Computational design of glycopeptides and cyclic peptides

An honors thesis submitted by

Julia R. Rogers

Tufts University

Department of Chemistry

May 2016

Advisor: Professor Yu-Shan Lin

Protein therapeutics are an exceptionally valuable class of drug that typically exhibit higher target specificity and potency compared to their small molecule counterparts. However, structural instability of protein therapeutics often plagues their pharmaceutical development. One way to improve issues including proteolytic degradation, aggregation, and denaturation is through glycoengineering. To fully utilize the benefits of glycosylation in protein design, the effects of glycosylation on protein structure, and subsequent activity, need to be understood to such a degree that glycans can be rationally incorporated into non-glycosylated proteins. Although specific protein-sugar interactions are thought to intrinsically contribute to structural stability, a limited number of interactions between a particular sugar and amino acid have been fully investigated. Computational methods allow efficient examination of sequences with various combinations of different protein-sugar interactions. Since the Protein Data Bank contains over 3,400 glycoprotein structures, structural bioinformatics offers an effective starting point to systematically discern naturally occurring, high frequency patterns of protein-sugar interactions associated with a type of protein structure. Additional computational analyses and experimental investigations of the identified protein-sugar interactions are used to further test conclusions from structural bioinformatics. This molecular level knowledge of how sugars mediate protein structure can further our understanding of the intrinsic biophysical properties of glycoproteins and enable glycosylation to be used to engineer stable protein structures.

Another modification that can improve the stability of peptide therapeutics is cyclization. Furthermore, cyclic peptide therapeutics can target aberrant protein-protein interactions, which are implicated in a variety of diseases including cancer, leukemia, and neurodegenerative diseases, with higher specificity and affinity than their linear peptide counterparts. Although natural cyclic peptide derivatives are under investigation as potential therapeutics, the rational design of cyclic peptide therapeutics remains underexplored since their structures are difficult to predict de novo. Indeed, cyclic peptide structural dynamics vary greatly compared to that of linear peptides: Although the structure of a linear

hexapeptide can be highly flexible, a cyclic hexapeptide typically only adopts a few conformations due to ring strain. Experimentally, cyclic hexapeptides have been shown to adopt structures with two β-turns. With a limited number of conformations, cyclic hexapeptides are ideal models to study how various types of β-turns interchange to produce different cyclic peptide structures. From analysis of molecular dynamics simulations of a highly flexible cyclic hexapeptide, cyclo-(GGGGGG), conformational switches were found to occur via coherent two-dihedral changes. These mechanistic insights into how different cyclic peptide structures interconvert have guided the development of simulation methods to more efficiently sample cyclic peptide structures.

Acknowledgements

Firstly, many thanks are owed to my advisor, Professor Yu-Shan Lin for never ending advice about both my research projects and also future research career. I am grateful that she not only has held me to high standards but has also helped me gain the skills needed to aim even higher. I am additionally grateful for the guidance of Professor Joshua Kritzer and Professor Samuel Thomas, both members of my senior honors thesis committee. I am very much appreciative of the valuable insights they have provided me and the tough questions they have prompted about both my research and future endeavors.

I would also like to thank all of the members of the Lin Lab who have made the past 2+ years truly memorable. I must thank Sean McHugh and Diana Slough in particular. Working with Sean to eventually converge on a complete, final answer to tough biophysical questions, I have learned many important skills from him and gained the friendship of a great scientist. Diana is always ready to answer my questions and assuage my concerns about current challenges with research and about future ones as I continue onto graduate school. Indeed, I look forward to seeing all she accomplishes upon obtaining her PhD. Additionally, many thanks are owed to Dr. Hongtao Yu for his technical expertise and help.

Finally, I want to thank the collaborating scientists who have aided the work described below. Indeed, the preliminary experimental characterization we performed would not have been possible without the lab of Matthew Shoulders at MIT and the guidance of Dr. Emmanuel Nekongo. Furthermore, thanks are also given to the collaborating lab of Joshua Price at Brigham Young University as we continue further wet-chemistry experiments.

Table of Contents

Chapter 1 – Rational design of glycosylated α-helices

1.1 – An introduction to the effects of glycosylation on protein structure

The structure of a protein is principally attributed to its primary sequence. Additionally, covalent modifications further enrich the diversity of protein chemistries, structures, and functions.[1] One of the most common modifications, N-linked glycosylation is naturally found on more than 50% of all proteins.[2,3] N-linked glycosylation occurs co-transnationally in the endoplasmic reticulum, creating a β-linkage between the core N-acetylglucosamine (GlcNAc) of the glycan to the nitrogen of the asparagine (Asn) side chain (**Figure 1.1**). N-linked glycosylation generally occurs in a conserved Asn-Xxx-Ser/Thr (serine/threonine) motif, where Xxx is any amino acid except proline (Pro) (**Figure 1.1**).[4] Glycosylation impacts not only the biological functions of glycoproteins[5-8] but also their folding efficiencies,[9-12] structures,[13-17] dynamics,[18,19] and stabilities.[20-23] N-glycosylation can intrinsically enhance protein stability and folding rate through, for instance, excluded volume, hydrophobic effects, as well as specific protein-sugar interactions.[8,11,24-27] However, the biophysical details underlying these characteristics of glycosylation have not been fully elucidated. A more complete molecular level understanding of the effects of glycosylation will enable the rational design of novel glycopeptide structures[22,27,28] and stable protein therapeutics.[29-31]

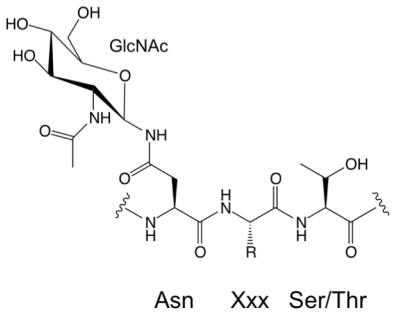

Figure 1.1 Chemical structure of N-glycosylated sequon.

Previous studies have provided multiple insights into the effects of glycosylation, including the significance of protein-sugar interactions on β-turn structures.[27,28,32] For example, glycosylation helps stabilize the human CD2 glycoprotein through essential protein-sugar interactions.[21,27] Properly folded CD2 contains a cluster of five positively charged lysines (Lys) that is destabilizing

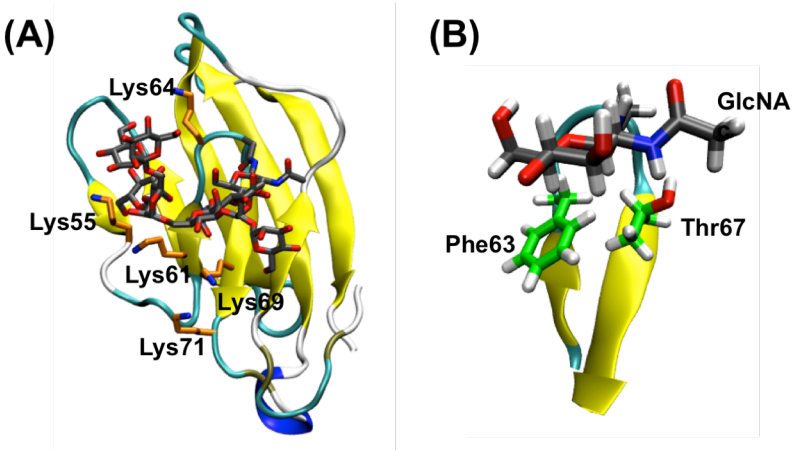

Figure 1.2 (A) Structure of human CD2 (PDB ID 1GYA) with the clustered Lys shown in orange and the glycan in gray. **(B)** The type I β-bulge in human CD2 stabilized by interactions with the core GlcNAc (shown in gray), Phe, and Thr (shown in green).

without the presence of the glycan to mediate their coulombic repulsion (**Figure 1.2A**).[21] Furthermore, the core GlcNAc attached to Asn at position *i* of a type I β-bulge, is involved in CH-π interactions with phenylalanine (Phe) at the *i*–2 position. The formation of a compact, hydrophobic core consisting of GlcNAc, Phe, and Thr at position *i*+2 additionally stabilizes the β-bulge motif (**Figure 1.2B**).[27] These specific protein-sugar interactions have been successfully engineered into non-glycosylated proteins to stabilize β-turns.[22,28,32]

N-linked glycans are most often found on turns, bends, and sites of secondary structure change.[33-36] Glycosylation sites are not as commonly found on α-helices, and glycosylation is usually considered α-helix breaking.[23,37-39] The close proximity of the side chains on α-helices allows for little tolerance of a bulky glycan. For example, incorporation of glycosylation sites into the center of α-helices in the bacterial immunity protein Im7 was destabilizing. In contrast, glycosylation sites introduced at the termini of those α-helices on Im7 showed no significant effect on stability.[37] Glycans likely require increased flexibility of nearby protein residues and reduced steric constraints to be accommodated without disrupting protein structure; α-helices do not typically provide such flexibility and, thus, are typically destabilized by glycosylation.

Despite evidence that glycosylation destabilizes α-helices, there are examples of naturally glycosylated helices and instances of proteins maintaining α-helicity upon glycosylation.[40-43] Therefore, we hypothesize that specific sequences form relatively stable α-helical structures when glycosylated. First, we predict the possible sugar orientations amenable with α-helical structure based on purely steric constraints. Then, we utilize structural bioinformatics to investigate protein-sugar interactions characteristic of α-helices naturally glycosylated according to the Swiss-Prot protein sequence database. From bioinformatics analysis, two modes of protein-sugar interactions, consistent with the orientations predicted from steric analysis, are identified for α-helical glycosylation sites. Finally, these two interaction modes are preliminarily investigated with circular dichroism experiments of a model α-helical peptide.

1.2 – Predictions of GlcNAc orientations on α-helices

Methods to evaluate steric constraints imposed on the glycan by α-helical structure. Because some effects of glycosylation, including α-helix destabilization, can be attributed to steric crowding, model α-helical glycopeptides were built and evaluated for steric clashes to determine allowable glycan orientations on helices. Ideal α-helices (ϕ = –57° and ψ = –47°) of 21 residues were made with Chimera,[44] and the central residue was glycosylated with the Glycam glycoprotein builder.[45] Different glycan orientations were specified by the dihedral angles, χ_1, χ_2, ψ_N, ϕ_N, ω_2, and ω_6. Definitions of these angles and the values used to generate different glycoprotein structures are given in **Figure 1.3**. The values used for χ_1 and χ_2 were selected based on the conformers of Asn observed in rotamer libraries, which catalogue the prevalence of a given side chain conformation, specified as a combination of dihedral angles, for each amino acid.[46-48] In comparison to the distinct conformers observed for χ_1 in these libraries, the χ_2 distribution does not exhibit specific preferences and is widely distributed about the N*t* conformer (χ_2 = 180°) since it involves the rotation of a planar amide relative to the tetrahedral Cβ.[47,49] Because ψ_N characterizes the amide bond between Asn and GlcNAc, the *trans* and *cis* conformers were used.

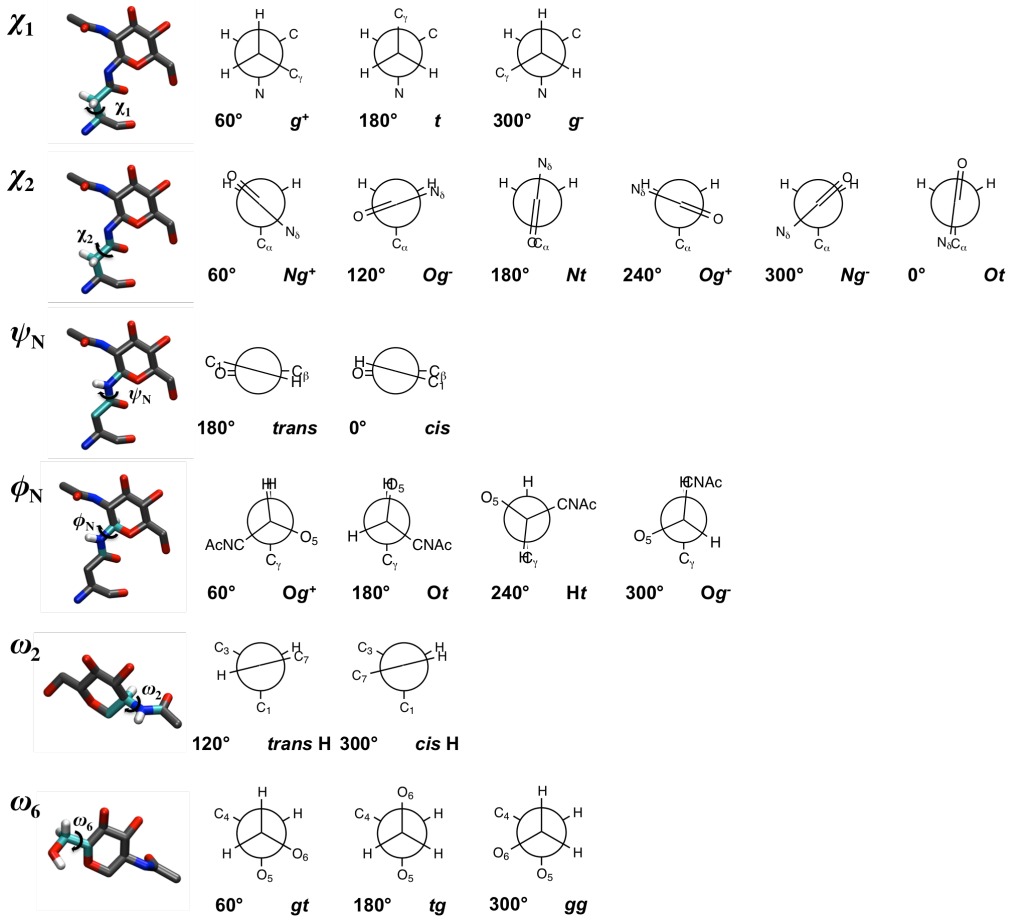

Figure 1.3 Dihedral angles used to characterize GlcNAc orientation. The atoms used to define each dihedral are colored cyan. The values for each dihedral used to create model sturctures and the corresponding Newman projections are shown.

Conformers of ϕ_N found intermediate between the O*t* (ϕ_N = 180°) and Og^- (ϕ_N = 300°) conformers are likely preferable for GlcNAc because unfavorable sterics between the N-acetyl group on C2 and the Asn side chain are limited. Indeed, an intermediate conformer with ϕ_N = 270° has previously been observed to be the most populated conformation.[33,50] The conformers used for ω_2 and ω_6 were chosen based on previous bioinformatic[50] and experimental results.[51] Combinations of all dihedrals in **Figure 1.3** resulted in a total of 864 model α-helical glycoproteins. Clashes, or overlap of van der Waals radii, of the peptide backbone with the Asn side chain and GlcNAc were calculated with Chimera's Find Clashes to identify conformations that would be sterically forbidden for α-helical glycopeptides.

