusarov, S.; Pujari, B. S.; Kovalenko, A., Efficient treatment of solvation shells in 3D molecular theory of solvation. *J. Comput. Chem.* **2012,** *33*, 1478–1494.

35. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theor. Comput.* **2012,** *8*, 3314–3321.

36. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, ., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. , Gaussian 09, Revision A.1, Gaussian, Inc.,: 2009.

37. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105*, 2999–3094.

38. Simon, S.; Duran, M.; Dannenberg, J. J., How does basis set superposition error change the potential surfaces for hydrogen bonded dimers? *J. Chem. Phys.* **1996,** *105*, 11024–11031.

39. Boys, S. F.; Bernardi, F., Calculation of Small Molecular Interactions By Differences of Separate Total Energies – Some Procedures with Reduced Errors. *Mol. Phys.* **1970,** *19*, 553–559.

40. Pálfi, V. K.; Perczel, A., How stable is a collagen triple helix? An ab initio study on various collagen and β-sheet forming sequences. *J. Comput. Chem.* **2008,** *29*, 1374–1386.

41. Mládek, A. t.; Sharma, P.; Mitra, A.; Bhattacharyya, D.; Šponer, J. i.; Šponer, J. E., Trans Hoogsteen/Sugar Edge Base Pairing in RNA. Structures, Energies, and Stabilities from Quantum Chemical Calculations. *J. Phys. Chem. B* **2009,** *113*, 1743–1755.

42. Berg, R. A.; Olsen, B. R.; Prockop, D. J., Titration and Melting Curves of the Collagen-like Triple Helices Formed from (Pro-Pro-Gly)10 in Aqueous Solution. *J. Biol. Chem.* **1970,** *245*, 5759–5763.

43. Shikata, T.; Yoshida, N.; Okuyama, K., Anomalous Dehydration Behavior of a Short Collagen Model Polypeptide, (l-Prolyl-l-ProlylGlycyl)5, in Aqueous Solution. *J. Phys. Chem. Lett.* **2009,** *1*, 412-416.

44. Raman, S. S.; Parthasarathi, R.; Subramanian, V.; Ramasami, T., Role of Length-Dependent Stability of Collagen-like Peptides. *J. Phys. Chem. B* **2008,** *112*, 1533–1539.

45. Xu, Y.; Dannenberg, J. J., Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProDAla, and ProProDSer Triads. *J. Am. Chem. Soc.* **2005,** *127*, 14130–14131.

46. Palfi, V. K.; Perczel, A., Stability of the Hydration Layer of Tropocollagen: A QM Study. *J. Comput. Chem.* **2010,** *31*, 764–777.

47. Shental-Bechor, D.; Levy, Y., Effect of glycosylation on protein folding: A dose book at thermodynamic stabilization. *Proc. Nat. Acad. Sci. U. S. A.* **2008,** *105*, 8256–8261.

48. Shental-Bechor, D.; Levy, Y., Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Curr. Op. Struct. Biol.* **2009,** *19*, 524–533.

49. Meyer, B.; Moeller, H., Conformation of glycopeptides and glycoproteins. *Topics Curr. Chem.* **2007,** *267*, 187–251.

50. Cheng, S. M.; Edwards, S. A.; Jiang, Y. D.; Grater, F., Glycosylation Enhances Peptide Hydrophobic Collapse by Impairing Solvation. *ChemPhysChem* **2010,** *11*, 2367–2374.

51. Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B., Effects of Glycosylation on Peptide Conformation: A Synergistic Experimental and Computational Study. *J. Am. Chem. Soc.* **2004,** *126*, 8421–8425.

52. Matsushita, T.; Ohyabu, N.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Hinou, H.; Nishimura, S.-I., Site-Specific Conformational Alteration Induced by Sialylation of MUC1 Tandem Repeating Glycopeptides at an Epitope Region for the Anti-KL-6 Monoclonal Antibody. *Biochemistry* **2012,** *52*, 402–414.

53. Ellis, C. R.; Maiti, B.; Noid, W. G., Specific and Nonspecific Effects of Glycosylation. *J. Am. Chem. Soc.* **2012,** *134*, 8184–8193.

54. Owens, N. W.; Stetefeld, J.; Lattova, E.; Schweizer, F., Contiguous O-Galactosylation of 4(R)-Hydroxy-L-proline Residues Forms Very Stable Polyproline II Helices. *J. Am. Chem. Soc.* **2010,** *132*, 5036–5042.

55. Naziga, E. B.; Schweizer, F.; Wetmore, S. D., Conformational Study of the Hydroxyproline-O-Glycosidic Linkage: Sugar-Peptide Orientation and Prolyl Amide lsomerization in (alpha/beta) Galactosylated 4(R/S)-Hydroxyproline. *J. Phys. Chem. B* **2012,** *116*, 860–871.

56. Erdmann, R. S.; Wennemers, H., Importance of Ring Puckering versus Interstrand Hydrogen Bonds for the Conformational Stability of Collagen. *Angew. Chem.-Int. Edit.* **2011,** *50*, 6835–6838.

57. Jenkins, C. L.; McCloskey, A. I.; Guzei, I. A.; Eberhardt, E. S.; Raines, R. T., O-acylation of hydroxyproline residues: effect on peptide-bond isomerization and collagen stability. *Biopolymers* **2005,** *80*, 1–8.

58. Bann, J. G.; Peyton, D. H.; Bachinger, H. P., Sweet is stable: glycosylation stabilizes collagen. *FEBS Lett.* **2000,** *473*, 237–240.

59. Bann, J. G.; Bachinger, H. P., Glycosylation/hydroxylation-induced stabilization of the collagen triple helix - 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix. *J. Biol. Chem.* **2000,** *275*, 24466–24469.

60. Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B., Amino acid propensities for the collagen triple-helix. *Biochemistry* **2000,** *39*, 14960–14967.

61. Brodsky, B.; Thiagarajan, G.; Madhan, B.; Kar, K., Triple-helical peptides: An approach to collagen conformation, stability, and self-association. *Biopolymers* **2008,** *89*, 345–353.

Chapter 7. Conclusions and Future Work

7.1. General Overview

This thesis set out to determine how hydroxyproline O–galactosylation affects the structure of model peptides and, by extension, the biological implications of this modification. This work was necessitated by a general lack of details on this type of glycosylation, as well as recent literature studies showing that Hyp glycosylation has a profound importance in many aspects of plant life.[1-3] In this concluding chapter, a summary of the work carried out in this thesis to accomplish these goals is provided, and avenues for future studies that can be carried out using the computational procedures used in this thesis are outlined.

7.2. Contributions from this Thesis

In order to explore the effects of Hyp glycosylation on peptide and protein structure, the glycosidic linkage must be characterized since no computational or experimental three dimensional data on its structure exists in the literature. Furthermore, an understanding of the linkage conformation is vital since it directs glycan orientation, and hence dictates sugar–peptide interactions. Additionally, experimental studies[4] on Ac–(4*R*/*S*–Hyp–[α/β–Gal])–OMe model compounds revealed that 4R–glycosylation of Hyp does not lead to a measurable change in the *trans* population of the prolyl ω dihedral that is important in the biologically-important PPII conformation. However, 4*S*–glycosylation favours the *trans* structure, and leads to an increase in the isomerization rate. Since the reasons for these observations were not clear from the experimental data, Chapter 3 of this thesis used MD simulations in explicit water and implicit solvent DFT calculations to investigate the experimental findings.

The data revealed the presence of two major conformations of the glycosidic linkage dihedrals for both α and β O–linked galactose sugars. However, glycosylation using β anomeric sugars produces structures with the sugar directed away from the peptide into solution, while α linkage results in more compact, stacked–like arrangements (Figure 3.6). This is interesting since plant HRGPs usually contain β O–glycosidic linkages. As mentioned in Chapter 1, two proposals have been made for the 3D structure of HRGPs. The "wattle blossom" model depicts a glycopeptide structure with globular sugars that point away from the peptide backbone, while the "twisted hairy rope" model suggests the sugars adopt a parallel-like fashion and interact with the protein backbone. The structure of the β O–glycosidic linkages determined in this thesis is more consistent with the "wattle blossom" model since the sugar moiety is directed away from the protein component in the glycopeptides (Figure 3.6). However, only monosaccharides were attached to the peptides considered in this thesis. Covalent linkage of a polysaccharide could lead to a more complex structure than predicted by the simple "wattle blossom" or "twisted hairy rope" models. In particular, while the structure of the β O–glycosidic is consistent with the "wattle blossom" model, oligosaccharide growth via linkage to the O6 hydroxyl for instance could place the remainder of the sugar in a parallel fashion to the peptide backbone, leading to an amalgam of both proposed models.

An explanation for the increase in the *trans* isomer stability observed for the 4*S* compound was provided by the simulations. Specifically, a hydrogen–bonding interaction between the sugar moiety and the peptide backbone occurs due to C^{γ}–endo puckering induced by the 4*S*–substituent. This interaction is sometimes mediated by a water molecule and organises the peptide backbone for the stabilizing $n \rightarrow \pi^*$ interaction that favours the *trans* isomer.