Glycan conformations sterically allowed on α-helices. Of the 864 non-glycosylated structures, 594 have clashes between the Asn side chain and the α-helix backbone. For α-helical Asn, $\chi_1 = 60°$ is known to be disfavored because it can result in steric hindrance with the proceeding turn of the α-helix.[47] Indeed, non-glycosylated ideal α-helices with $\chi_1 = 60°$ have clashes (**Figure 1.4A**). The majority of model helical glycopeptides have steric clashes (661 out of 864 structures). Fewer conformations are allowed for glycosylated helices relative to non-glycosylated helices, supporting the general consensus that glycosylation is often poorly tolerated on α-helices.[23,37-39] Similar to non-glycosylated structures, all glycoprotein structures with $\chi_1 = 60°$ have clashes (**Figure 1.4B**). In addition to clashes with the Asn side chain, 257 of the 288 glycoproteins with $\chi_1 = 60°$ also have clashes with GlcNAc. Thus, natural α-helical glycosylation sites are predicted to have either $\chi_1 \approx 180°$ or $\chi_1 \approx 300°$ to accommodate a glycan (**Figure 1.4C**).

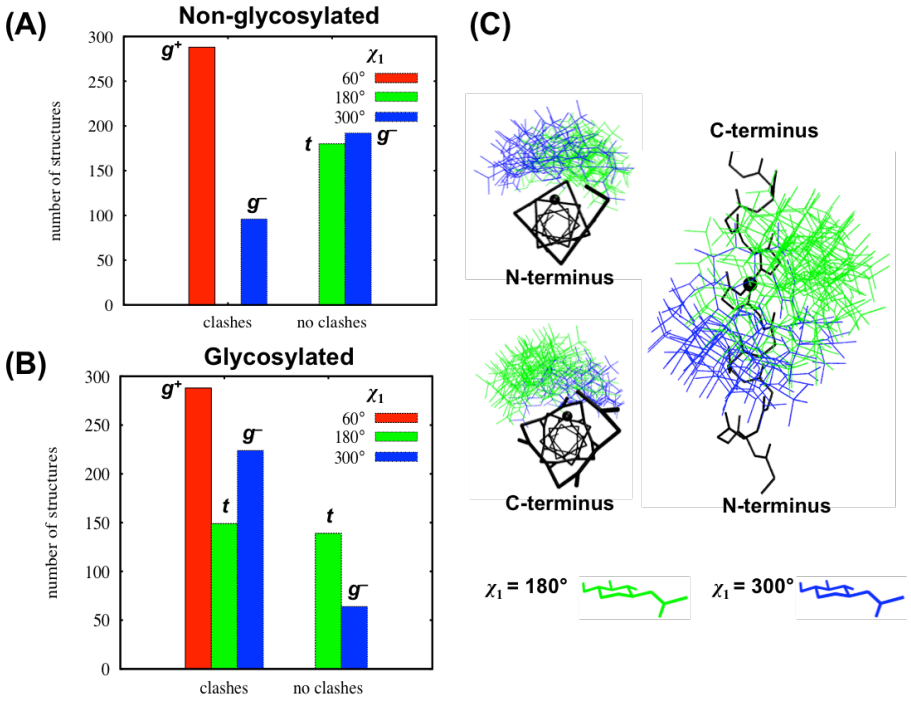

Figure 1.4 (A) Number of model structures with clashes or no clashes between Asn and the helix backbone grouped according to χ_1. **(B)** Number of model glycosylated structures with clashes or no clashes between GlcNAc and the Asn side chain with the helix backbone grouped according to χ_1. **(C)** Alignment of model structures without any clashes. Top left, view down the helix axis from the N-terminus. Bottom left, view from the C-terminus. Right, view looking down onto the sugar. The peptide backbone is colored black with a sphere marking the Asn C_α.

1.3 – Methods for structural bioinformatics analysis

Glycoprotein data set. Glycoproteins were first identified in the non-redundant, manually annotated Swiss-Prot sequence database (accessed December 2, 2014).[52] Glycosylation sites were identified from a feature line key of 'CARBOHYD' excluding sites with the 'glycation' modifier to include only natural glycosylation sites. Swiss-Prot entries were then mapped to Protein Data Bank (PDB) structures (accessed December 15, 2014).[53] A total of 3,402 glycoprotein PDBs were obtained. PDB sites previously identified from the Swiss-Prot sequences were then verified to be glycosylated in the PDB by checking the CONNECT record for Asn to be β-linked to GlcNAc (residue name NAG in the PDB). PDBs with unreasonably long bond lengths between sugars in a glycan or between the core sugar and glycosylated Asn were excluded from further analysis. These errors were verified with pdb-care, an online tool designed to check that sugar residue names and CONNECT records agree with the sugar structure in the PDB coordinates.[54] To ensure accurate evaluations of interactions between the protein and core GlcNAc in later analysis, PDB sites lacking atoms from either the core GlcNAc or an amino acid with atoms within 5 Å of the core GlcNAc were also excluded from further analysis. To avoid bias from redundant PDBs, each PDB site was weighted accordingly so that sites identified from the same Swiss-Prot site were altogether only counted as one glycosylation site in analyses. The final dataset consists of 1,961 weighted N-glycosylation sites.

Protein secondary structure determination. The secondary structure of each protein was determined with both DSSP[55] and STRIDE[56] algorithms. Residues classified as α-helical by both DSSP and STRIDE are considered α-helical. To investigate glycosylation sites at the center of α-helices, glycosylation sites on helical Asn's with at least four additional α-helical residues towards both termini were identified and will be referred to as 9-residue helices (**Figure 1.5**). Helical glycosylation sites with at least four α-helical residues towards the N-terminus (furthermore termed N-terminal helices) and sites with at least four α-helical

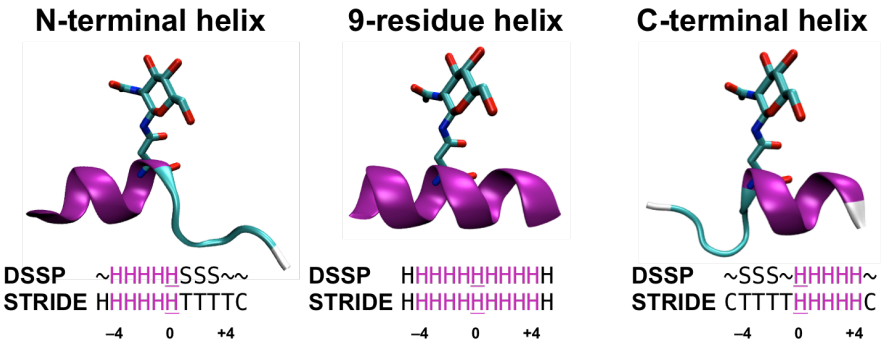

Figure 1.5 Illustration of N-terminal, 9-residue, and C-terminal helix classifications. DSSP and STRIDE secondary structure assignments used to determine each classification are given. Residues considered in helix classification are colored purple. Underlined residues are glycosylated. Secondary structure assignment for DSSP: H = α-helix, S = bend, ~ = not specified structure. Secondary structure assignment for STRIDE: H = α-helix, T = turn, C = coil.

residues towards the C-terminus (termed C-terminal helices) were also analyzed for protein-sugar interactions that would allow an α-helix to be maintained either N-terminal to the glycosylation site or C-terminal, respectively. It is noted that glycosylation sites classified as 9-residue helices are also classified as both N-terminal and C-terminal helices. These three structural classifications are illustrated in **Figure 1.5**.

Analysis of GlcNAc orientation. The orientation of the core, N-linked GlcNAc along the α-helix can be characterized by the Asn side chain dihedral angles χ_1 and χ_2. To identify common GlcNAc orientations, dihedral principal component analysis (dPCA)[57,58] of χ_1 and χ_2 and subsequent clustering was performed. The principal components obtained from dPCA describe the main underlying differences between structures; the first principal component (PC1) captures the widest variance of the structures, and the second principal component (PC2) captures the second widest variance orthogonal to the first. Following dPCA, a grid-based density peak-based clustering[59] in the 2D principal subspace was utilized to identify common orientations of GlcNAc relative to the α-helix and to quantify the number of sites with a given GlcNAc orientation. The results of this analysis are compared to the predictions based on steric arguments presented in section 1.2 – Predictions of GlcNAc orientations on α-helices.

Evaluation of protein-sugar interactions. To identify the most favorable protein-sugar interactions for a given secondary structure, the interaction energy of each amino acid within 3.5 Å of the core GlcNAc was calculated with GROMACS 4.6.1.[60] Due to the lack of hydrogen in many PDB structures, hydrogen was first added to all residues within 3.5 Å of the sugar, the glycosylated Asn, and the core GlcNAc. The positions of the hydrogen were then refined through steepest descent energy minimization. The interaction energy between each individual amino acid and the core GlcNAc was calculated as a difference between the free energy of the dimer and the sum of the sugar and amino acid monomers. For the calculation, parameters from the Charmm22* force field[61] for the amino acids and from the Charmm36 force field[62] for GlcNAc were utilized. To take into the account effects of water molecules, which are not always included in PDB coordinates, Generalized Born with solvent accessible surface area (GBSA) implicit solvent[63] was utilized in the energy calculations. Protein-sugar interactions with negative interaction energies are considered to be favorable interactions.

1.4 – Common GlcNAc orientations on α-helices

Out of the 1,961 N-glycosylation sites, fifty 9-residue helices were found. Two common orientations of GlcNAc on these 9-residue helices were identified from dPCA of χ_1 and χ_2 and subsequent density peak-based clustering (**Figure 1.6**). These two clusters have similar populations (28.1 sites in cluster 1 and 21.9 sites in cluster 2)[†], suggesting the two orientations are equally preferred on α-helical glycoproteins. In both clusters, GlcNAc is generally aligned atop the helix axis. The two clusters differ in the direction which GlcNAc is angled towards, either towards the N-terminus of the helix (cluster 1) or the C-terminus (cluster 2). These spatial differences result from different χ_1 conformers: Cluster 1 has a χ_1 distribution centered about 280°, nearly the g^- conformer; cluster 2 has a χ_1

[†] Weighted counts for a given classification can be non-whole numbers since all PDB sites mapped to the same Swiss-Prot sequence may not have the exactly the same secondary structures of residue position. For this very reason (i.e. because we wanted to evaluate all possible structures for a given sequence), weighting was performed.

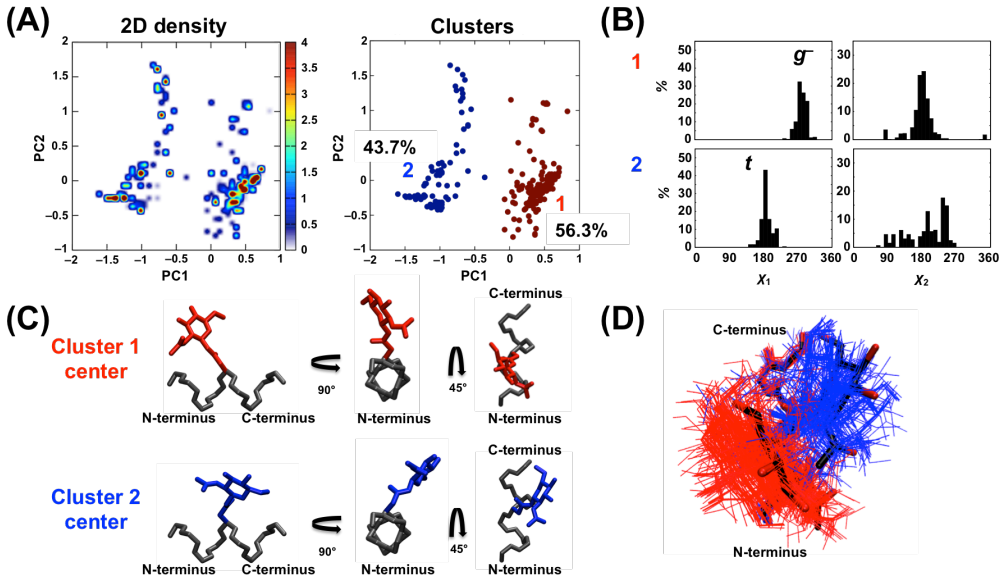

Figure 1.6 **(A)** 2D density plot and cluster assignments in principal component space from dPCA of glycosylated 9-residue helices. **(B)** Distributions of χ_1 and χ_2 for both clusters. **(C)** Structure of each cluster center shown from the side, down the helix axis from the N-terminus, and looking down onto the sugar. **(D)** Alignment of all 9-residue helices shown from above looking down onto the sugar with the helix axis perpendicular. Structures in cluster 1 are colored red, cluster 2 are blue. The two cluster centers are shown in black.

distribution centered about 180°, the *t* conformer. Both clusters have similar average χ_2 values around 187°, but cluster 2 has a wider, flatter distribution than cluster 1. Thus, the two different orientations result primarily from different χ_1 conformers. Consistent with predictions based on solely steric effects (see the above section 1.2 – Predictions of GlcNAc orientations on helices), no glycosylation sites with $\chi_1 = 60°$ are observed on 9-residue helices.

1.5 – Protein-sugar interactions of α-helical glycoproteins

To identify the protein-sugar interactions that could locally stabilize the α-helix and be used in helical glycopeptide design, we focus our analysis on sequential interactions as opposed to spatial interactions arising from tertiary contacts (**Figure 1.7**). Sequential protein-sugar interactions involve residues from position –10 to +10 around the glycosylated Asn at position 0 in the sequence. Since we identified two common GlcNAc orientations for α-helical glycosylation sites (mode #1 with $\chi_1 = 280°$ or mode #2 with $\chi_1 = 180°$) that spatially place the sugar at opposite ends of the α-helix, we thus hypothesize that favorable protein-

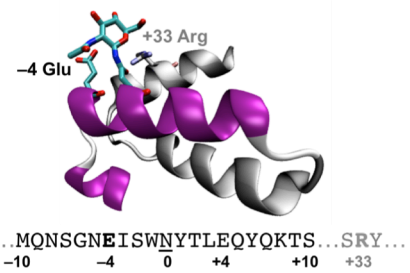

Figure 1.7 Example of a sequential interaction (–4 Glu shown in cyan), which involves a residue within a 10 residue radius of the glycosylated Asn in the sequence (resides from –10 to +10 positions are colored by secondary structure assignment according to STRIDE), compared to an example of a spatial interaction (+33 Arg shown in gray), which involves a residue outside of that 10 residue radius (residues before –10 and after +10 positions are shown in gray). (PDB ID 4KX8).

sugar interactions will utilize one of these two modes. Additionally, the majority (60%) of the 9-resiude helices with only sequential interactions (11.2 sites have only sequential interactions out of 50.0 sites that are 9-residue helices) have only one protein-sugar interaction; from structural bioinformatics analysis, only one protein-sugar interaction mode can be investigated on a single static structure. To evaluate sequential protein-sugar interactions likely to utilize mode #1, in which the sugar is orientated towards the N-terminus of the α-helix relative to the glycosylation site, we focus our analysis on N-terminal helices; to evaluate interactions likely to utilize mode #2, in which the sugar is orientated towards the C-terminus of the helix relative to the glycosylation site, we focus on C-terminal helices (**Figure 1.4C**) (for definitions of N-terminal and C-terminal helices see "Protein secondary structure determination" in the section 1.3 – Methods for structural bioinformatics analysis).