The work carried out in Chapter 3 provided the first molecular level structural information in the literature on the conformation of the Hyp–O–glycosidic linkage. The preferred orientation of the sugar and (hydroxy) proline rings determined in Chapter 3 indicate that intra (to peptide) and inter (to solvent) interactions could occur in oligopeptides that include the glycosylated Hyp residue. Such interactions could stabilize the PPII conformation of the peptide backbone, implying that the glycosylation observed in biological HRGPs plays a structural role. Indeed, the results obtained in Chapter 4 using advanced MD sampling (REMD, ABMD and HT–REMD) of the Ac–$(Pro)_9$–NH_2, Ac–$(Hyp)_9$–NH_2 and Ac–(Hyp–[β–Gal]$)_9$–NH_2 peptides demonstrate the existence of intramolecular sugar–peptide and sugar–sugar hydrogen–bonding interactions in the contiguously glycosylated peptide (Figure 4.2). Additionally, it was determined that sugar–solvent interactions significantly stabilize the PPII conformation in Ac–(Hyp–[β–Gal]$)_9$–NH_2 relative to the unglycosylated and hydroxylated peptides. These results explain the very high melting temperature (70°C) observed experimentally for Ac–(Hyp–[β–Gal]$)_9$–NH_2 and were used to propose a likely mode of action for the covalently–attached sugars that are present in the contiguously glycosylated segments of plant hydroxyproline rich glycoproteins (HRGPs)[3,5] namely stabilizing the PPII conformation of the peptide backbone.

Chapter 5 uses the effective suite of techniques applied in Chapter 4 to study the effects of non–contiguous glycosylation. This study was motivated in part by the presence of non–contiguously glycosylated sequences in biological HGRPs.[5] Additionally, synthetic compounds with the same glycosylation patterns (Table 5.1) were developed by Schweizer and coworkers as a model for this kind of natural glycosylation.[6] Specifically, the $(Pro)_9$, $(POP)_9$, (POP)–($PO^{\alpha}P$)–(POP), (POP)–($PO^{\beta}P$)–(POP), $(PO^{\alpha}P)_3$ and $(PO^{\beta}P)_3$ peptides were considered. Experimental results obtained from CD spectroscopy suggested that the attachment of one sugar moiety ((POP)–(PO$^{\alpha/\beta}$P)–(POP)) stabilizes the PPII conformation to a greater extent than glycosylation with three sugars ((PO$^{\alpha/\beta}$P)$_3$). This is in contrast to the stability observed for the contiguously glycosylated Hyp peptide considered in Chapter 4, and the reason for this difference was not clear from the experimental data. The molecular modeling implemented in Chapter 5 revealed that intramolecular sugar–peptide hydrogen-bonding interactions that favour the *cis* conformation (Figure 5.2) of certain residues are introduced upon glycosylation. Furthermore, more such contacts exist in the triply glycosylated (PO$^{\alpha/\beta}$P)$_3$ relative to the singly glycosylated (POP)–(PO$^{\alpha/\beta}$P)$_3$–(POP), which may be the reason for the relative decrease in PPII character observed experimentally for the non-contiguously glycosylated peptides simulated in Chapter 5. These results for non-contiguous glycosylation are similar to previous literature results, which show that non-covalent interactions can tune *cis/trans* stability in Pro containing peptides and proteins.[20] For example, aromatic amino acid residues in a Pro-Pro-Xaa (Xaa = aromatic residue) sequence can greatly increase the *cis* population of the preceding Pro residue.[20] The significant increase in the *cis* population of Pro in non-contiguously glycosylated compounds studied in this thesis suggests that such modifications may provide a way to introduce structural bias into oligoproline peptides (proteins).

The results in Chapters 3 to 5 further support literature proposals that glycosylation provides structural stability to the PPII conformation adopted by HRGPs.[2-3] Additionally, the results show that the location of the glycosylation is vital to conformational stability, at least when a monosaccharide is attached to non-contiguously glycosylated Hyp oligopeptide. However, the steric demands of a complex polysaccharide may compensate for the lack of contiguous glycosylation. In this light, it is interesting to note that it has been proposed in the literature that clusters of non-contiguous Hyp are glycosylated with arabinogalactan polysaccharides, while contiguously glycosylated Hyp contain smaller arabinose sugars.[5]

While all models presented above (Chapters 3 – 5) are associated with naturally-occurring glycosylation in plants, Chapter 6 aims to explain the lack of such modifications in collagen (animals). Interestingly, while threonine[7-8] and lysine[9-10] residues in collagen can be glycosylated in nature, Hyp–O–glycosylation has not been observed in collagen, despite containing a large number of hydroxyproline residues and being structurally similar to HGRPs.[3,11] Therefore, triple helices of the $(POG)_7$, $(POG)_3$–$(PO^{\beta}G)$–$(POG)_3$, $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, $[(POG)$–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ CMPs with varying degrees of glycosylation, as well as their monomeric strands, were simulated using MD. Furthermore, the $(POG)_3$ and (POG)–$(PO^{\beta}G)$–(POG) triple helical segments were optimized using implicit solvent DFT calculations. The same CMPs were synthesized by Schweizer and coworkers and characterized using CD and NMR spectroscopies.[6] Experiments show that glycosylation destabilizes the triple helix. In particular, a slight reduction in the melting temperature (~2°C) was observed in $(POG)_3$–$(PO^{\beta}G)$–$(POG)_3$ compared to the unmodified CMP. However, all other glycosylated CMPs did not form a triple helix.

Analysis of MD data on the triple helices demonstrates that vital interstrand hydrogen-bonding interactions are preserved in all glycosylated CMPs. Moreover, new intra (α and β-glycosylated CMPs) and intermolecular (α-glycosylated CMP) hydrogen bonds involving the sugar hydroxyl groups and the peptide backbone are introduced that could stabilize the triple helix. Therefore, no conclusions could be drawn about CMP stability based on triple helical data alone. However, MD simulation on the corresponding monomer strands indicate that glycosylation stabilizes the monomer conformations, where the level of stability is related to the degree of glycosylation. Molecular mechanics analysis of the energetics of monomer to triple helix formation showed that glycosylation provides a better electrostatic, entropic and solvation environment for the monomer compared to the triple helix in the CMPs. This finding explains the lack of triple helices in the $(POG)_3$–$(PO^{\alpha}G)$–$(POG)_3$, $[(POG)$–$(PO^{\beta}G)]_3$–(POG) and $(PO^{\beta}G)_7$ CMPs.

The study of the glycosylated triple helices and their monomer strands provide a rational for the lack of Hyp glycosylation in vertebrate collagen. In contrast to the stabilizing effects of contiguous glycosylation of Hyp observed in plants, the biologically required triple helix is destabilized by Hyp glycosylation in vertebrate collagen. In particular, HRGPs make use of glycosylation to stabilize the PPII structure through intramolecular and intermolecular interactions, which justifies the presence of this double post-translational modification. Glycosylation with a single sugar moiety per stand ($(POG)_3$–$(PO^{\beta}G)$–$(POG)_3$), where the β-anomeric orientation of the sugar points it away from the helix allows for triple helix formation, albeit with a small reduction in melting temperature. However, more than three glycosylated Hyp residues per strand cannot be accommodated, and thereby leads to loss of the triple helix. These results are in contrast to *O*-acylation of Hyp which led to only a moderate destabilization of the triple helix.[19] For instance, contiguous *O*-acylation of Hyp in $[Pro\text{-}Hyp\text{-}Gly]_{10}$ as in $[Pro\text{-}Hyp(C(O)CH_3)Gly]_{10}$ results in only a 11 °C drop in melting point of the triple helix. This destabilization was attributed to steric interactions caused by 30 acetate groups in the trimeric complex. Also, an interesting comparison is found in the fact that the β-D-galactosylation of Thr in $Ac\text{-}[Gly\text{-}Pro\text{-}Thr]_{10}\text{-}NH_2$ collagen model peptides is required for triple helix formation, where the carbohydrate was attributed with shielding the triple helix from strong interactions with water molecules, thereby stabilizing interchain N-H to C=O hydrogen bonding interactions.[7-8] In conclusion, it appears that the absence or presence of Hyp-glycosylation provides a rationale for the observed different aggregation states of the PPII helix in vertebrate and plant species, respectively.

Overall, the work presented in this thesis has used a blend of different computational theories and techniques to provide important structural details about Hyp O–glycosylation and explained the results of several experimental investigations. In particular, sophisticated MD techniques, such as REMD, ABMD, HT–REMD and umbrella sampling, were implemented. Additionally, electronic structure information was obtained using PCM–B3LYP calculations on select conformations identified using MD simulations. The level of detailed information obtained using these approaches demonstrate that the suite of computational protocols implemented in this thesis can be applied to many problems that involve elucidating the structural details of oligopeptides. Furthermore, the work in this thesis highlights the effectiveness of a combined approach to biochemical problems, which should be applied to further expand on the work carried out in this thesis.