Protein-sugar interactions of N-terminal helices. Of the 120.3 glycosylated N-terminal helices we identified, 47% of those sites have sequential residues within 3.5 Å of GlcNAc. The majority (71%) of sites with only sequential interactions have only one protein-sugar interaction. Consistent with results obtained for glycosylated 9-residue helices, only one protein-sugar interaction likely mediates α-helical structure N-terminal to the glycosylation site. Evaluation of the interaction energy for sequential interactions sorted by position of the interacting residue indicated that the most favorable interaction (interacting residue that most commonly has a negative interaction energy in contrast to a positive interaction energy) occurs with a residue at the –4 position (**Figure**

1.8A). Due to the *i*–4 and *i*+4 hydrogen bonding pattern of α-helices, residues at the –4 position are aligned along the same side of the α-helix as GlcNAc and poised to interact. Of the interactions between GlcNAc and –4 position residues, interactions with –4 glutamate (Glu) are the most favorable (**Figure 1.8B**). Additionally, these glycosylated N-terminal helices with a –4 Glu within 3.5 Å of GlcNAc have a distinct preference for mode #1 with χ_1 = 280° (**Figure 1.9**). With mode #1, the sugar is oriented towards the N-terminus of the α-helix and Glu at position –4. Favorable –4 Glu-GlcNAc interactions are electrostatic in nature, involving H-bonding between a carboxylate oxygen of the Glu side chain and the HN of the GlcNAc N-acetyl group (**Figure 1.10**). Therefore, –4 Glu-GlcNAc interactions may help maintain α-helical structure N-terminal to the glycosylation site through interaction mode #1.

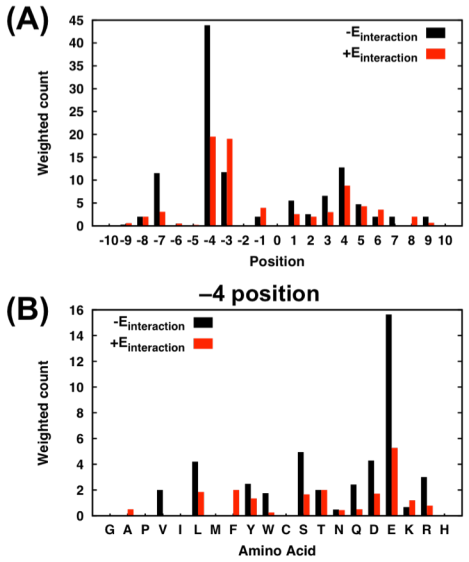

Figure 1.8 Number of N-terminal helices with negative or positive interaction energies **(A)** according to sequence position of the interacting residue and **(B)** at the –4 position.

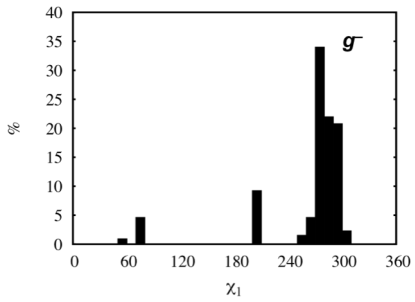

Figure 1.9 χ_1 distribution for N-terminal helices with –4 Glu within 3.5 Å of GlcNAc.

Example structures with –4 Glu-GlcNAc interactions are shown in **Figure 1.10**. Studies of the glycoproteins shown in **Figures 1.10A** and **1.10B** have illustrated the biological advantages conferred by glycosylation, prompting the question: is the –4 Glu-GlcNAc interaction we identified on each of these glycoproteins also important to obtain these beneficial effects of glycosylation? **Figure 1.10A** illustrates a glycosylation site on a phytase, an enzyme that

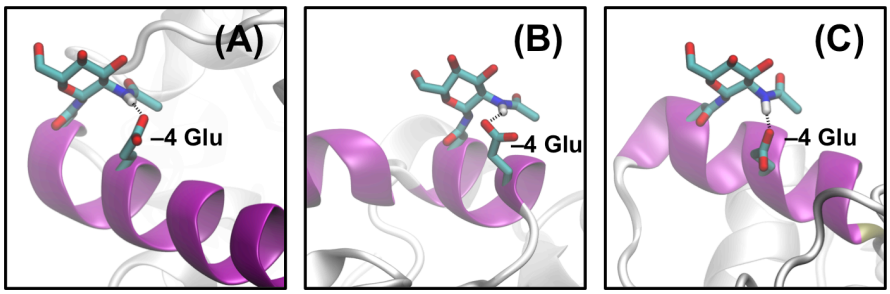

Figure 1.10 Examples of N-terminal helices with –4 Glu-GlcNAc interactions. **(A)** A glycosylation site on a phytase (PDB ID 3K4P). **(B)** A glycosylation site on phospholipase B-like protein 1 (PBD ID 4BWC). **(C)** A glycosylation site on formylglycine-generating enzyme (PDB ID 2AFY).

catalyzes the hydrolysis of phytate to release stored phosphorous. Because phytase activity is limited in the gastrointestinal tracts of simple-stomached animals such as swine and poultry, the design of thermostable phytases for inclusion in animal feed is of much interest.[64,65] The glycans on phytases have been shown to increase their thermostability and to affect the optimum pH for catalytic activity.[66,67] Another example, phospholipase B-like protein 1 (PLBD1), a lysosomal protein that aids macromolecular degradation, is shown in **Figure 1.10B**. Improper lysosomal protein function can lead to a number of inherited lysosomal storage disorders. Transporting the necessary proteins to the lysosome is thus key; typically transport occurs via the mannose-6-phosphate dependent pathway. The glycosylation site specifically shown in **Figure 1.10B** is conserved among the families of PBLD1 and a PBLD1 paralog and is crucial for proper transport of PLBD1 to the lysosome.[68] To see if the –4 Glu is also conserved, Swiss-Prot was searched for proteins of the phospholipase B-like family, and the sequences were aligned with Clustal Omega 1.2.1[69]. Of the 13 proteins with this conserved glycosylation site, nine also contained a –4 Glu (**Table 1**). It is interesting to consider the possibility that both –4 Glu and the glycosylation site necessary for proper protein localization may have been conserved to help maintain the structure of the PLBD1 and lead to proper protein localization. In light of our structural bioinformatics identification of key –4 Glu-GlcNAc interactions on α-helices, mutational studies of the –4 Glu on both phytases and PLBD1 could further investigate the importance of our conclusions and also

Table 1. Sequence alignment of PLBD1 family in Swiss-Prot around conserved glycosylation site (bolded and underlined). –4 Glu is bolded and colored red. Swiss-Prot accession code Q6P4A8 is shown in **Figure 1.10B**.

Swiss-Prot accession code	Descriptor	Sequence	Last residue number
Q8VCI0	PLBL1 mouse	YWASYNIPFHKTIY**N**WSGYPL	418
Q8NHP8	PLBL2 human	YWASYNIPSF**E**TVF**N**ASGLQA	442
F8S101	PLB *Crotalus adamanteus*	YWPSYNIPFDKEIY**N**MSGYGE	422
Q9GL30	PLBL1 bovine	YWPSYNIPFH**E**KVY**N**WSGYPI	414
Q5U2V4	PLBL1 rat	YWASYNIPFHKTVY**N**WSGYPL	418
Q4QQW8	PLBL2 rat	YWASYNIPYF**E**SVF**N**ASGLQA	438
Q55BJ6	PLBLB *Dictyostelium discoideum*	YWPSYNIPYF**E**TIF**N**MSGFND	433
Q6P4A8	PLBL1 human	YWPSYNVPFH**E**KIY**N**WSGYPL	417
Q3TCN2	PLBL2 mouse	YWASYNIPYF**E**TVF**N**ASGLQA	447
Q55FN1	PLBLG *Dictyostelium discoideum*	YWKSYNIPYF**E**SIY**N**ISGYSE	456
Q2KIY5	PLBL2 bovine	YWASYNIPSF**E**SVF**N**ASGLPA	442
Q54M94	PLBLC *Dictyostelium discoideum*	YWPSFNVPYF**E**TIS**N**MSGFNY	412
F8J2D3	PLB *Drysdalia coronoides*	YWPSYNIPFHKVIY**N**MSGYRE	422

provide further insights into the effects of glycosylation on these proteins. The effects of glycosylation on the formylglycine-generating enzyme (FGE) shown in **Figure 1.10C** have not been investigated. FGE oxidizes a conserved cysteine residue in all sulfatases. Reduction of sulfatase activity results in severe diseases such as mucopolysacchridoses and X-linked ichthyosis. Elimination of FGE activity leads to the fatal disease multiple sulfatase deficiency, the severity of which is determined by FGE instability.[70-72] Since glycans can increase protein stability and in light of our findings about –4 Glu-GlcNac interactions on α-helices, it could be interesting to investigate if any disease phenotypes are connected to mutations of glycosylation and/or the –4 Glu.

Protein-sugar interactions of C-terminal helices. Next, we looked for sequential interactions on glycosylated C-terminal helices to detect interactions that might locally stabilize the portion of the α-helix C-terminal to the glycosylation site. Of the 77.3 glycosylated C-terminal helices we identified, 48% of those sites have sequential residues within 3.5 Å of GlcNAc. Evaluation of the interaction energy for sequential interactions sorted by position of the interacting residue is shown in **Figure 1.11A**. In comparison to the number of protein-sugar interactions found on N-terminal helices (**Figure 1.8A**), fewer protein-sugar

interactions are found on C-terminal helices. All 15.6 structures with interactions between –4 position residues and GlcNAc are also helical towards the N-terminus relative to the glycosylation site and classified as 9-residue helices. The most favorable interaction C-terminal to the glycosylation site occurs with a residue at the +4 position (**Figure 1.11A**). –4 interactions locally mediate helical structure N-terminal to the glycosylation site, whereas +4 interactions locally mediate helical structure C-terminal to the glycosylation site. Again due to the hydrogen bonding pattern of α-helices, residues at the +4 position are properly positioned to favorably interact with GlcNAc. Of the interactions between GlcNAc and +4 position residues, interactions with +4 Glu are the most favorable (**Figure 1.11B**). Favorable +4 Glu-GlcNAc interactions involve a carboxylate oxygen of the Glu side chain and the HN of the GlcNAc N-acetyl group, similarly to –4 Glu-GlcNAc interactions observed on glycosylated N-terminal helices. In contrast to N-terminal helices with a –4 Glu-GlcNAc interaction, these glycosylated C-terminal helices with a +4 Glu within 3.5 Å of GlcNAc have a distinct preference for mode #2 with χ_1 = 180° (**Figure 1.12**). With mode #2, the sugar is oriented towards the C-terminus of the helix and Glu

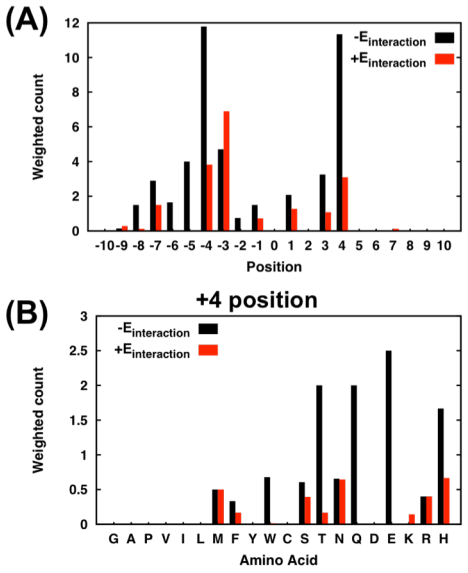

Figure 1.8 Number of C-terminal helices with negative or positive interaction energies **(A)** according to sequence position of the interacting residue and **(B)** at the +4 position.

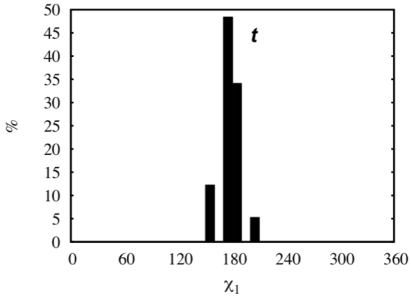

Figure 1.9 χ_1 distribution for C-terminal helices with +4 Glu within 3.5 Å of GlcNAc.

at position +4. Therefore, +4 Glu-GlcNAc interactions may help maintain α-helical structure C-terminal to the glycosylation site through interaction mode #2.

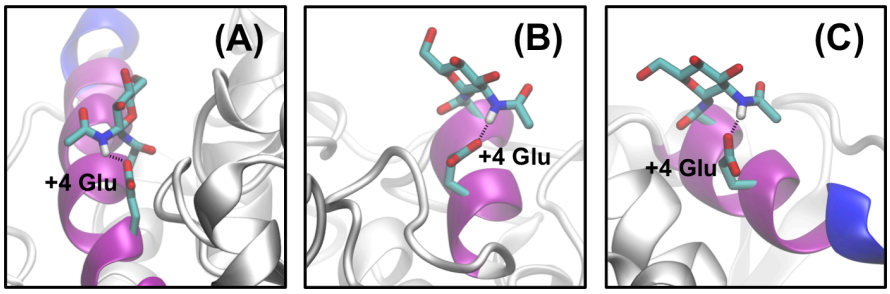

Figure 1.13 Examples of C-terminal helices with +4 Glu-GlcNAc interactions. **(A)** A glycosylation site on angiotensin-I-converting enzyme (PDB ID 2C6N). **(B)** A glycosylation site on the α isoform of the human folate receptor (PBD ID 4LRH). **(C)** A glycosylation site on neprilysin (PDB ID 1DMT).

Example structures with +4 Glu-GlcNAc interactions on C-terminal helices are shown in **Figure 1.13**. A glycosylation site on angiotensin-I-converting enzyme (ACE) is shown in **Figure 1.13A**. ACE regulates blood pressure by cleaving the tensor peptides angiotensin I and bradykinin and is also a therapeutic target for cardiovascular and renal disease. The glycans on ACE are important for its stability and activity.[73,74] The particular glycosylation site shown in **Figure 1.13A** is one of two glycans that most significantly contribute to the thermal stability of the enzyme.[73] Another example in which glycans are important for proper protein expression, cellular localization, and activity is the α isoform of the human folate receptor (αFR) (**Figure 1.13B**). Folate receptors mediate cellular uptake of folate, or vitamin B, and are overexpressed in cancer cells. All three consensus glycosylation sites, including the site illustrated in **Figure 1.13B**, play roles in expression efficiency and help maintain the enzyme's active conformation.[75,76] Similarly to the effects of glycosylation observed for αFR, altered glycosylation of neutral endopeptidase, also known as neprilysin, (NEP) also affects its transport to the cell surface, protein stability, and enzyme activity (**Figure 1.13C**). NEP is a zinc metallopeptidase that degrades small regulatory peptides. NEP also cleaves Aβ, linking the enzyme to the pathogenesis of Alzheimer disease and hereditary inclusion-body myopathy.[77,78] As glycosylation serves important roles in the stability of these example enzymes, it is interesting

to consider the role of the +4 Glu-GlcNAc interactions we identified on these proteins to aid enzyme expression, stability, and ultimately biological function.

From structural bioinformatics analysis, we identified two specific protein-sugar interactions characteristic of α-helical glycoproteins: (1) –4 Glu-GlcNAc interactions utilize interaction mode #1 (χ_1 = 280°) to mediate α-helical structure N-terminal to the glycosylation site. (2) +4 Glu-GlcNAc interactions utilize interaction mode #2 (χ_1 = 180°) to mediate α-helical structure C-terminal to the glycosylation site. We thus suggest that rational design of α-helical glycopeptides should incorporate –4 Glu or +4 Glu to help the peptide maintain helicity upon glycosylation. From structural bioinformatics it is unclear the potential structural effects of incorporating both –4 Glu or +4 Glu into a glycopeptide sequence. However, because both interaction modes involve the same sugar moiety but different orientations and spatial locations of the sugar along the helix, both –4 Glu-GlcNAc and +4 Glu-GlcNAc cannot occur simultaneously. Instead, one interaction mode could dominate and only stabilize a portion of the α-helix, or the sugar could switch between both interaction modes leading to either further maintenance or destabilization of the full α-helix.

1.6 – Comparison to structural bioinformatics analysis of other types of glycosylation and secondary structures

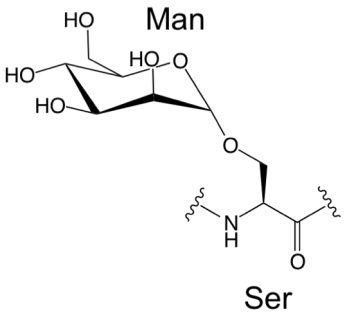

Figure 1.14 Chemical structure of O-linked glycosylation with a core mannose (Man). Ser is shown as an example.

α-helices O-glycosylated with Mannose. Because both favorable modes of Glu-GlcNAc interactions on helices involve the N-acetyl group of the sugar, these interactions are expected to be specific to N-linked GlcNAc. O-linked glycosylation, the second most common type of natural glycosylation, involves a linkage between a variety of core sugars and Ser or Thr (**Figure 1.14**).[4] To probe the specificity of –4 Glu-GlcNAc and +4

Glu-GlcNAc interaction for N-linked GlcNAc, O-linked glycosylation sites were investigated with structural bioinformatics. The most common core O-linked sugar found in our data set is mannose (Man). We identified only six 9-residue helices with O-linked Man. As seen for N-glycosylated 9-residue helices, no glycosylation sites with χ_1 = 60° are observed (**Figure 1.15**). In contrast to N-linked N-terminal helices, favorable protein-sugar interactions on O-linked N-terminal helices were most frequently found at the –3 position with the aromatic residues tryptophan and phenylalanine (**Figure 1.16**). Indeed, the protein-Man interactions on α-helices differ from the –4 Glu-GlcNAc and +4 Glu-GlcNAc interactions found on α-helices. We were unable to identify any individual interaction favorable for C-terminal helices glycosylated with O-linked Man. Thus, favorable Glu-GlcNAc interactions seem specific to the N-acetyl group of GlcNAc. However, the very limited number of O-linked N-terminal or C-terminal helices (16 glycosylation sites) limits the statistical significance of these observations.

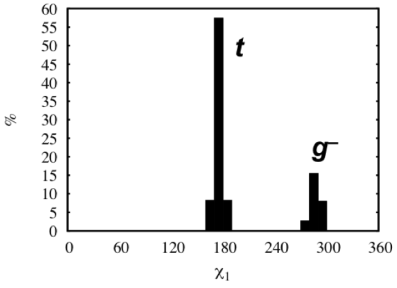

Figure 1.15 χ_1 distribution for 9-residue helices O-glycosylated with Man.

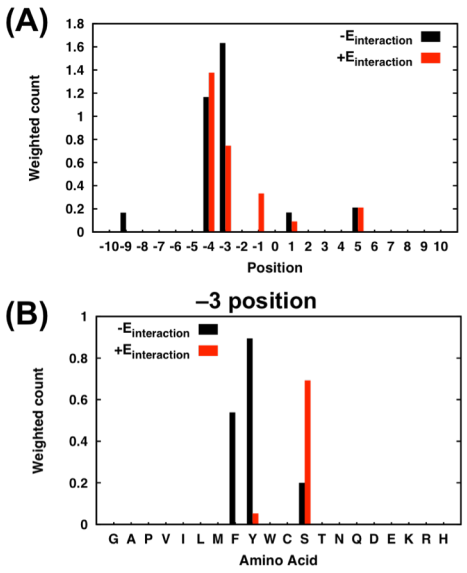

Figure 1.16 Number of N-terminal helices with negative or positive interaction energies **(A)** according to sequence position of the interacting residue and **(B)** at the –3 position.

β-turns N-glycosylated with GlcNAc. Additionally, we investigated N-glycosylation sites found on all types of β-turns to compare the specificity of –4 and +4 Glu-GlcNAc interactions for α-helical structure. For this analysis, turns were by classified based on the ideal

Table 2. Ideal dihedrals of β-turn types.

β-turn type	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	–60°	–30°	–90°	0°
II	–60°	120°	80°	0°
I'	60°	30°	90°	0°
II'	60°	–120°	–80°	0°

Table 3. Weighted counts of N-linked glycosylation sites on each position of a type of β-turn with sequential and/or spatial interactions.

β-turn type	Position of glycosylated Asn	Without amino acids within 3.5 Å	With sequential amino acids within 3.5 Å	With spatial amino acids within 3.5 Å	With both sequential and spatial amino acids within 3.5 Å
I	***i***	28	45	16	26
	***i*+1**	14	4	9	2
	***i*+2**	47	33	47	21
	***i*+3**	41	29	32	15
II	***i***	2	2	0	0
	***i*+1**	6	2	6	1
	***i*+2**	17	3	9	3
	***i*+3**	3	3	1	0.3
I'	***i***	0.3	0	0	0
	***i*+1**	11	1	11	1
	***i*+2**	1	2	0.3	1
	***i*+3**	0	1	0.1	1
II'	***i***	1	3	2	0.3
	***i*+1**	1	0	3	0
	***i*+2**	8	1	5	2
	***i*+3**	1	1	0	2

backbone dihedrals of each turn type, allowing for ±30° variation from each ideal dihedral (**Table 2**). Glycosylation sites at positions from *i* to *i*+3 on the β-turns were identified. The number of glycosylation sites found at each position of each turn type are given in **Table 3** based on the types of interactions (sequential and/or spatial) observed at each site. N-glycosylated type I β-turns are most common of all turn types observed. Since positions *i*, *i*+2, and *i*+3 of type I β-turns are most frequently glycosylated, we investigate the protein-sugar interaction that are most common at those sites (**Figure 1.17**). For glycosylation sites at position *i*, favorable interactions at a given position typically have more or a comparable number of unfavorable interactions also (**Figure 1.17A**). Because interactions at the +3 position have the greatest number of interactions with

negative energies, we look further into the particular amino acids involved in these interactions (**Figure 1.17B**). Favorable +3 position interactions are most commonly found to occur with histidine (His). In contrast to Glu-GlcNAc interactions on α-helices that consistently involve the N-acetyl group of GlcNAc,

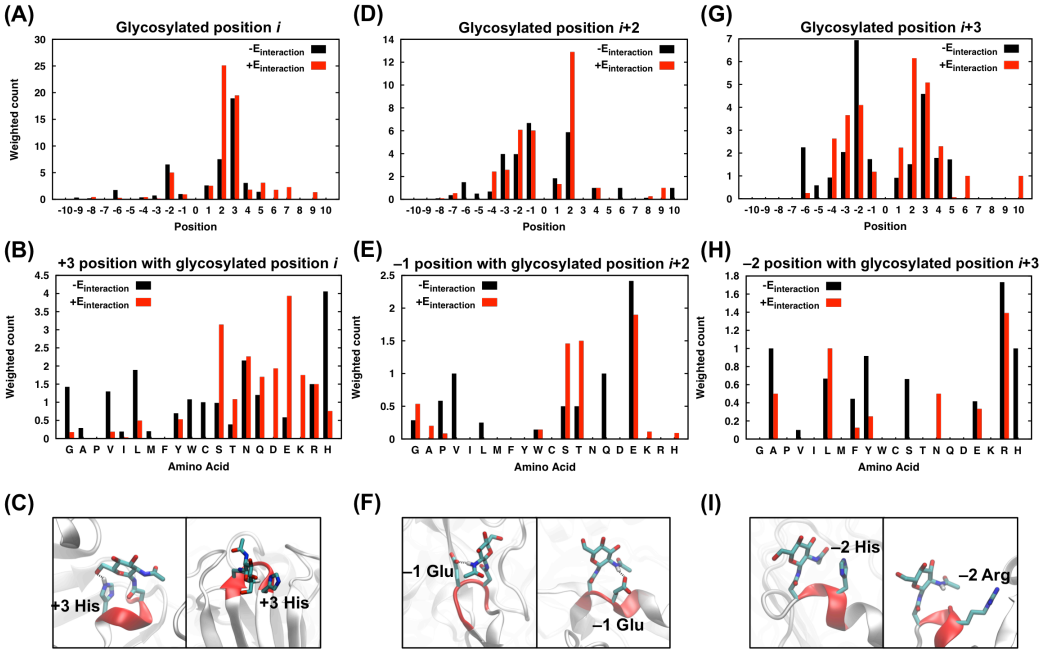

Figure 1.17 For glycosylation sites at position *i* of type I β-turns umber, **(A)** the number of negative or positive protein-sugar interaction energies based on position and **(B)** at the +3 position are given. **(C)** Examples of +3 His-GlcNAc interactions on PDB ID 2CH8 (left) and PDB ID 4KKZ (right). For glycosylation sites at position *i*+2 of type I β-turns umber, **(D)** the number of negative or positive protein-sugar interaction energies based on position and **(E)** at the –1 position are given. **(F)** Examples of –1 Glu-GlcNAc interactions on PDB ID 4E1G (left) and PDB ID 2HR7 (right). For glycosylation sites at position *i*+3 of type I β-turns umber, **(G)** the number of negative or positive protein-sugar interaction energies based on position and **(H)** at the –2 position are given. **(I)** Examples of –2 His-GlcNAc interactions on PDB ID 3J0A (left) and –2 Arg-GlcNAc interactions on PDB ID 3KS9 (right).

+3 His-GlcNAc interactions can involve different moieties of GlcNAc. For example, in **Figure 1.17C**, both H-bonding between His and O6 of GlcNAc and CH-π interactions between His and H5 of GlcNAc are shown. Similar to the results obtained for glycosylation sites at position *i* when the interaction energies are sorted based on the position of the interacting residue, glycans at position *i*+2 are not found to favorably interact with residues at a particular position more than they unfavorably interact (**Figure 1.17D**). Because interactions at the –1 position have the greatest number of interactions with negative energies, we further investigate these interactions, and find –1 Glu-GlcNAc interactions to most

commonly have negative interaction energies (**Figure 1.17E**). –1 Glu-GlcNAc interactions involve H-bonding between the Glu side chain and the N-acetyl group of GlcNAc (**Figure 1.17F**). Even though these –1 Glu-GlcNAc interactions are electrostatic like the –4 and +4 Glu-GlcNAc interactions found on α-helices, the Glu position needed for these favorable interaction to occur differs. For glycosylation sites at position *i*+3 of type I β-turns, interaction with –2 position residues most frequently have negative interaction energies (**Figure 1.17G**). Both –2 His and –2 arginine (Arg) are most commonly involved in these interactions (**Figure 1.17H**). Both of these interactions may actually involve interactions between the N-acetyl group of GlcNAc and the carbonyl oxygen of the peptide backbone as opposed to the side chain characteristic of either residue (**Figure 1.17I**). Overall, the most commonly favorable interaction on type I β-turns differ from those observed on α-helices.

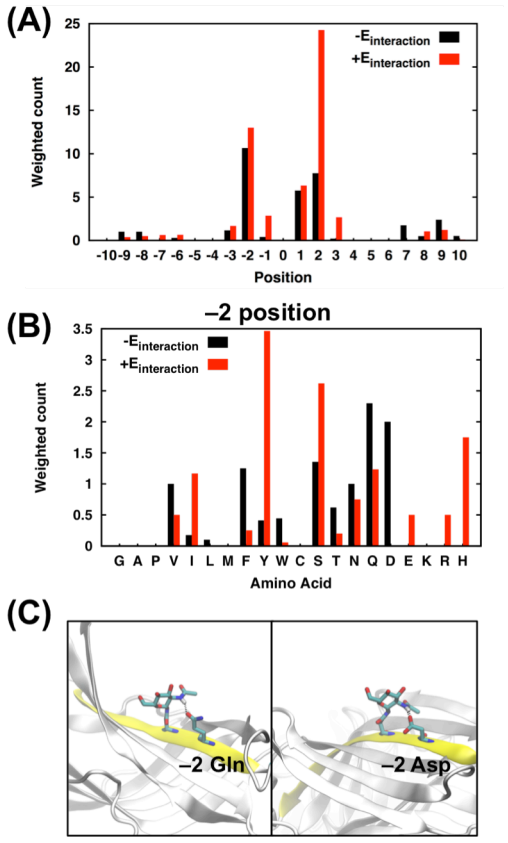

Figure 1.18 Number of negative and positive interaction energies for N-glycosylated β-sheets **(A)** sorted by position and **(B)** at the –2 positions. **(C)** Examples of –2 Gln-GlcNAc interactions on PDB ID 4D01 (left) and –2 Asp-GlcNAc interactions on PDB ID 1DHK.

β-sheets N-glycosylated with GlcNAc. Finally, we investigated N-glycosylation sites on β-sheet structures. Glycosylation sites with at least five residues centered on the glycosylated Asn in extended structures according to both DSSP and STRIDE were considered as glycosylated β-sheets. Of the 167 N-glycosylated β-sheets identified, 69.9 sites have at least one sequential interaction. Analysis of the interaction energies for these sequential interactions found at these sites indicates that although interactions with residues at most

positions are more commonly unfavorable than favorable, the most protein-sugar interactions with negative interaction energies occur at position –2 (**Figure 1.18A**). Both –2 glutamine (Gln) and –2 aspartate (Asp) can favorable interact with GlcNAc (**Figure 1.18B**). Both –2 Gln-GlcNAc and –2 Asp-GlcNAc interactions involve H-bonds with the N-acetyl group of GlcNAc (**Figure 1.18C**). Although the nature of these interaction are similar to the –4 Glu-GlcNAc and +4 Glu-GlcNAc interactions observed on α-helices, the protein-sugar interactions on these β-sheets involve residues not found to commonly interact with GlcNAc on α-helices. Additionally, since glycosylation sites found on β-sheets may be found on tertiary structures consisting of many anti-parallel β-sheets (**Figure 1.18C**), other residue-residue interactions likely help stabilize the glycosylated secondary structure.