7.3. Future Work

The structural information about Hyp O–glycosylation described above demonstrate that the methodologies used in this thesis can be applied to other systems to yield valuable insight into peptide structure. In this light, several avenues can be pursued to continue the work on Hyp glycosylation. For instance, it is known that extensins (a class of HRGPs) are glycosylated with five–membered arabinose sugars.[12-13] Like the arabinogalactan proteins, the extensin subfamily of HRGPs have been implicated in cell wall assembly and growth in plants.[1-3] They commonly contain the X–$(Hyp)_n$ sequence motif (n is typically 4–6), where the serine amino acid residue is frequent in the X position and is glycosylated with short arabinose polysaccharides. This particular O–glycosidic linkage should be characterized and compared to the Hyp–galactose linkage considered in this thesis. Thereafter, the effects of contiguous and non–contiguous arabinosylation should be evaluated. These investigations should evaluate whether the consequences of glycosylation are sugar specific or vary depending on the type of attached sugar. Additionally, including Ala and Ser interruptions should be considered since the $(Ala/Ser–Hyp)_n$ sequence repeat has been observed in several HRGPs.[5]

It would be interesting to consider the effects of expanding the carbohydrate moiety attached to the non–contiguously glycosylated compounds to more complex sugars, such as the polysaccharides identified by Kieliszewski and coworkers.[14-15] The data presented in this thesis provides evidence that intramolecular hydrogen bonding reduces PPII character in non–contiguously glycosylated oligoprolines upon attachment of a simple monosaccharide sugar. However, bulkier polysaccharides could alter this pattern and may lead to stronger PPII helices in non-contiguously as well as contiguously glycosylated Hyp containing oligopeptides (proteins).

While experimental data and modeling show that multiple Hyp glycosylation does not allow triple helix formation in the CMPs discussed in this thesis, collagen obtained from the cuticle of certain worms only form a triple helices when threonine residues are highly glycosylated.[7-8] However, there is a reduced occurrence of Hyp residues in such collagen. A computational study using the methodologies implemented in Chapter 6 of this thesis will provide the structural and energetic basis for these observations and explain why a glycosylated threonine per triplet of residues is required for collagen stability. Additionally, the associated study will provide vital information about how glycosylated threonine compensates for a reduced number of Hyp residues in such collagen.

As mentioned in previous Chapters, some lysine residues of human collagen are glycosylated.[9,16] A galactose monomer or a glucose–galactose disaccharide is found to be the most frequent covalently linked sugars to (hydroxy)lysine.[9] The level of glycosylation varies among collagen types, with fibril forming collagen types I and III much less glycosylated than type II or IV collagen. This has led to the hypothesis that the degree of glycosylation is responsible for fibril size differences between types I and II collagen.[21] This notion is supported by the work of Torre-Blanco *et al*.[21] In their study, monomer strands of type I collagen from human fibroblasts was produced at an elevated temperature of 40.5°C, which hinders folding and allows for more post translational modifications before triple helix assembly. No measureable increase in thermal stability due to hydroxylation/glycosylation was observed, but the time required for collagen triple helix assembly increased. This implies that folding of the monomer strands is affected as N(C)-proteinase (which cleaves peptide terminals before assembly) is known to require a native peptide conformation for enzymatic activity.[23] Computational studies similar to what is implemented in this thesis will provide details about the effects lysine glycosylation on the triple helix and monomer strands of the collagen and help rationalize the experimental findings.

Additionally, it was observed in the experimental study by Torre-Blanco and coworkers described above[21] that the type I collagen fibrils obtained were much thinner in the sample prepared at 40.5 °C. This is in consonance with observations of thinner fibers for type IV collagen, which is naturally more glycosylated. An interesting project would study how lysine glycosylation influences collagen fibril formation. Here, a sequence from collagen will be glycosylated to various degrees, thereafter MD simulations will examine how self-association into fibrils depends on the degree of glycosylation. The results from the

MD study will explain the thinner fibers obtained at 40.5 °C and thereby provide further insight into the structural consequences of glycosylation which is the theme of this thesis.

7.4. Conclusions

The body of work presented in this thesis has revealed important structural information about the effects of naturally occurring Hyp glycosylation in the HRGPs of plants, as well as explained the natural lack of Hyp glycosylation in the collagen of animals. Interestingly, these effects are varied, and can be stabilizing or destabilizing to the PPII conformation that is important for collagen and HRGP structure. This heterogeneity of glycan action has been previously noted in the literature for other N(O)-glycosylated proteins[17] and entails that predicting the outcome of a specific glycosylation is difficult. Therefore, computational strategies, such as those applied in this thesis, in collaboration with experimental work will be very useful for unraveling the various effects of glycosylation in nature. This is particularly important since over half of animal proteins are glycosylated.[18]

7.5. References[a]

1. Velasquez, S. M.; Ricardi, M. M.; Dorosz, J. G.; Fernandez, P. V.; Nadra, A. D.; Pol-Fachin, L.; Egelund, J.; Gille, S.; Harholt, J.; Ciancia, M.; Verli, H.; Pauly, M.; Bacic, A.; Olsen, C. E.; Ulvskov, P.; Petersen, B. L.; Somerville, C.; Iusem, N. D.; Estevez, J. M., O-Glycosylated Cell Wall Proteins Are Essential in Root Hair Growth. *Science* **2011,** *332*, 1401-1403.

2. Cannon, M. C.; Terneus, K.; Hall, Q.; Tan, L.; Wang, Y. M.; Wegenhart, B. L.; Chen, L. W.; Lamport, D. T. A.; Chen, Y. N.; Kieliszewski, M. J., Self-assembly of the plant cell wall requires an extensin scaffold. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 2226-2231.

3. Kieliszewski, M. J.; Lamport, D. T. A.; Tan, L.; Cannon, M. C., Hydroxyproline-rich glycoproteins: form and function. *Annu. Plant Rev.* **2011,** *41*, 321-342.

[a] ACS referencing style was implemented throughout this thesis.

4. Owens, N. W.; Lee, A.; Marat, K.; Schweizer, F., The Implications of (2S,4S)-Hydroxyproline 4-O-Glycosylation for Prolyl Amide Isomerization. *Chem.-Eur. J.* **2009,** *15*, 10649-10657.

5. Kieliszewski, M. J., The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **2001,** *57*, 319-323.

6. E. B. Naziga et. al., Conformational Analysis of Non-contiguously Glycosylated Oligoprolines, In preparation. **2013**.

7. Bann, J. G.; Bachinger, H. P., Glycosylation/hydroxylation-induced stabilization of the collagen triple helix - 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix. *J. Biol. Chem.* **2000,** *275*, 24466-24469.

8. Bann, J. G.; Peyton, D. H.; Bachinger, H. P., Sweet is stable: glycosylation stabilizes collagen. *FEBS Lett.* **2000,** *473*, 237-240.

9. Torreblanco, A.; Adachi, E.; Hojima, Y.; Wootton, J. A. M.; Minor, R. R.; Prockop, D. J., Temperature induced posttranslational over-modification of the type I procollagen: effects of over-modification of the protein on the rate of cleavage by procollagen N-proteinase and on self assembly of collagen into fibrils. *J. Biol. Chem.* **1992,** *267*, 2650-2655.

10. Schegg, B.; Hulsmeier, A. J.; Rutschmann, C.; Maag, C.; Hennet, T., Core Glycosylation of Collagen Is Initiated by Two beta(1-O)Galactosyltransferases. *Mol. Cell. Biol.* **2009,** *29*, 943–952.

11. Shoulders, M. D.; Raines, R. T., Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009,** *78*, 929–958.

12. Kieliszewski, M. J.; Lamport, D. T. A., Extensin - repetitive motifs, functional sites, posttranslational codes and phylogeny. *Plant J.* **1994,** *5*, 157–172.

13. Lamport, D. T. A., Hydroxyproline-O-glycosidic linkage of the plant cell wall glycoprotein extensin. *Nature (London)* **1967,** *216*, 1322–1324.

14. Tan, L.; Qiu, F.; Lamport, D. T. A.; Kieliszewski, M. J., Structure of a hydroxyproline (Hyp)-arabinogalactan polysaccharide from repetitive Ala-Hyp expressed in transgenic Nicotiana tabacum. *J. Biol. Chem.* **2004,** *279*, 13156–13165.

15. Tan, L.; Varnai, P.; Lamport, D. T. A.; Yuan, C. H.; Xu, J. F.; Qiu, F.; Kieliszewski, M. J., Plant O-Hydroxyproline Arabinogalactans Are Composed of Repeating Trigalactosyl Subunits with Short Bifurcated Side Chains. *J. Biol. Chem.* **2010,** *285*, 24575–24583.

16. Spiro, R. G., Characterization and quantitative determination of hydroxylysince-linked carbohydrate units of several collagens. *J. Biol. Chem.* **1969,** *244*, 602–612.

17. Varki, A., Biological Roles of Oligosaccharides - All of the Theories are Correct. *Glycobiology* **1993,** *3*, 97–130.

18. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta-Gen. Subj.* **1999,** *1473*, 4–8.