1.7 – Limitations to structural bioinformatics analysis

The generality of the conclusions from structural bioinformatics analysis depends on the size and presence of any biases in the PDB glycoprotein data set. The statistical significance is limited by the small number of glycoproteins in the PDB: Although glycans are found on more than 50% of all proteins,[2,3] the number of glycoproteins in the PDB at the time of accession (3,402 PDB structures) pales in comparison to the 105,000+ deposited PDB structures and does not match the natural frequency of glycosylation. Difficulties in fully characterizing the complex structure of glycans, which naturally consist of many monosaccharides connected through a number of different linkage types, may explain this limited number of glycosylated glycoproteins in the PDB.[5] Furthermore, the flexibility of glycans makes structural characterization with X-ray crystallography, the method used to solve the majority of structures deposited in the PDB, quite challenging.[79]

Indeed, another potential bias to consider is the effect of crystallization on the glycoprotein structures. All but three glycosylation sites in the data set used for structural bioinformatics analysis were solved by X-ray crystallography. Crystallization propensity typically depends on low levels of side chain entropy, which results in reduced entropic cost upon crystallization.[80-82] For this reason, the

structural heterogeneity of glycans can hinder the crystallization of glycoproteins.[79] Additionally, crystallization propensity can depend on the formation of favorable electrostatic interactions, which lead to an enthalpic gain upon crystallization.[80,81] Because intermolecular interactions can be altered upon crystallization[83] and electrostatic interactions are favored for crystallization to occur, the two identified protein-sugar interaction modes involving electrostatic interactions with Glu could be biased by crystallization effects. To check if the observed frequency of Glu-GlcNAc interactions may be biased due to crystallization, we compared the frequency of –4 and +4 position Glu around glycosylation sites in the PDB to the frequency in the Swiss-Prot database, which is not biased by any crystallization effects. A –4 Glu is found at 5.94% of glycosylation sites in the PDB and at 5.96% of glycosylation sites in Swiss-Prot; a +4 Glu is found at 5.05% of glycosylation sites in the PDB and at 5.93% of glycosylation sites in Swiss-Prot. These frequencies reasonably agree with the expected occurrence of Glu based on its natural abundance (6.3%).[84] The similar frequencies found in the PDB and Swiss-Prot indicate that crystallization effects likely do not bias conclusions about favorable Glu-GlcNAc interactions from structural bioinformatics analysis.

1.8 – Preliminary experimental investigation of α-helical glycoproteins

To further investigate the utility of the two identified protein-sugar interaction modes for α-helical glycopeptide design, we incorporate –4 and/or +4 Glu into a model peptide sequence and characterize the structure of the peptides with circular dichroism (CD). A sequence composed of primarily alanine, which has been previously used to investigate the effects of the center amino acid on the peptide's α-helicity,[85] was chosen as model peptide sequence. The peptide sequences investigated are given in **Table 4**. Glu was incorporated at either the –4 or +4 position relative to the central Asn to test the ability of –4 Glu-GlcNAc and +4 Glu-GlcNAc interactions, respectively, to maintain α-helical structure upon glycosylation. Since it is unclear from structural bioinformatics analysis how α-helical structure will be affected by glycosylation when interactions with both –4

and +4 Glu are possible, a peptide with both –4 and +4 Glu was also investigated. CD experiments in 20 mM sodium phosphate buffer at pH 7 were performed in duplicate for each peptide.

Table 4. Peptide sequences investigated. Underlined residues are glycosylated with GlcNAc. % helicity of each peptide determined from CD experiments is also given.[86]

	Sequence	% Helicity
N22	Ac–AOOAAAAAANAAAAAAOOAGGY–NH_2	19.50
N22g	Ac–AOOAAAAAA<u>N</u>AAAAAAOOAGGY–NH_2	7.08
EN22	Ac–AOOAAEAAANAAAAAAOOAGGY–NH_2	33.81
EN22g	Ac–AOOAAEAAA<u>N</u>AAAAAAOOAGGY–NH_2	27.93
NE22	Ac–AOOAAAAAANAAAEAAOOAGGY–NH_2	23.98
NE22g	Ac–AOOAAAAAA<u>N</u>AAAEAAOOAGGY–NH_2	22.45
ENE22	Ac–AOOAAEAAANAAAEAAOOAGGY–NH_2	42.88
ENE22g	Ac–AOOAAEAAA<u>N</u>AAAEAAOOAGGY–NH_2	41.35

The CD spectra are shown in **Figure 1.19**, and the percent helicity[86] of each peptide is given in **Table 4**. Minima at 208 and 222 nm are characteristic of α-helical structure. The structure of N22, the peptide sequence lacking Glu, is significantly altered upon glycosylation. In contrast, the structures of EN22, NE22, and ENE22, the peptide sequences with at least one Glu, are not altered by glycosylation. Incorporation of either –4 Glu or +4 Glu allows the peptide to maintain its structure upon glycosylation, supporting the hypothesized role of –4 Glu-GlcNAc and +4 Glu-GlcNAc interactions to help maintain the α-helicity of glycopeptides. ENE22g, which has the potential for both –4 Glu-GlcNAc and +4 Glu-GlcNAc interactions, also maintained its structure. Although both interaction modes cannot occur simultaneously as determined from structural bioinformatics analysis, the sugar may flip between interaction modes with both –4 Glu and +4

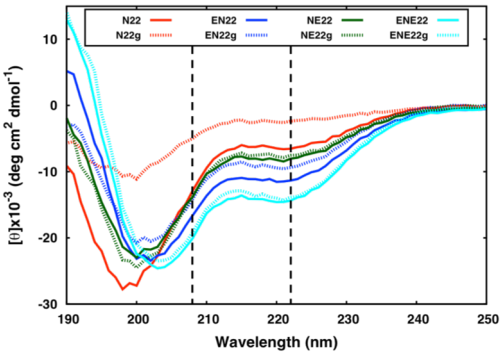

Figure 1.19 CD spectra of investigated peptides. The spectra shown in dashed lines are from glycosylated peptides. 208 and 22 nm are indicated with dashed lines.

Glu, thus helping the peptide to maintain its structure through both interactions. Our preliminary experimental findings support our predictions from structural bioinformatics.

1.9 – Conclusions

Glycans have been shown to improve protein stability and folding efficiency,[8,13,14,19,22,87] characteristics desirable to include in protein engineering and therapeutic design. However, an incomplete understanding of biophysical effects of glycosylation on protein structure has limited the rational design of novel glycoproteins. We thus sought to investigate principles that further our understanding of how protein-sugar interactions impact α-helical protein structure and that are useful for helical glycopeptide design. From purely steric arguments, we illustrated that a limited number of glycan conformations are amenable to α-helical structure and predicted two sugar conformations possible for α-helical glycoproteins. Indeed, these two conformations (mode #1 with χ_1 = 280° and mode #2 with χ_1 = 180°), were the only conformations observed on naturally glycosylated α-helices in the PDB. Further structural bioinformatics analysis resulted in the identification of protein-sugar interactions characteristic of α-helical glycoproteins that utilized these two modes: (1) Mode #1 is utilized in –4 Glu-GlcNAc interactions to mediate α-helical structure N-terminal to the glycosylation site. (2) Mode #2 is utilized in +4 Glu-GlcNAc interactions to mediate α-helical structure C-terminal to the glycosylation site. To illustrate the potential utility of these two interaction modes for α-helical glycopeptide design, we experimentally characterized the effects of glycosylation on peptide structures including –4 Glu and/or +4 Glu. Our preliminary experimental results supported the utility of incorporating –4 Glu or +4 Glu to maintain α-helical structure upon glycosylation.

1.10 – Future directions

Further experimental investigation with CD and NMR experiments will more completely probe the effects of both –4 Glu-GlcNAc and +4 Glu-GlcNAc

interactions on α-helical peptide structure. Since the non-glycosylated peptide with only Asn incorporated (N22) exhibited limited α-helical structure, further experiments will be performed on a longer polyalanine sequence of Ac–OOAAAAAAAAAAAAAAAAAAOOY–NH_2, expected to exhibit further α-helicity.[88-90] The sequence without any Asn incorporated will also be investigated to ensure the wild type peptide is α-helical. Additionally, variants with –4 and +4 aspartate (Asp) will also be investigated. Since Asp and Glu have the same side chain terminal functional group, comparisons between the peptides with Glu and with Asp can help shed light on the importance of side chain length and specificity of Glu-GlcNAc interactions to mediate α-helical structure. From structural bioinformatics analysis, –4 Asp-GlcNAc and +4 Asp-GlcNAc interactions are not frequently observed and, thus, not predicted to maintain helical structure as seen for –4 Glu-GlcNAc and +4 Glu-GlcNAc interactions. Furthermore, variants N-glycosylated with glucose (Glc) instead of GlcNAc will allow us to probe the chemical nature of the Glu-GLcNAc interactions. Since the Glu-GlcNAc interactions identified from structural bioinformatics analysis depend on the N-acetyl group, proteins glycosylated with Glc are not expected to maintain α-helicity.

Concurrently, molecular dynamics (MD) simulations of the experimentally characterized peptides and their glycosylated variants will be performed. The simulations will allow further dynamical information about the protein-sugar interactions to be obtained. Replica exchange MD (REMD) simulations will be used to enhance the sampling of the system. In REMD simulations, multiple replicas are run in parallel at different temperatures. During the simulations, exchanges between replicas are attempted so that higher temperature replicas that sample many different configurations can enhance the sampling of the lower, biologically-relevant temperature replicas. From these atomistic simulations, we can gain further molecular details about the interactions modes more difficult or impossible to obtain from CD experiments. After ensuring the equilibrium structural ensemble for each peptide obtained from simulation agrees with experimental results, the prevalence of each Glu-GlcNAc interaction and each

conformer of Asn χ_1, which characterizes the two interaction modes, can be determined. As the structural ensembles of each peptide may likely be composed of multiple structural states, the key residue contacts unique to each state can be identified through Bayes factor analysis;[91] by determining the importance of a given interaction for a particular structural state, the contacts key to glycosylated α-helices can be identified. From structural bioinformatics analysis, we expect –4 and +4 Glu-GlcNAc contacts to be key to glycosylated α-helices. Furthermore, by comparing contacts observed in the glycosylated α-helices compared to non-glycosylated α-helices, it can be determined if any native contacts are disrupted by glycosylation. Overall, a two pronged approach with both wet experiments and simulations can provide further insights into the molecular importance of Glu-GlcNAc interactions to mediate α-helical structure.

Another possible way to investigate the importance of the protein-sugar interactions identified from structural bioinformatics is through evolutionary analysis. In section 1.5 – Protein-sugar interactions of α-helical glycoproteins, the possibility that a –4 Glu may have evolved with a biologically important glycosylation site on phospholipase B-like protein is suggested. A more extensive investigation of the evolution of this glycoprotein sequence and others would allow the significance of the identified protein-sugar interactions to be probed from a perspective orthogonal to the proposed wet experiments and simulations.

Additionally, our structural bioinformatics analysis can be applied to identify favorable protein-sugar interactions characteristic of other secondary structures. For instance, we identified N-glycosylation sites found on all types of β-turns and β-sheets. As glycosylation is also thought to occur more often at changes of secondary structure,[33-36] investigations of glycosylation sites between different secondary structures, for example between an α-helix and a turn, with structural bioinformatics analysis could also probe the impact of protein-sugar interactions on super-secondary structures. Such investigations could be initial steps to better understand the effects of glycosylation on larger protein structures. Indeed, further knowledge of how protein-sugar interactions that impact the full structure of a

protein, including protein-sugar interactions arising from tertiary contacts, will be needed to rationally incorporate glycans onto larger biologics.

Chapter 2 – Understanding the conformational dynamics of cyclic peptides

2.1 – An introduction to cyclic peptides and approaches to simulate their structures

Protein-protein interactions (PPIs) are responsible for numerous biological processes in cellular development and the maintenance of homeostasis.[92] Abnormal PPIs can result in numerous diseases such as cancers, bacterial and viral infections, and neurodegenerative diseases.[93] Thus, targeting PPIs offers a useful avenue to design new therapeutics.[94] However, the shallow and large surfaces (~1,500–3,000 $Å^2$)[95,96] characteristic of protein-protein interfaces poses challenges for the development of small-molecule PPI modulators: Small-molecule drugs typically bind in smaller (~300–1,000 $Å^2$)[97,98] groves and pockets of a protein target. One of the most promising candidates for PPI inhibition is cyclic peptides. With sizes intermediate of small-molecule drugs and larger biologics, cyclic peptides can bind protein surfaces with high affinity and specificity. Additionally, cyclization can lead to reduced proteolytic degradation, preorganization of the peptide into a bioactive conformation, and improved membrane permeability. Cyclic peptides can exhibit improved stability, selectivity, and bioavailability compared to their linear counterparts.[99-106] Even though cyclic peptides have been successfully applied to modulate PPIs, and some are being further investigated for therapeutic applications,[104,107-111] the rational design of novel cyclic peptide therapeutics remains limited. One of the challenges hindering the development of new cyclic peptide PPI modulators is the difficulty of efficiently predicting their structures with current computational methods.

Both the structures and dynamics of cyclic peptides can vary greatly from those of linear peptides and larger proteins, for which many simulation methods have been developed. In comparison to highly flexible linear hexapeptides, cyclic

hexapeptides typically adopt either one or a very limited number of structures due to ring strain.[112-119] Furthermore, since the ring strain of small cyclic peptides creates large free energy barriers between different conformations, cyclic peptide structures can be very difficult to efficiently sample with standard molecular dynamics (MD) simulations.

One enhanced sampling method that can efficiently provide converged structural ensembles is metadynamics (META) simulations. In a META simulation, the free energy of the system is biased along a given reaction coordinate or collective variable (CV). By depositing Gaussian hills to fill the free energy profile of the system, large free energy barriers can be overcome and the full free energy profile can be sampled.[120-122] The choice of CV greatly affects the performance of a META simulation; if the chosen CV properly captures the slow degrees of freedom such that the projection of the free energy onto the selected CV contains the barriers hardest for the system to overcome, then converged results can be quickly obtained through META simulations. Thus, the most challenging aspect of META simulations is proper CV selection, which typically requires knowledge of the chemistry fundamental to the system of interest.

Full structural ensembles of cyclic peptides have been obtained with META and bias-exchange META (BE-META) simulations.[123-126] In BE-META simulations, the system is biased along multiple CVs simultaneously by running multiple replicas in parallel. A different CV is used for each replica, and exchanges between replicas are attempted during the simulation to further enhance sampling.[122] Our lab has previously utilized BE-META simulations to provide converged structural descriptions for the model cyclic octapeptide cyclo-(YNPFEEFF).[126] By biasing ϕ, ψ, and χ dihedrals, converged results were obtained in 300 ns. The performance of this method can be greatly improved by selecting CVs based upon a better understanding of the mechanisms that cyclic peptides utilize to switch conformations. To gain such mechanistic understanding, conformational switches observed in MD simulations of cyclo-(GGGGGG) are analyzed and found to occur via one of two mechanisms. These two mechanisms are observed during conformational changes of the more complex cyclic peptides

cyclo-(AAAAAA) and cyclo-(YNPFEEGG). These mechanistic insights allowed us to improve our CV selection for enhanced sampling of cyclic peptide structures using META simulations, reducing the computational time needed to obtain a converged structural ensemble for cyclic peptides.