19. C. L. Jenkins, A. I. McCloskey, I. A. Guzei, E. S. Eberhardt, R. T. Raines, O-acylation of hydroxyproline residues: effect on peptide-bond isomerization and collagen stability. *Biopolymers* **2005**, *80*, 1–8

20. Zondlo, N. J., Aromatic–Proline Interactions: Electronically Tunable CH/π Interactions. *Acc. Chem. Res.* **2012,** *46*, 1039–1049.

21. Notbohm, H.; Nokelainen, M.; Myllyharju, J.; Fietzek, P. P.; Muller, P. K.; Kivirikko, K. I., Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro. *J. Biol. Chem.* **1999,** *274*, 8988-8992.

22. Torreblanco, A.; Adachi, E.; Hojima, Y.; Wootton, J. A. M.; Minor, R. R.; Prockop, D. J., Temperature induced posttranslational over-modification of the type I procollagen: effects of over-modification of the protein on the rate of cleavage by procollagen N-proteinase and on self assembly of collagen into fibrils. *J. Biol. Chem.* **1992,** *267*, 2650-2655.

23. Hojima, Y.; McKenzie, J. A.; Vanderrest, M.; Prockop, D. J., Type I procollagen N-proteinase from chick-embryo tendons - purification of a new 500-KDA form of the enzyme and identification of the catalytically active polypeptides. *J. Biol. Chem.* **1989,** *264*, 11336-11345.

Appendix A

Conformational Study of the Hydroxyproline–O–Glycosidic Linkage: Sugar–Peptide Orientation and Prolyl Amide Isomerization in α(β)–Galactosylated 4R(S)–Hydroxyproline.

List of Contents

Table A1: PCM–B3LYP/6–311++G(d,p) calculated activation parameters (kJ/mol) for *cis-trans* isomerization in Hyp and hyp compounds[a].

Compound	$\Delta G_{TS\text{-}Cis}$		$\Delta H_{TS\text{-}Cis}$		$\Delta E_{TS\text{-}Cis}$	
	Minimum 1	Minimum 2	Minimum 1	Minimum 2	Minimum 1	Minimum 2
Hyp	81.7		77.5		79.5	
Hyp(α-Gal)	80.8	82.6	76.8	78.7	78.5	80.6
Hyp(β-Gal)	83.1	82.0	78.3	77.2	79.4	79.2
hyp	84.7	77.5	82.2	74.4	83.8	76.2
hyp(α-Gal)	80.1	79.5	76.0	74.8	78.0	76.9
hyp(β-Gal)	81.7	83.8	78.1	79.0	79.9	81.0

[a]For unglycosylated hyp, minimum 2 refers to the conformation with an intramolecular hydrogen bond between the C^{δ} hydroxyl and the peptide backbone. In glycosylated models, minima refer to the conformation about the glycosidic linkage defined by φ_g and ψ_g(see Scheme 3 and Figures 4 and 5). In glycosylated hyp compounds, minimum 2 has an intramolecular hydrogen bond between sugar and peptide.

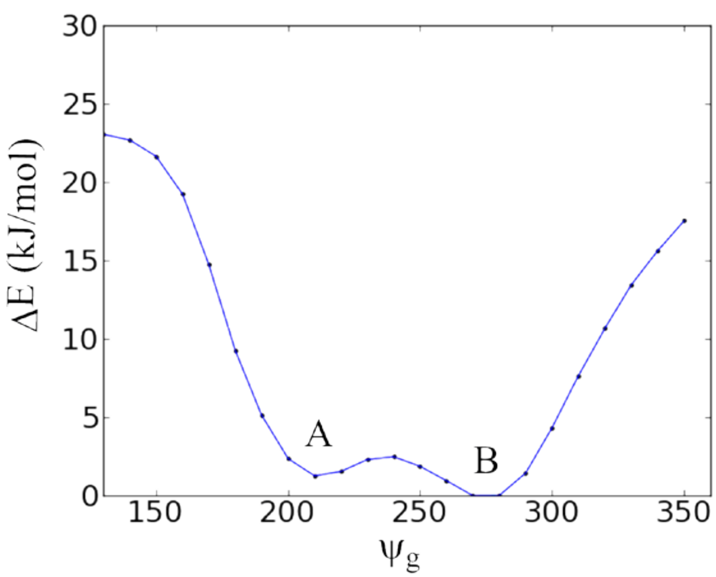

Figure A1. B3LYP/6-31G(d,p) PES scan of the glycosidic linkage ψ_g dihedral angle of Hyp–β–Gal

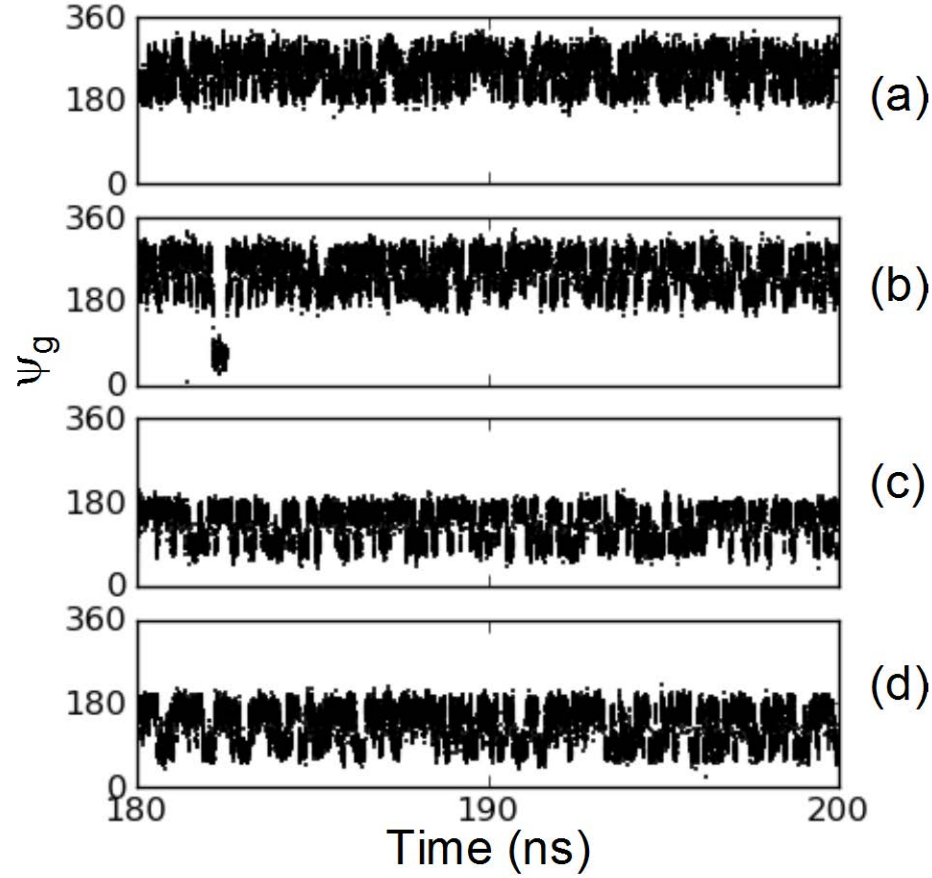

Figure A2. Evolution of the glycosidic linkage ψ_g dihedral angle in the last 20 ns of the MD simulation in (a) Hyp–α–Gal, (b) Hyp–β–Gal, (c) hyp–α–Gal and (d) hyp–β–Gal.

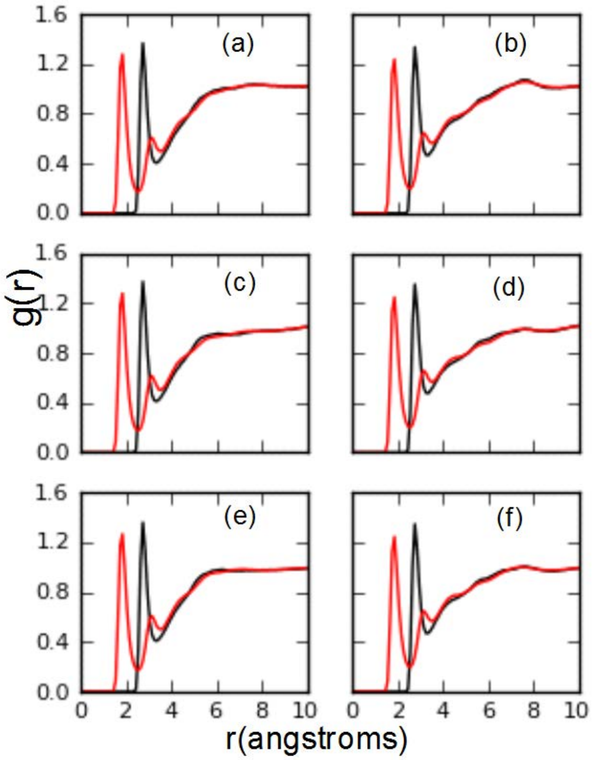

Figure A3. RDFs for interaction between a water oxygen (OW, black), water hydrogen (HW, red) and the peptide backbone O1 (left) or O (right) of the trans isomer of (a) Hyp (b) Hyp (c) Hyp–α–Gal, (d) Hyp–α–Gal, (e) Hyp–β–Gal and (f) Hyp–β–Gal.