2.2 –Methods to analyze cyclic peptide conformational switches

To gain such fundamental understanding of how cyclic peptides switch conformation, I analyzed MD simulations performed by a graduate coworker (Sean M. McHugh) of a highly flexible cyclic hexapeptide composed of all glycine residues using the RSFF1 force field[127] with TIP4P-EW[128] water at 300 K. Experimentally, cyclic hexapeptides with all *trans* peptide bonds adopt stable structure with two β-turns (ideal dihedrals for β-turn types are given in **Figure 2.1A**).[113,114,119,129-139]

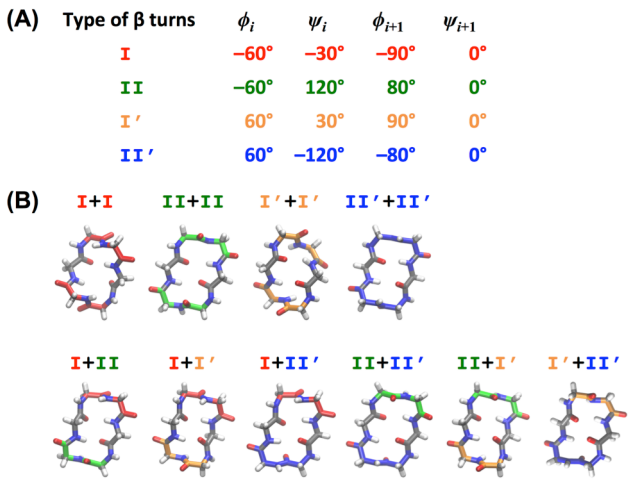

Figure 2.1 (A) Ideal dihedrals for each β-turn type. **(B)** All ten theoretically possible combinations of two β-turn structures shown for cyclo-(GGGGGG).

Structures with combinations of (1) a type I β-turn + a type I β-turn;[119] (2) I + II; [119,137] (3) I + I';[129,140] (4) I + II';[115,131] (5) II + II;[135,137] (6) II + II';[115] (7) II' + II'[114] have been characterized experimentally. With four β-turn types, a total of ten different combinations of structures with two β-turns are theoretically possible (**Figure 2.1B**). The three combinations of β-turns that have not been observed (II+I', I'+I' and I'+II') may be due to the small number of cyclic hexapeptides that have been characterized, or may be due to very small populations of these turn combinations among other structures being difficult to detect. With a limited number of stable conformations, cyclic hexapeptides are thus an exemplary system to investigate mechanisms of conformational change. Because glycine lacks a side chain and is achiral, we

hypothesize that numerous conformational switches can be observed between turn types via MD simulations of cyclo-(GGGGGG).

To investigate the details of conformational changes of cyclo-(GGGGGG) and cyclo-(AAAAAA) and since cyclic hexapeptide structures have been experimentally characterized by two β-turns (**Figure 2.1**), I assigned the β-turn type(s) found in each frame of the trajectory. A β-turn type was assigned to two consecutive residues if the ϕ and ψ backbone dihedrals of both residues were within 30° of the ideal values for a given β-turn type. Every consecutive pair of residues was tested for a β-turn. To gain a more complete understanding of the mechanisms underlying conformational changes, the magnitude of dihedral angle changes during transitions between β-turn types is calculated.

In contrast to the cyclic hexapeptides cyclo-(GGGGGG) and cyclo-(AAAAAA), the conformations of the octapeptide cyclo-(YNPFEEGG) are not solely characterized by β-turns. Thus, conformational states of cyclo-(YNPFEEGG) were assigned based on the clusters obtained from dihedral principal component analysis (dPCA)[57,58] and a subsequent peak-based density clustering[59] in 3D principal component space.

2.3 – Analysis of cyclo-(GGGGGG)

Two sets of MD simulations of cyclo-(GGGGGG) were preformed using the RSFF1 force field[127] and TIP4P-EW[128] water by a graduate student in the lab (Sean M. McHugh). Many conformational switches of cyclo-(GGGGGG) can be observed during 1 μs MD simulations (**Figure 2.2**), and a converged structural ensemble, in which structures with all 10 possible combinations of two β-

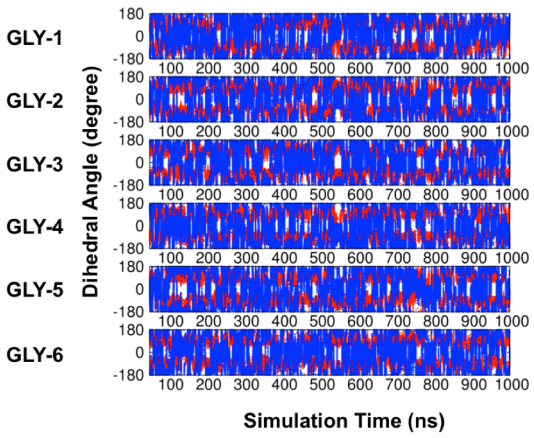

Figure 2.2 Trajectory of ϕ (red) and ψ (blue) of an MD simulation of cyclo-(GGGGGG). (Credit: Sean M. McHugh)

turns are present, can be obtained from MD simulations in a reasonable amount of time,[141] suggesting that conformational changes are also likely sufficiently sampled.

Using these 1 μs MD simulations, I analyzed how cyclo-(GGGGGG) transitions from one combination of β-turns to a different combination of β-turns. Examples of observed conformational changes are shown in **Figure 2.3A**. Conformational changes proceed according to two different pathways: (1) The positions of the β-turns remains the same but two of the residues involved in a β-turn transition to a different type of β-turn (blue boxes in **Figure 2.3**). (2) The positions of the β-turns change (red boxes in **Figure 2.3**). During conformational changes via this pathway, the *i*+2 residues of the β-turns in the initial structure become the *i*+1 residues of the β-turns in the final structure (or equivalently *i*+1

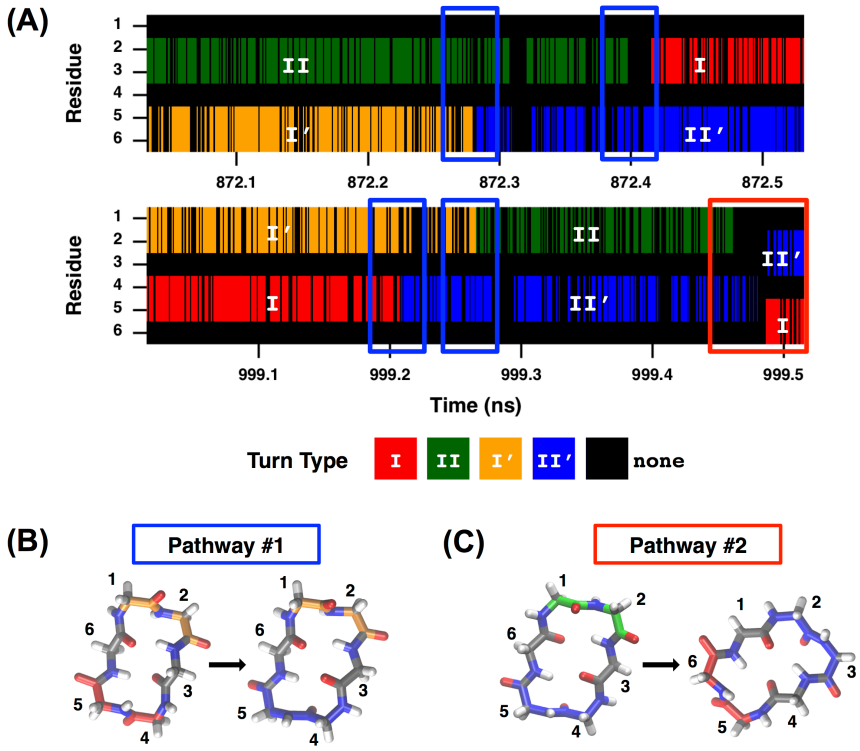

Figure 2.3 (A) Examples of β-turn analysis from MD simulations of cyclo-(GGGGGG). Two pathways of conformational changes are identified: (1) in-position change of β-turn (blue boxes) and (2) shifted-position change of β-turn type (red box). β-turn types are colored as shown in **Figure 2.1**. Illustrations of conformational changes via **(B)** pathway #1 and **(C)** pathway #2.

residues become *i*+2 residues). For cyclo-(GGGGGG), conformational switches occur via pathway #1 approximately four times more than switches via pathway #2.

Mechanisms of pathway #1. In-position changes of β-turn type (pathway #1) typically involve changes between I' ⇔ II, I ⇔ II', I ⇔ II, and I' ⇔ II'. These four pairs of β-turn types each have two similar ideal dihedral angles (**Figure 2.1**), reducing the total number of changes that are likely necessary to complete a conformational change. To gain a more complete understanding of the mechanisms underlying conformational changes, I analyzed the degree of dihedral angle changes during conformational transitions. Pathway #1 is found to proceed via two different mechanisms. The first mechanism is characterized by large coupled dihedral changes of the *i*+1 residue of a β-turn. The ϕ and ψ angles of the *i*+1 residue quickly change by a large magnitude, while the other backbone dihedrals of cyclo-(GGGGGG) remain relatively constant. For example, as shown in **Figure 2.4A**, a type I β-turn at residues 1 and 2 switches to a type II' β-turn through coupled changes of ϕ_1 and ψ_1. Mechanism 1 underlies conformational changes of I' ⇔ II and I ⇔ II'. These two pairs of turn types have similar ideal dihedrals for the *i*+2 residue such that larger changes at the *i*+1 position are needed to complete these conformational switches (**Figure 2.4B**). In contrast, the second mechanism is characterized by large coupled dihedral changes of the ψ angle of the *i*+1 residue and the ϕ angle of the *i*+2 residue of a β-turn. For example, as shown in **Figure 2.4C**, a type II' β-turn at residues 4 and 5 switches to a type I' β-turn through coupled changes of ψ_4 and ϕ_5. Mechanism 2 underlies conformational changes of I ⇔ II and I' ⇔ II'. These two pairs of turn types have similar ideal ϕ_{i+1} and ψ_{i+2} angles such that larger changes at the ψ_{i+1} and ϕ_{i+2} angles are needed to complete these conformational transitions (**Figure 2.4D**). For cyclo-(GGGGGG), in-position shifts of β-turn type occur via mechanism 1 (coupled changes of ϕ_i and ψ_i) approximately ten times more often than via mechanism 2 (coupled changes of ψ_i and ϕ_{i+1}). Since the definitions of ϕ and ψ of a single residue both include Cα, changes of ϕ_i are more likely to be coupled to

changes of ψ_i and vice versa, possibly accounting for the higher frequency of mechanism 1.

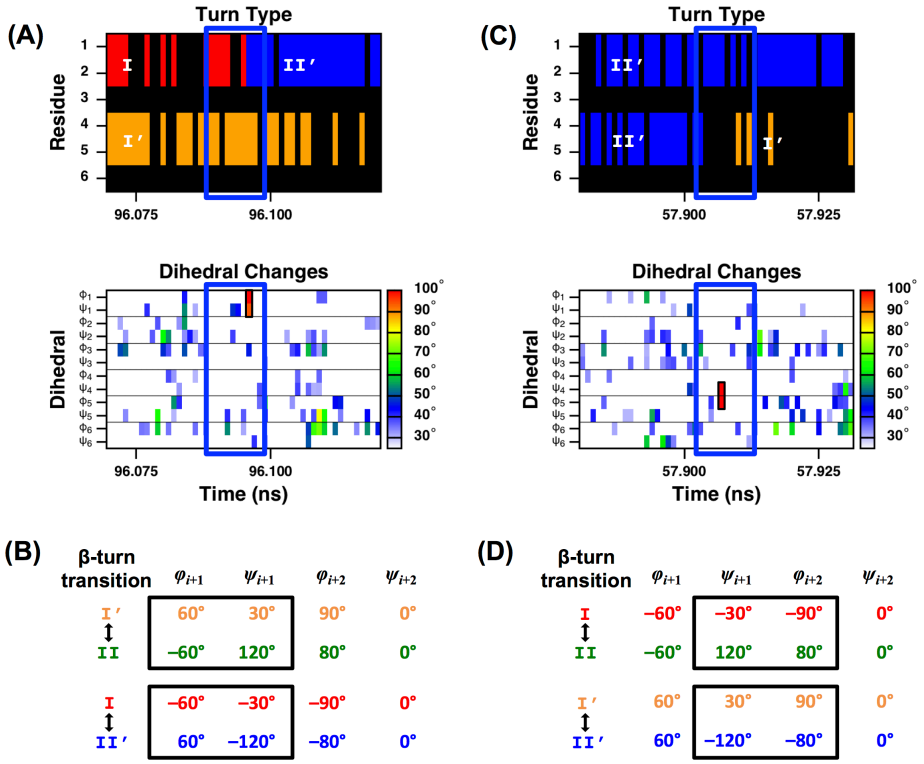

Figure 2.4 Example of a conformational change of cyclo-(GGGGGG) following pathway #1 via **(A)** mechanism 1. The dihedral angle analysis shows the degree of dihedral angle changes from the previous frame. Coupled ϕ_1 and ψ_1 angle changes are observed during the transition from a type I β-turn to a type II' β-turn at residues 1 and 2. **(B)** In the first mechanism of pathway #1, the ϕ and ψ angles of the same residue change during conversion between type I' and II β-turns, or type I and II' β-turns. **(C)** Example of a conformational change via mechanism 2. Coupled ψ_4 and ϕ_5 angle changes are observed during conversion from a type II' β-turn to a type I' β-turn at residues 4 and 5. **(D)** In the second mechanism, the ψ angles of a residue and the ϕ angle of the following residue change to enable conversion between type I and II β-turns, or type I' and II' β-turns.

Mechanisms of pathway #2. Shifted-position changes of β-turn types (pathway #2) typically require more time to complete conformational switches than via pathway #1. The patterns of dihedral changes during transitions via pathway #2 are also more complex (**Figure 2.5**). Since the positions of the β-turns change, all six residues are affected by the conformational switch. Even though changes in many dihedrals are required for a complete conformational change, large coupled changes of the ψ angle of one residue and the ϕ angle of the next residue limit full transitions. The dihedral angle changes underlying pathway #2

primarily proceed via mechanism 2 (**Figure 2.4C**). Thus, the reduced frequency of conformational changes via pathway #2 during simulations of cyclo-(GGGGGG) is consistent with the observed reduced frequency of mechanism 2.