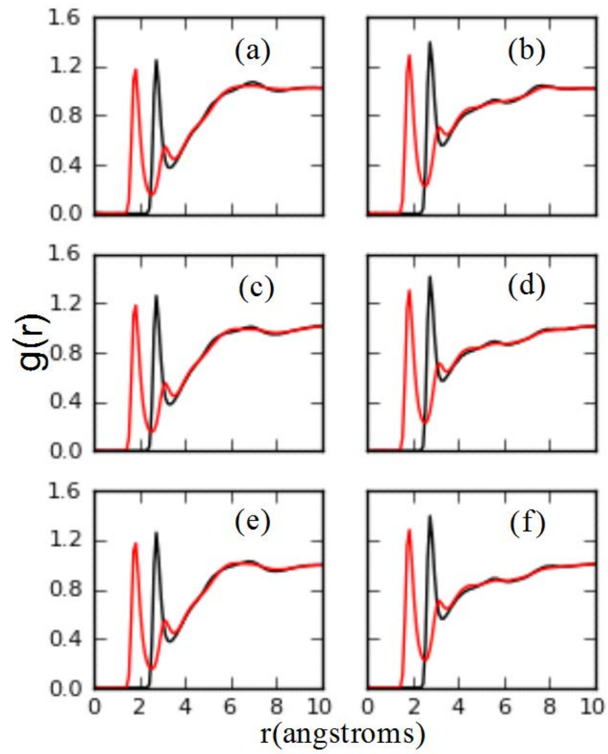

Figure A4. RDFs for interaction between a water oxygen (OW, black), water hydrogen (HW, red) and the peptide backbone O1 (left) or O (right) of the cis isomer of (a) Hyp (b) Hyp (c) Hyp–α–Gal, (d) Hyp–α–Gal, (e) Hyp–β–Gal and (f) Hyp–β–Gal.

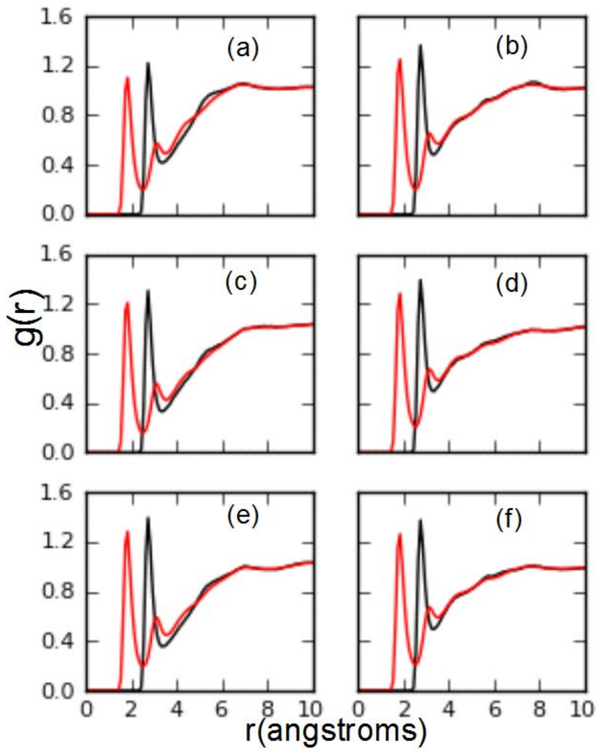

Figure A5. RDFs for interaction between a water oxygen (OW, black), water hydrogen (HW, red) and the peptide backbone O1 (left) or O (right) of the trans isomer of (a) hyp (b) hyp (c) hyp–α–Gal, (d) hyp–α–Gal, (e) hyp–β–Gal and (f) hyp–β–Gal.

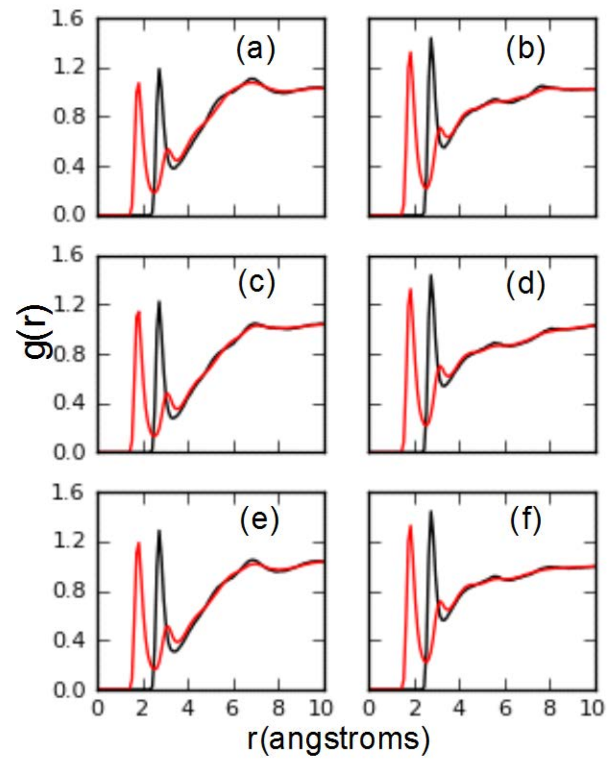

Figure A6. RDFs for interaction between a water oxygen (OW, black), water hydrogen (HW, red) and the peptide backbone O1 (left) or O (right) of the cis isomer of (a) hyp (b) hyp (c) hyp–α–Gal, (d) hyp–α–Gal, (e) hyp–β–Gal and (f) hyp–β–Gal.

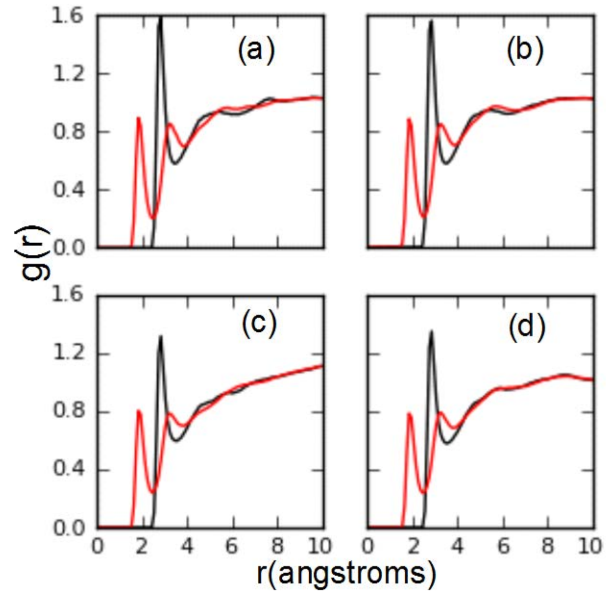

Figure A7. RDFs for interactions between a water oxygen (OW, black), water hydrogen (HW, red) and the peptide O^{δ} of Hyp (a, b) or hyp (c, d) in the *trans* (a, c) or *cis* (b, d) conformation.

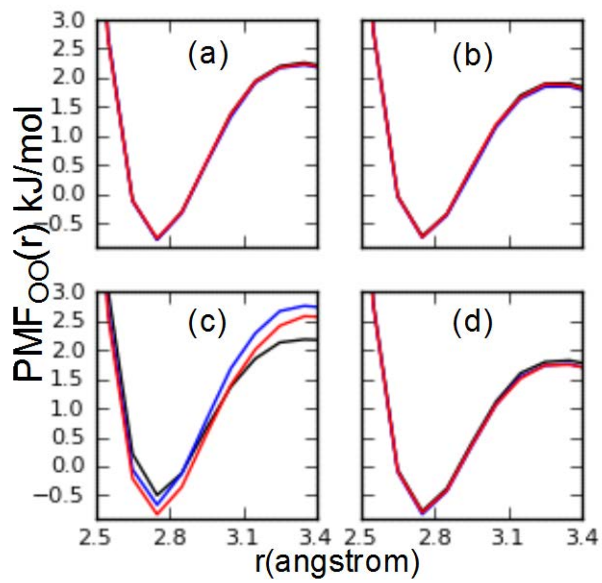

Figure A8. Potential of mean force (PMF) between water oxygen atoms (OW) and the backbone O1 (a, c) or O (b, d) atom the unglycosylated (black), α-glycosylated (blue) or β-glycosylated (red) derivative of Hyp (a, b) or hyp (c, d).