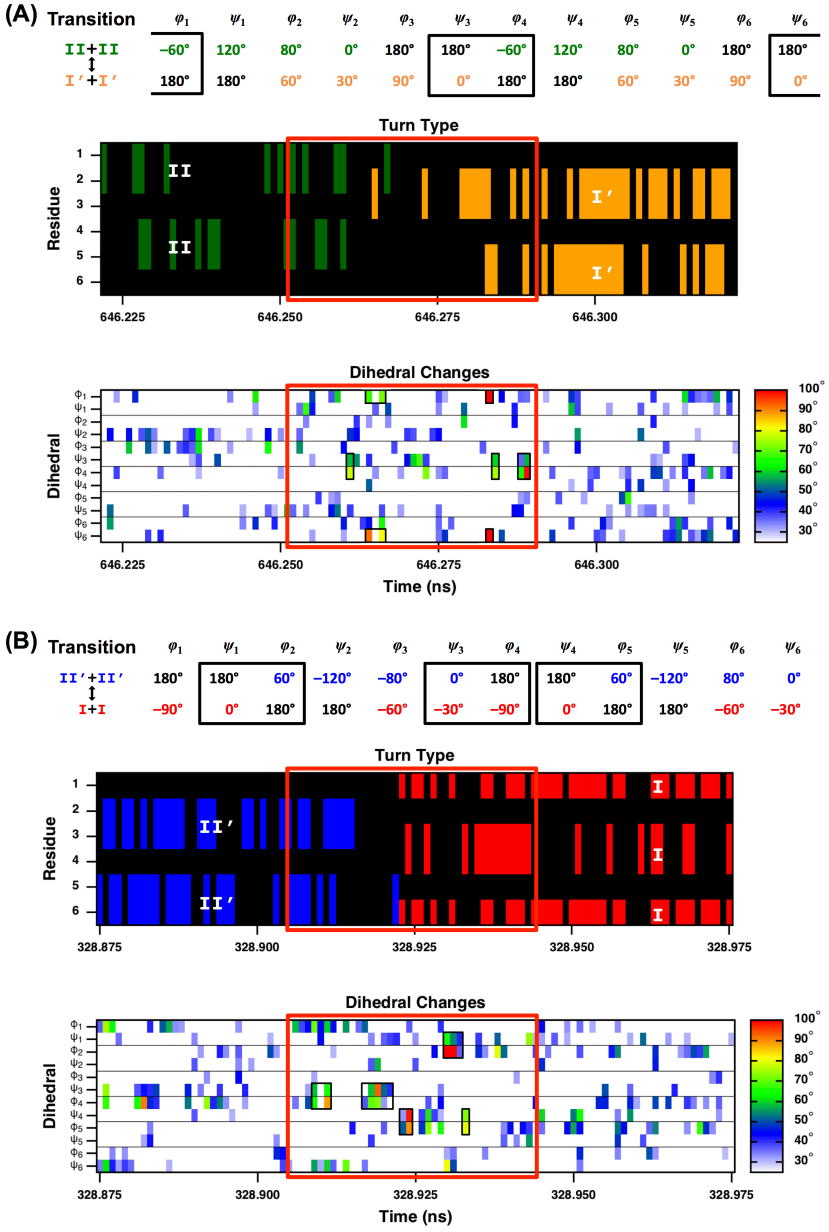

Figure 2.5 Examples of conformational changes of cyclo-(GGGGGG) via pathway #2. **(A)** A conformation with a type II turn at residues 1 and 2 + a type II turn at residues 4 and 5 changes into a conformation with a type I' at residues 2 and 3 + a type I' turn at residues 5 and 6. The dihedral angle analysis shows the degree of dihedral angle changes from the previous frame. Coupled two-dihedral changes, for example, ψ_6 and ϕ_1 at ~646.28 ns, are observed to complete the transition (highlighted by black boxes). **(B)** A conformation with a type II' turn at residues 2 and 3 + a type II' turn at residues 5 and 6 change into a conformation with a type I at residues 3 and 4 + a type I turn at residues 6 and 1. Coupled two-dihedral changes, for example, ψ_4 and ϕ_5 at ~328.925 ns, are observed.

Mechanisms observed with different force fields. During MD simulations of cyclo-(GGGGGG) using the RSFF1 force field, conformational changes occur via coupled two-dihedral changes of either ϕ_i and ψ_i (mechanism 1) or ψ_i and ϕ_{i+1} (mechanism 2). RSFF1, a residue-specific force field with backbone and side chain parametrization based on the protein coil library, was developed to reproduce the Ramachandran plots of each of the 20 individual amino acids.[127] Individual descriptions of each individual amino acid may be especially important to simulate cyclic peptides; with only a few residues, cyclic peptides are typically composed of a few types of amino acids, suggesting that each individual amino

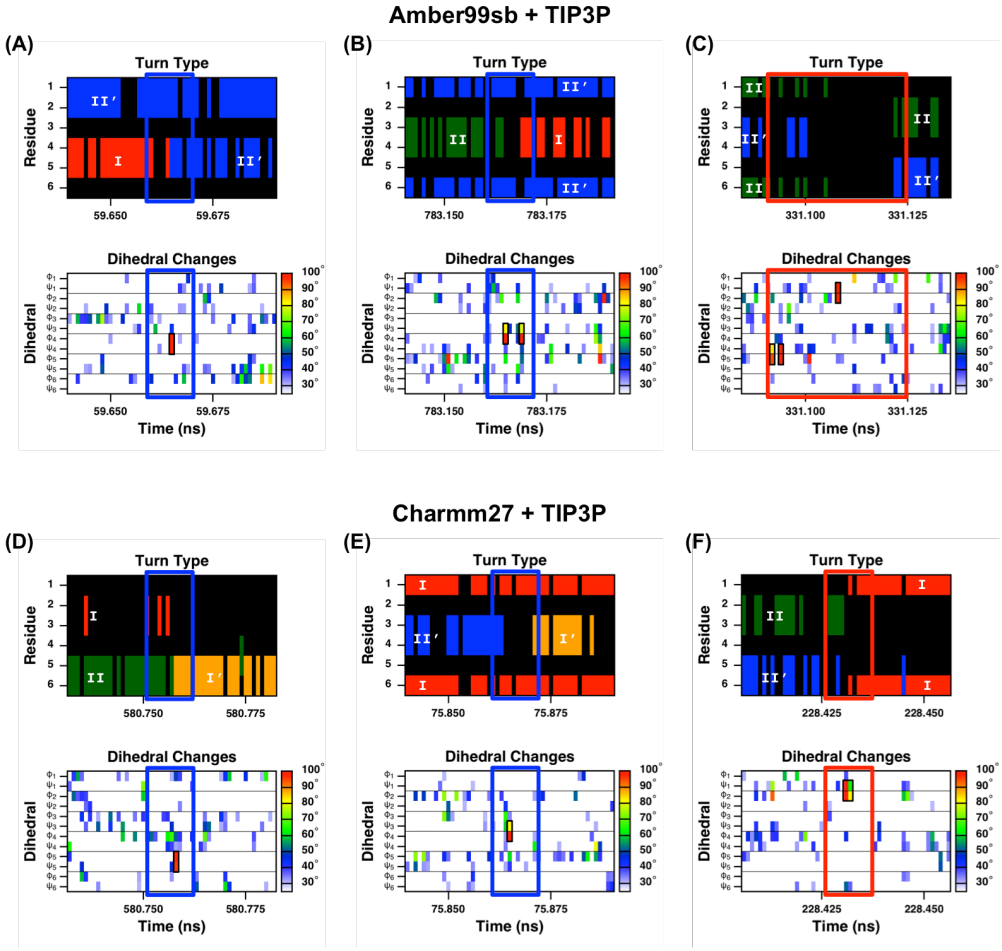

Figure 2.6 Examples of coupled two-dihedral changes from MD simulations of cyclo-(GGGGGG) using **(A-C)** Amber99sb, and **(D-F)** Charmm27. For each example, the results of β-turn analysis are shown above the results of dihedral angle analysis. **(A, D)** Conformational changes occur via pathway #1, mechanism 1. **(B, E)** Conformational changes occur via pathway #1, mechanism 2. **(C, F)** Conformational changes occur via pathway #2, mechanism 2.

acid needs to be accurately parametrized. Additionally, RSFF1 was chosen since in our previous study of cyclo-(YNPFEEGG), the force field well recapitulated the structure of the YNPF region of the cyclic peptide, the region crucial for binding the EH domain of EHD1.[142] However, the RSFF1 force field also overstabilized the β/PPII region of the Ramachandran plot and failed to describe the noncanonical ϕ/ψ backbone dihedrals of Glu-5.[126] To determine if our observations are inherent to the dynamics of the cyclic peptide and not dependent on the force field used for the simulations, MD simulations of cyclo-(GGGGGG) using the Amber99sb force field[143] with TIP3P water[144] and using the Charmm27 force field[145] with TIP3P water[144] were performed by a graduate student in the lab (Sean M. McHugh). My analysis of both simulations showed that conformational changes occur via coupled two-dihedral changes consistent with mechanism 1 and mechanism 2 (**Figure 2.6**). Thus, even though thermodynamic descriptions of the structural ensemble of a small cyclic peptide may be force field dependent, the mechanisms of conformational change appear force field independent.

2.4 – How other cyclic peptides switch conformation

Two mechanisms of coupled two-dihedral changes underlie conformational changes of cyclo-(GGGGGG): Coupled changes of ϕ and ψ of one residue characterize mechanism 1, and coupled changes of ψ of one residue and ϕ of the next residue characterize mechanism 2. I also investigate the mechanisms resulting in conformational changes of the slightly more complex cyclic hexapeptide cyclo-(AAAAAA) (**Figure 2.7**). Both pathway #1 (**Figure 2.7A** and **B**) and pathway #2 (**Figure 2.7C**) occur. As for cyclo-(GGGGGG), pathway #1 proceeds via mechanism 1 for switches of I' ⇔ II and I ⇔ II', and pathway #2 proceeds via mechanism 2 for switches of I ⇔ II and I' ⇔ II'. The duration of shifted-position changes of β-turn type for cyclo-(AAAAAA) are also are longer and more complex than those for pathway #1.

Additionally, the conformational switches of the larger and more heterogeneous cyclo-(YNPFEEGG) are analyzed by classifying conformational states base on dPCA and clustering. Because pathways are defined based on β-

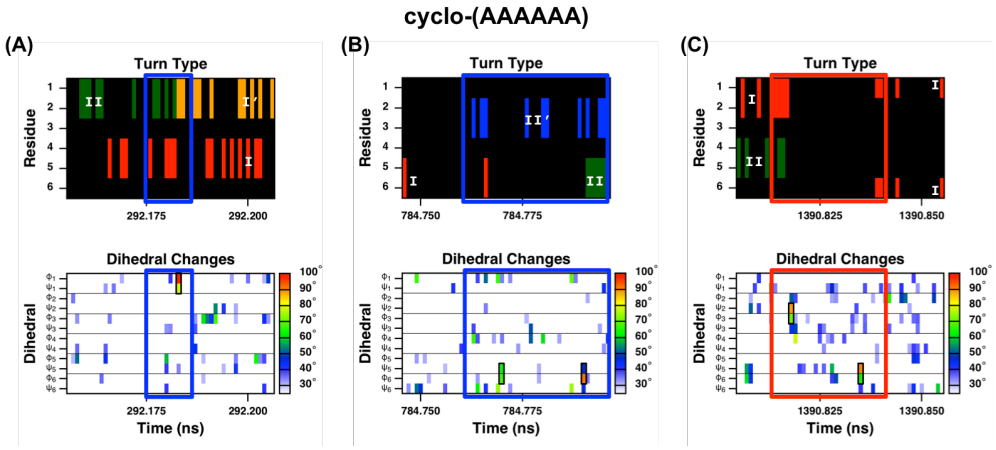

Figure 2.7 Examples of coupled two-dihedral changes from MD simulations of cyclo-(AAAAAA) performed by a graduate student in the lab (Sean M. McHugh) of cyclo-(AAAAAA). **(A)** A conformational change occurs via pathway #1, mechanism 1 involving coupled ϕ_1 and ψ_1 angle changes. **(B)** A conformational change occurs via pathway #1, mechanism 2 involving coupled ψ_5 and ϕ_6 angle changes. **(C)** A conformational change occurs via pathway #2 in which a type I at residues 1 and 2 + a type II at residues 4 and 5 switches to a type I at residues 2 and 3 + a type II' at residues 5 and 6 over ~0.26 ns. As examples, coupled dihedral changes between ψ_2 and ϕ_3 and between ψ_5 and ϕ_6 are shown from the beginning of the transition in which the type I β-turn at residues 1 and 2 first changes to a type I at residues 6 and 1 before completing the transition to a conformation with two β-turns.

turn types, they cannot be observed for cyclo-(YNPFEEGG). Even so, the same two mechanisms of coupled two-dihedral changes are also observed during conformational transitions of cyclo-(YNPFEEGG) (**Figure 2.8**). Changes localized either to ϕ and ψ of one residue (mechanism 1) (**Figure 2.8A**), or to ψ of one residue and ϕ of the next residue characterize (mechanism 2) (**Figure 2.8B**) underlie conformational switches. In the analysis of dihedral changes during conformational switches of cyclo-(YNPFEEGG), changes in χ angles were also considered to determine if changes in side chain orientations are also necessary for conformational changes. No large χ angle changes occur simultaneously with the coupled ϕ and ψ angle changes of mechanism 1 and mechanism 2 during conformational switches. Therefore, the two mechanisms of coupled two-dihedral changes observed for cyclo-(GGGGGG) appear to be inherent to conformational switches of more complex cyclic peptides of six to eight residues.

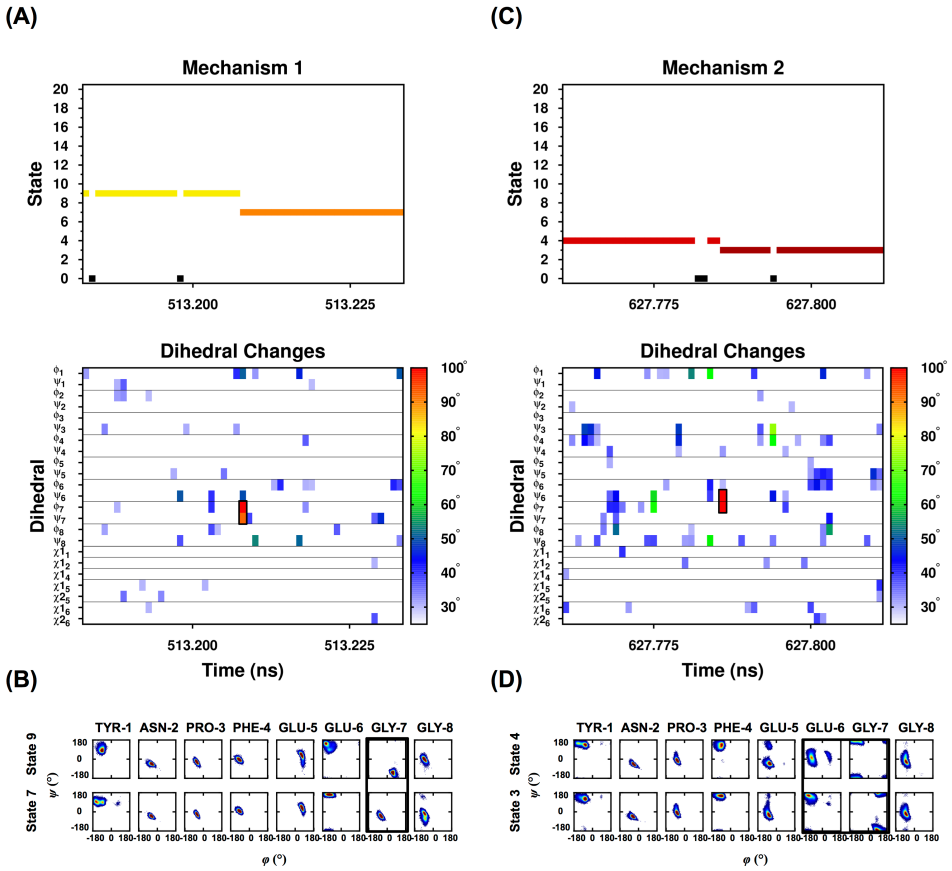

Figure 2.8 Examples of conformational changes of cyclo-(YNPFEEGG) from MD simulations performed by a graduate coworker (Sean M. McHugh). **(A)** Changes of conformational state following mechanism 1. The state index is defined based on the cluster analysis of the dPCA results; frames unassigned to a cluster are given state index 0. Coupled ϕ_7 and ψ_7 angle changes are observed during the transition. (**B**) Ramachandran plots for the backbone dihedral angles of the two states involved in the conformational switch (states 9 and 7). Black box indicates which residue undergoes the largest degree of dihedral angle change. **(C)** Changes of conformational state following mechanism 2. Coupled ψ_6 and ϕ_7 angle changes are observed during the transition. (**D**) Ramachandran plots for the backbone dihedral angles of the two states involved in the conformational switch (states 4 and 3).