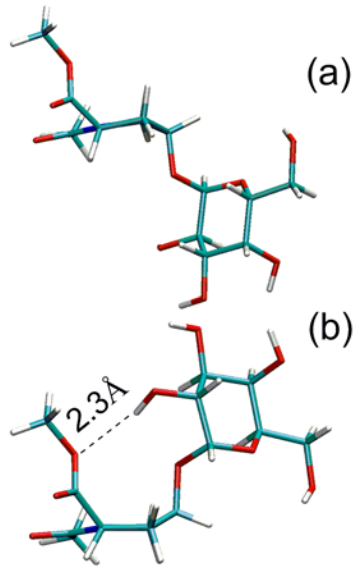

Figure A9. The α_R conformation of the ψ dihedral angle in (a) Hyp–α–Gal and (b) hyp–α–Gal, illustrating the weak contact between the sugar C2 hydroxyl group and the ester oxygen of the peptide in the hyp glycopeptide.

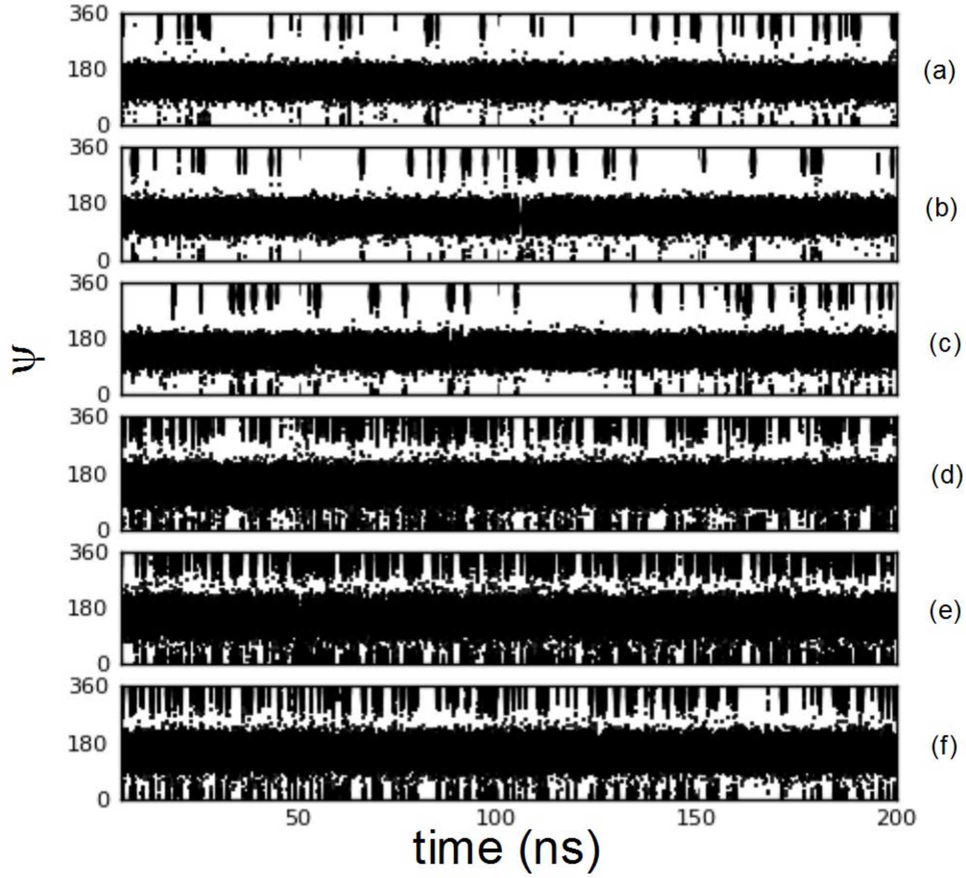

Figure A10. Evolution of the peptide backbone ψ dihedral over a 200 ns MD simulation of (a) Hyp, (b) hyp, (c) Hyp–α–Gal, (d) hyp–α–Gal, (e) Hyp–β–Gal and (f) hyp–β–Gal.

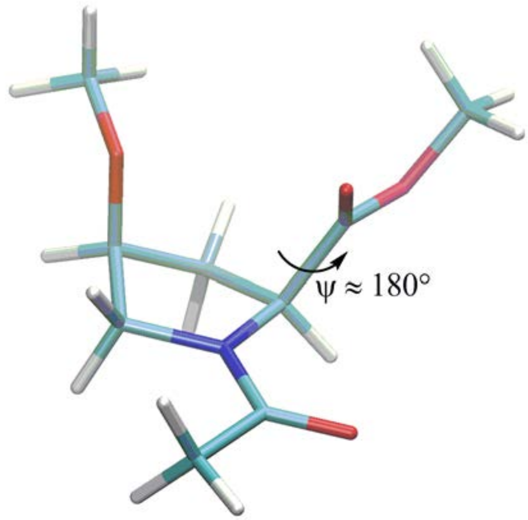

Figure A11. PCM-B3LYP/6-311++G(d,p) optimized structure of *trans* 4(S)-O-methylated hydroxyproline.

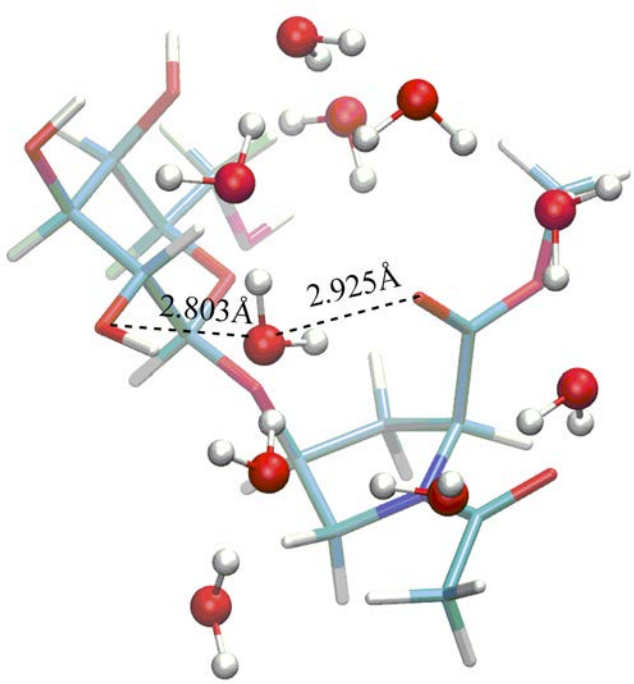

Figure A12. B3LYP/6-31G(d) optimized structure of hyp–β–Gal surrounded by 10 discrete water molecules including a bridging water molecule.

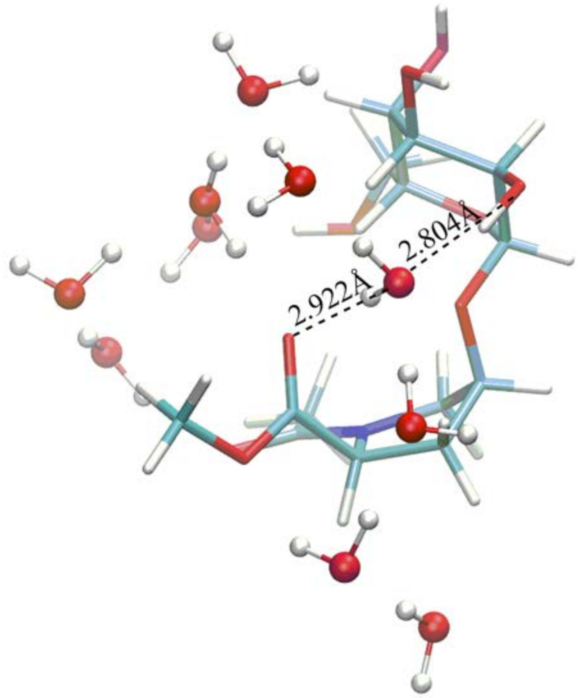

Figure A13. B3LYP/6-31G(d) optimized structure of hyp–α–Gal surrounded by 10 discrete water molecules including a bridging water molecule.

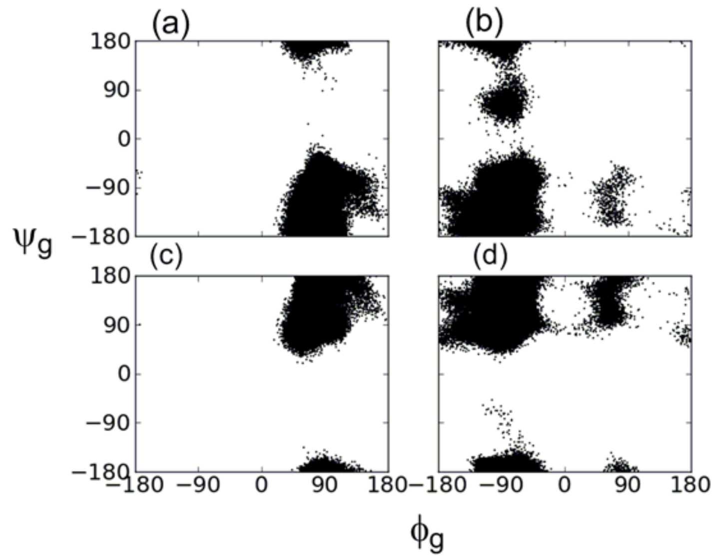

Figure A14. Distribution of glycosidic linkage φ_g and ψ_g dihedral angles (deg.) obtained from the MD trajectory of the *trans* isomer of (a) Hyp–α–Gal, (b) Hyp–β–Gal, (c) hyp–α–Gal and (d) hyp–β–Gal.

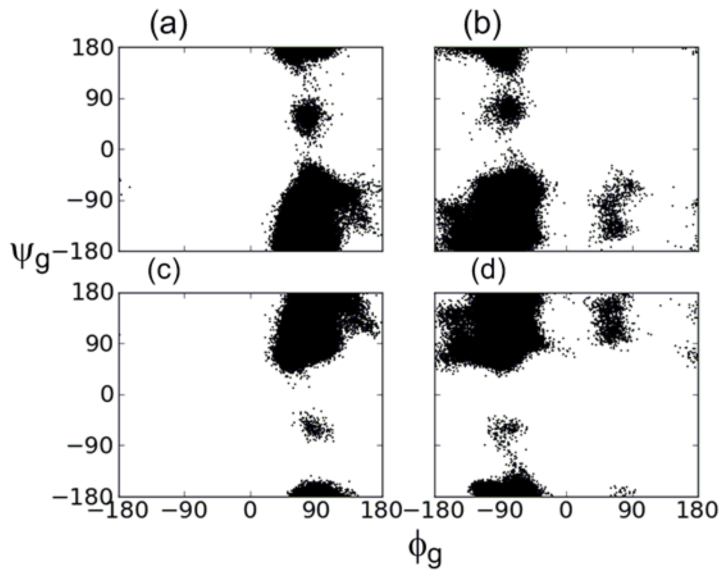

Figure A15. Distribution of glycosidic linkage φ_g and ψ_g dihedral angles (deg.) obtained from the MD trajectory of the *cis* isomer of (a) Hyp–α–Gal, (b) Hyp–β–Gal, (c) hyp–α–Gal and (d) hyp–β–Gal.

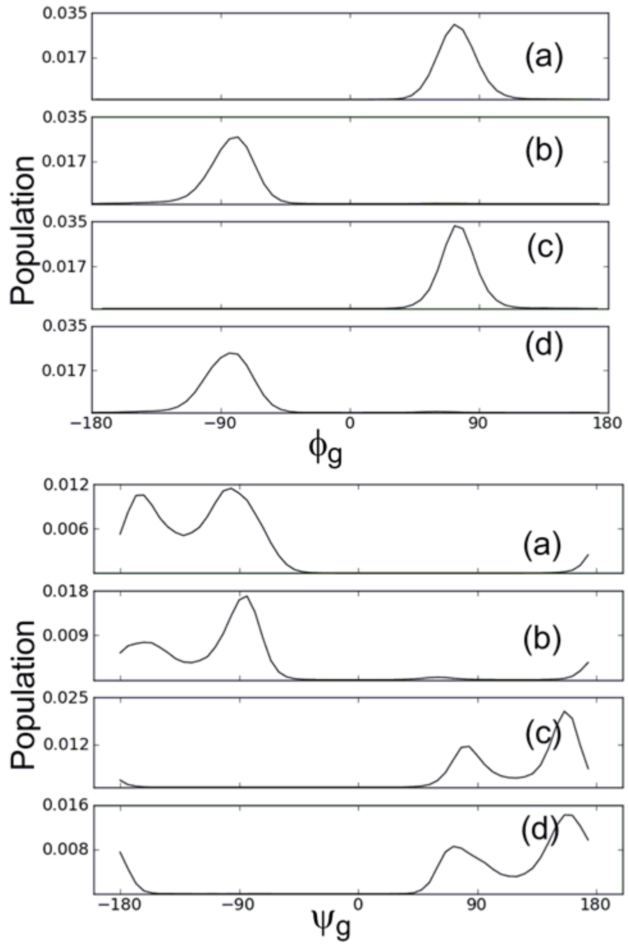

Figure A16. Histograms of the φ_g (top) and ψ_g (bottom) glycosidic linkage dihedral angles (deg.) obtained from the MD trajectory of the *trans* isomer of (a) Hyp–α–Gal, (b) Hyp–β–Gal, (c) hyp–α–Gal and (d) hyp–β–Gal (d).

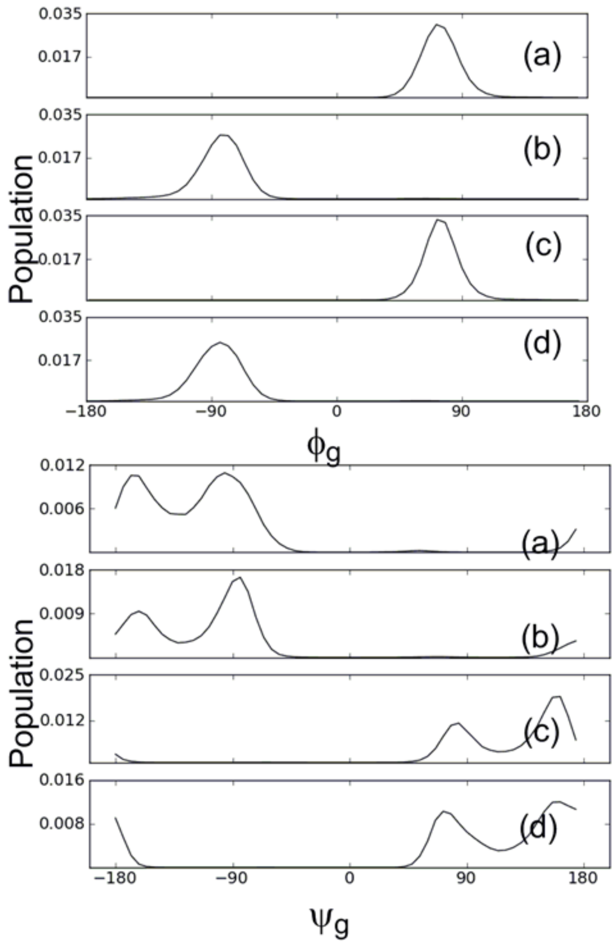

Figure A17. Histograms of the φ_g (top) and ψ_g (bottom) glycosidic linkage dihedral angles (deg.) obtained from the MD trajectory of the *cis* isomer of (a) Hyp−α−Gal, (b) Hyp−β−Gal, (c) hyp−α−Gal and (d) hyp−β−Gal (d).

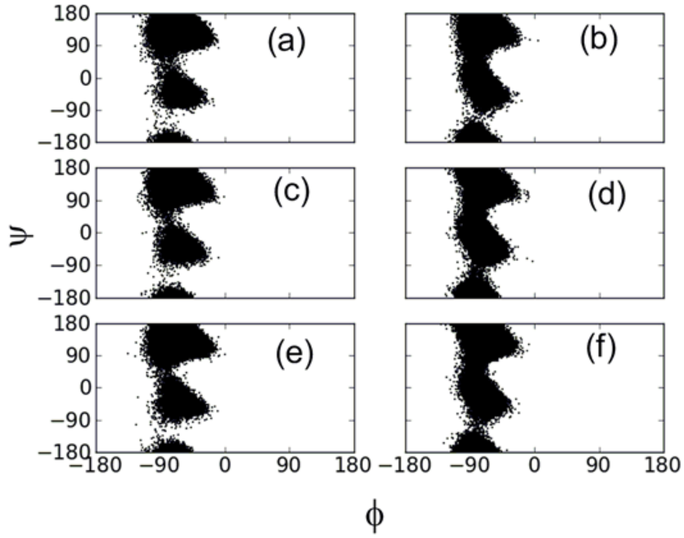

Figure A18. Distributions of peptide backbone φ and ψ dihedral angles (deg.) in the *trans* conformation of (a) Hyp, (b) hyp, (c) Hyp–α–Gal, (d) Hyp–β–Gal, (e) hyp–α–Gal and (f) hyp–β–Gal.

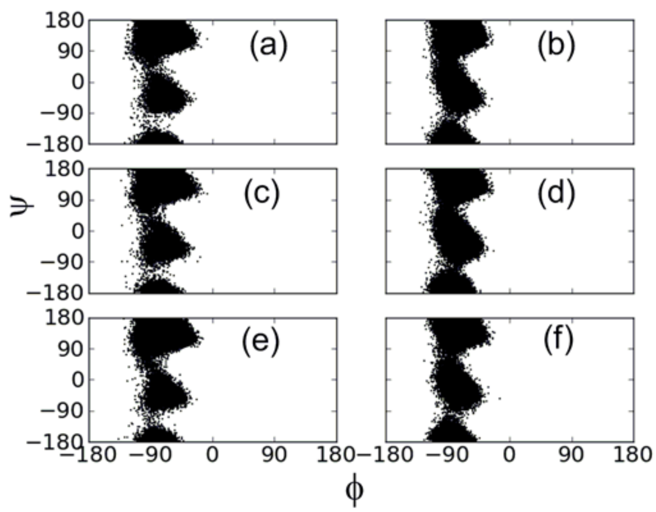

Figure A19. Distributions of peptide backbone φ and ψ dihedral angles (deg.) in the *cis* conformation of (a) Hyp, (b) hyp, (c) Hyp–α–Gal, (d) Hyp–β–Gal, (e) hyp–α–Gal and (f) hyp–β–Gal.

Appendix B

Solvent Interactions Stabilize the Polyproline II Conformation in Glycosylated Oligoprolines

List of Contents

Table B1: Intramolecular hydrogen bonding in Ac-[Hyp-(β-D-Gal)]$_9$-NH$_2$ in the PPII conformation observed during MD simulation at 300 and 552 K.[a,b]

300K			552K		
Acceptor H res@atom	Donor res@atom	% occupied	Acceptor H res@atom	Donor res@atom	% occupied
5@O6'	4@H2O	13.04			
7@O	10@H6O	11.69			
9@O6'	8@H2O	10.73			
2@O	5@H6O	10.54			
8@O6'	7@H2O	10.31			
4@O	6@H6O	9.72			
10@O6'	9@H2O	9.28			
6@O	9@H6O	9.28			
6@O6'	6@H4O	8.85			
1@O	4@H6O	8.57			
3@O	5@H6O	8.55			
5@O	8@H6O	8.08			
4@O6'	3@H2O	8.07			
2@O	4@H6O	8.06			
4@O6'	4@H4O	7.31			
7@O	9@H6O	7.11	7@O	9@H6O	5.15
6@O	8@H6O	6.83			
1@O	3@H6O	6.39			
4@O	7@H6O	6.32			
3@O	6@H6O	6.17			
6@O6'	5@H2O	5.64			
8@O	10@H6O	5.27			
7@O6'	6@H2O	5.12			

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH$_2$) in the 9-mer peptides.

Table B2: Intermolecular hydrogen-bonding interactions in Ac-[Hyp-(β-D-galactose)]$_9$-NH$_2$ in the PPII conformation observed during MD simulations at 300 K.[a,b]

Acceptor H	Donor	
res@atom	res@atom	% occupied
10@O	solvent	144.92
9@O	solvent	124.44
1@O	solvent	119.44
8@O	solvent	105.35
2@O6'	solvent	98.23
2@O	solvent	97.26
3@O6'	solvent	95.81
10@O2'	solvent	94.74
5@O	solvent	94.35
4@O	solvent	94.1
9@O2'	solvent	93.71
8@O2'	solvent	92.79
3@O	solvent	92.46
7@O2'	solvent	91.37
6@O2'	solvent	91.35
4@O2'	solvent	90.4
5@O2'	solvent	90.15
7@O6'	solvent	88.87
2@O2'	solvent	87.92
3@O2'	solvent	87.31
6@O	solvent	87.2
4@O6'	solvent	86.67
10@O6'	solvent	86.57
9@O6'	solvent	85.41
8@O6'	solvent	85.38
6@O6'	solvent	85.19
7@O	solvent	84.62
4@O3'	solvent	84.25
5@O6'	solvent	83.17
3@O3'	solvent	82.95
6@O3'	solvent	82.21
9@O3'	solvent	81.58
8@O3'	solvent	81.56
2@O3'	solvent	81.45
7@O3'	solvent	81.15
10@O3'	solvent	80.98
5@O3'	solvent	79.22

solvent	2@H3O	77.7
solvent	9@H3O	77.6
solvent	3@H3O	77.55
solvent	10@H3O	77.53
solvent	4@H3O	77.26
solvent	8@H3O	77.23
solvent	7@H3O	77.1
solvent	5@H3O	77.09
solvent	10@H2O	76.86
solvent	6@H3O	76.17
solvent	2@H4O	74.84
solvent	2@H2O	74.74
solvent	3@H4O	73.74
solvent	5@H2O	71.97
solvent	7@H4O	71.86
solvent	10@H4O	71.62
solvent	2@H6O	70.85
solvent	6@H2O	70.75
solvent	3@H2O	70.67
solvent	9@H4O	70.21
solvent	4@H4O	70.18
solvent	9@H2O	69.15
solvent	5@H4O	68.87
solvent	7@H2O	68.47
solvent	8@H2O	67.71
4@O4'	solvent	66.95
3@O4'	solvent	66.87
solvent	8@H40	66.05
solvent	4@H20	65.94
2@O4'	solvent	65.46
solvent	6@H40	63.49
10@O4'	solvent	63.48
6@O4'	solvent	63.15
9@O4'	solvent	63.15
5@O4'	solvent	62.9
7@O4'	solvent	62.81
8@O4'	solvent	62.81
solvent	3@H6O	62.65
solvent	7@H6O	58.41
solvent	8@H6O	55.48
solvent	6@H6O	54.19

solvent	4@H6O	53.34
solvent	9@H6O	53.02
solvent	5@H6O	51.45
solvent	10@H6O	50.46
solvent	11@HN1	21.52

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.

Table B3: Hydrogen-bonding interactions in Ac-[Hyp-(β-D-Gal)]$_9$-NH$_2$ in the PPI conformation observed during MD simulation.[a,b]

Acceptor H	Donor	
res@atom	res@atom	% occupied
	Intramolecular	
5@O	6@H6O	85.1
6@O6'	3@H4O	73.78
8@O	9@H6O	70.42
2@O	3@H6O	58.41
9@O6'	6@H4O	57.95
6@O	7@H6O	56.49
7@O6'	4@H4O	35.2
9@O	10@H6O	32.85
3@O	4@H6O	30.65
10@O6'	7@H4O	21.45
5@O6'	2@H2O	21.23
8@O6'	5@H2O	10.83
7@O	8@H6O	10.78
10@O	8@H2O	10.27
3@O4'	2@H6O	9.57
2@O2'	5@H6O	8.92
8@O6'	9@H4O	7.28
5@O6'	2@H3O	5.94
	Intermolecular	
1@O	solvent	144.14
3@O6'	solvent	100.77
4@O6'	solvent	100.49
2@O6'	solvent	94.04
9@O2'	solvent	93.48
10@O2'	solvent	93.34
4@O	solvent	89.53
8@O2'	solvent	89.22
6@O2'	solvent	86.46
7@O2'	solvent	85.8
3@O2'	solvent	85.36
10@O	solvent	84.92
2@O	solvent	84.67
9@O3'	solvent	83.37
8@O3'	solvent	83.18
solvent	5@H4O	82.46
4@O2'	solvent	81.7

10@O3'	solvent	81.66
solvent	3@H3O	79.34
3@O	solvent	79.01
solvent	6@H3O	181
8@O6'	solvent	78.5
solvent	8@H3O	77.99
solvent	10@H3O	77.96
solvent	9@H3O	77.88
2@O3'	solvent	77
solvent	9@H2O	76.81
solvent	4@H3O	75.86
solvent	10@H4O	75.35
solvent	3@H2O	75.14
solvent	10@H2O	74.78
6@O4'	solvent	74.7
solvent	7@H3O	74.52
5@O2'	solvent	74.33
7@O	solvent	73.82
solvent	6@H2O	73.28
7@O3'	solvent	73.01
5@O6'	solvent	72.81
4@O3'	solvent	72.78
solvent	8@H4O	72.25
solvent	4@H2O	72.11
3@O3'	solvent	71.88
6@O3'	solvent	71.76
solvent	2@H4O	71.7
7@O4'	solvent	71.61
solvent	9@H4O	71.53
solvent	2@H3O	71.5
solvent	5@H2O	70.75
4@O4'	solvent	70.56
10@O6'	solvent	69.49
5@O4'	solvent	69.43
solvent	8@H2O	67.04
solvent	7@H2O	66.9
solvent	5@H3O	65.97
10@O4'	solvent	65.86
8@O4'	solvent	64.33
5@O3'	solvent	63.75
2@O4'	solvent	63.54
3@O4'	solvent	62.86

9@O4'	solvent	59.8
solvent	7@H4O	59.77
solvent	8@H6O	56.65
7@O6'	solvent	55
solvent	2@H6O	54.63
9@O	solvent	51.99
solvent	5@H6O	51.56
2@O2'	solvent	50.31
solvent	2@H2O	49.54
solvent	4@H4O	45.1
solvent	4@H6O	42.32
solvent	10@H6O	42.05
9@OD1	solvent	36.19
10@OD1	solvent	35.24
6@O	solvent	34.47
8@OD1	solvent	31.15
2@OD1	solvent	30.94
9@O6'	solvent	30.54
solvent	6@H4O	27.98
solvent	7@H6O	26.5
8@O	solvent	19.49
6@O6'	solvent	17.12
solvent	3@H6O	16.42
solvent	9@H6O	14.48
solvent	3@H4O	14.41
solvent	11@HN1	14.03
4@OD1	solvent	12.72
5@OD1	solvent	11.09
7@OD1	solvent	10.86
5@O	solvent	8.75
solvent	6@H6O	5.42

[a]A hydrogen bond is considered to be present when the donor–receptor distance is 3.5 Å or less, and the donor–donor hydrogen acceptor angle is 120° or greater. [b]Residue numbering is from 1 (Ac) to 11 (NH_2) in the 9-mer peptides.