2.5 – An improved simulation method for cyclic peptides

Based on the two observed mechanism of conformational change identified from my analysis, a new set of CVs for efficient META simulations of cyclic peptides are developed and implemented by a graduate coworker (Sean M. McHugh). To target the mechanisms of coupled two-dihedral changes, each replica is biased along two CVs concurrently in the new BE-2D-META simulations. Two types of 2D biases are applied: (1) The ϕ and ψ angles of each

residue are used as 2D coordinates ($\phi_i \times \psi_i$) to capture the changes limiting conformational changes via mechanism 1, and (2) the ψ angle of each residue and the ϕ angle of the next residue are used as 2D coordinates ($\psi_i \times \phi_{i+1}$) to capture the changes limiting conformational changes via mechanism 2.

To test the efficiency of the new BE-2D-META method using CVs based on the two mechanisms of conformational change, the converging behavior of the new method compared to MD and BE-META simulations for cyclo-(AAAAAA) and cyclo-(YNFPEEGG) are compared. Although converged results could not be obtained from 1.5 μs MD for cyclo-(AAAAAA), BE-2D-META simulations converge in 100 ns, faster than BE-META simulations biasing the 12 backbone dihedrals, and BE-2D-META simulations biasing 6 ($\phi_i \times \psi_i$) and 6 ($\psi_i \times \phi_{i+1}$), which converge in 400 ns. Furthermore, the new BE-2D-META method is tested on the more complex cyclo-(YNPFEEGG). The efficiency of BE-META simulations biasing 18 ϕ, ψ, and χ dihedrals[126] is compared to BE-2D-META simulations biasing 8 ($\phi_i \times \psi_i$) and 8 ($\psi_i \times \phi_{i+1}$) for cyclo-(YNPFEEGG). **Figure 2.9A** shows how the conformational density profiles for the four simulations change as simulation time increases. For the BE-META simulations, the location of the most populated cluster for S1 and S2 do not agree until after 350 ns. The population of this most populated cluster in S1 and S2 starts to converge during the remaining 350–500 ns. In contrast, the most populated state is identified within 150 ns of BE-2D-META simulations. The population of this most populated cluster remains fairly constant during the remaining 200–500 ns of BE-2D-META simulations. The normalized integrated product plot in **Figure 2.9B** also illustrates the faster converging behavior of BE-2D-META simulations compared to BE-META simulations of cyclo-(YNPFEEGG). Therefore, by targeting the mechanisms of coupled two-dihedral changes that underlie conformational switches of cyclic peptides, complete structural ensembles can be efficiently obtained for cyclic peptides.

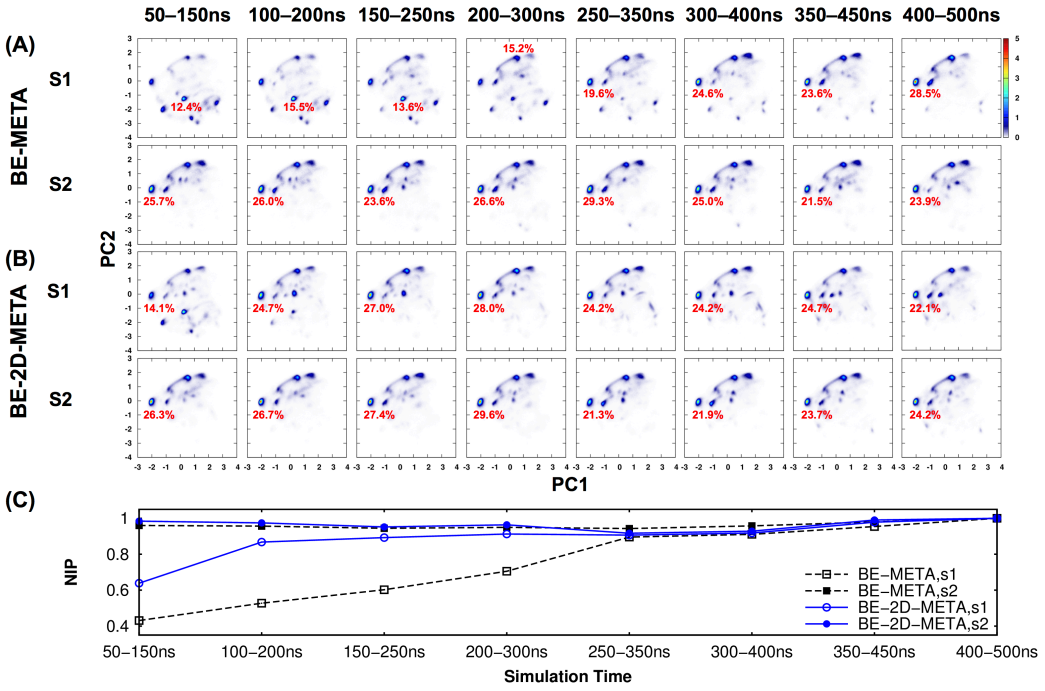

Figure 2.9 Conformational density profiles as a function of the first two largest principal components of cyclo-(YNPFEEGG). **(A)** 500-ns BE-META simulations and **(B)** 500-ns BE-2D-META simulations (in 100-ns increments) starting from two initial configurations. The most populated cluster for each time increment is indicated by the given population shown in red. Principal component analysis was performed on all four trajectories simultaneously. (**C**) Normalized integrated product (NIP) of the density profile and the converged result at different simulation times. Density profiles without any overlap have NIP values of zero; density profiles with perfect overlap have NIP values of one.[141] (Credit: Sean M. McHugh)

2.6 – Conclusions

Although cyclic peptides have great potential to selectively mediate protein–protein interactions for both mechanistic investigations of PPIs and therapeutic treatments of various diseases, the rational design of new cyclic peptides remains limited. To improve upon current simulation methods for cyclic peptides, I first investigate how small cyclic peptides change conformations. From analysis of MD simulations of cyclo-(GGGGGG), cyclo-(AAAAAA), and cyclo-(YNPFEEGG), conformational changes are observed to occur via mechanisms of coupled two-dihedral changes, either coupled changes of ϕ_i and ψ_i or coupled changes of ψ_i and ϕ_{i+1}. These mechanistic insights into the conformation changes inherent to small cyclic peptides allowed us to develop an improved set of CVs for efficient metadynamics simulations. Complete structural ensembles of cyclic

peptides can be more quickly obtained with the new method, improving upon simulation methods for structure predictions of cyclic peptides.

2.7 – Future directions

Having developed an improved simulation method to sample cyclic peptide structures, the next steps towards the rational design of new cyclic peptide structures includes elucidating initial sequence structure relationships. Indeed the converged structural ensembles for cyclo-(GGGGGG) and cyclo-(AAAAAA) differ: While all ten combinations of β-turns are observed for cyclo-(GGGGGG) with a type I + I' being the most populated (**Figure 2.10**), only eight combinations are observed for cyclo-(AAAAAA) with a type I + II being the most populated and combinations I' + I' and II + I' missing (**Figure 2.11**). To understand these structural preferences, the free energy difference between structures with different combinations of β-turns will be decomposed into enthalpic and entropic contributions.[146] Additionally, water-peptide interactions will be considered as additional reasoning to explain the differing structural preferences. As conformational switches of cyclo-(AAAAAA) require more time than corresponding switches of cyclo-(GGGGGG) (**Figure 2.7** and **Figures 2.4** and **2.5**), water dynamics during conformational transitions will also be investigated.

In addition to performing further analysis on the simulations of cyclo-(AAAAAA) compared to those of cyclo-(GGGGGG), the structural ensembles of cyclic hexapeptides with all combinations of Gly and Ala residues will also be investigated. Since many sequences will need to be simulated, additional enhanced sampling methods will be compared for efficiency. Furthermore, since ten structures with combinations of two β-turns are known to be thermodynamically favorable for cyclic hexapeptides, the problem is somewhat simplified: Only the energetic differences between these ten structures for each sequence need to be determined. One approach consists of using the RMSD to one of ten reference structures with a combination of β-turns to bias each of ten replicas for BE-META simulations. Additionally, the utility and efficiency of

other approaches, such as the confinement method[147-150] and variational method,[151-154] will be investigated The converging behavior of these methods will be compared to our newly developed BE-2D-META method determine the most efficient simulation method to determine further sequence-structure relationships for small cyclic peptides.

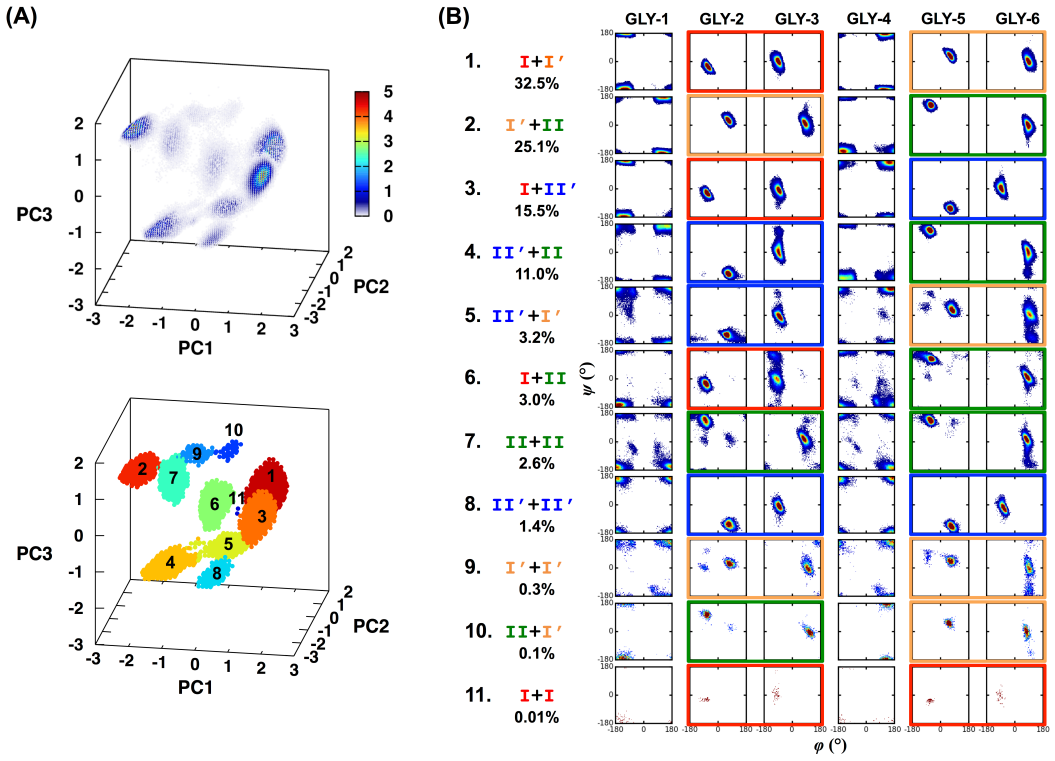

Figure 2.10 (A) Conformational density profile in 3D principal component space (top) and subsequent cluster analysis (bottom) from MD simulations of cyclo-(GGGGGG) using the 50–1000 ns. **(B)** Ramachandran plots for each cluster. β-turns are boxed and colored by turn type as shown in **Figure 2.1**. (Credit: Sean M. McHugh)

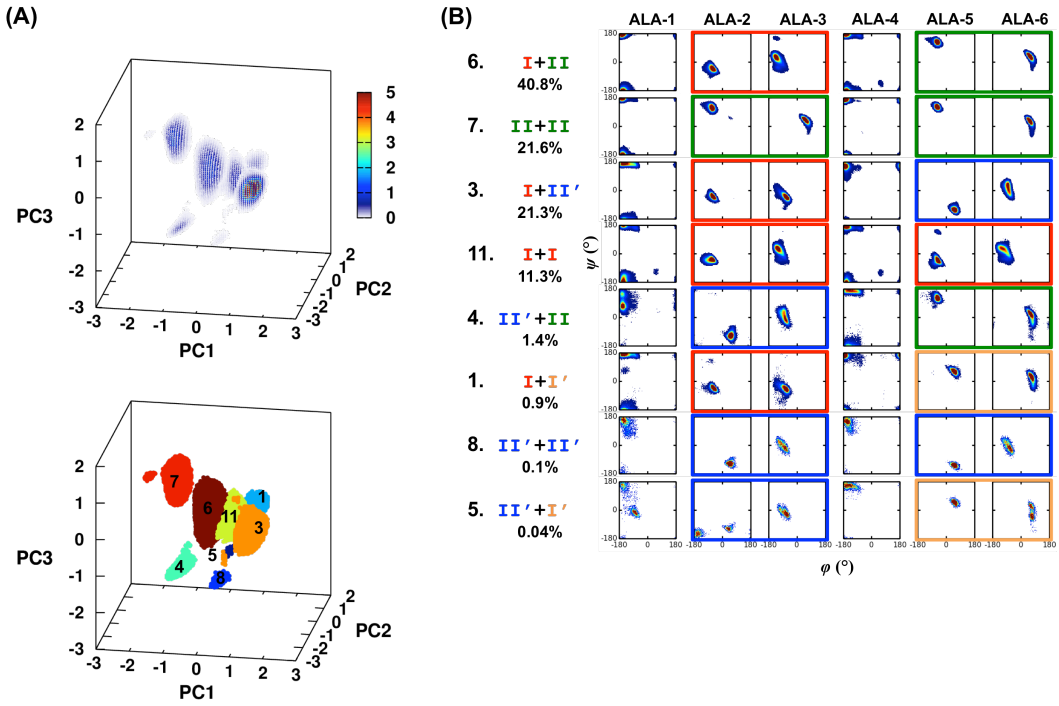

Figure 2.11 (A) Conformational density profile in 3D principal component space (top) and subsequent cluster analysis (bottom) from BE-2D-META simulations of cyclo-(AAAAAA) using 50–1000 ns. PC1, PC2, and PC3 were determined from MD simulations of cyclo-(GGGGGG). **(B)** Ramachandran plots for each cluster. Clusters are numbered according to the clusters observed for cyclo-(GGGGGG) in **Figure 2.4**. β-turns are boxed and colored by turn type as shown in **Figure 2.1**.

References

(1) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. *Sci. Rep.* **2011**, *1*, 90.
(2) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4.
(3) Spiro, R. G. *Glycobiology* **2002**, *12*, 43r.
(4) *Essentials of Glycobiology*; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2009.
(5) Defaus, S.; Gupta, P.; Andreu, D.; Gutierrez-Gallego, R. *Analyst* **2014**, *139*, 2944.
(6) Freeze, H. H.; Eklund, E. A.; Ng, B. G.; Patterson, M. C. *Annu. Rev. Neurosci.* **2015**, *38*, 105.
(7) Hart, G. W.; Copeland, R. J. *Cell* **2010**, *143*, 672.
(8) Hebert, D. N.; Lamriben, L.; Powers, E. T.; Kelly, J. W. *Nat. Chem. Biol.* **2014**, *10*, 902.
(9) Parodi, A. J. *Biochem. J.* **2000**, *348 Pt 1*, 1.
(10) Rickert, K. W.; Imperiali, B. *Chem. Biol.* **1995**, *2*, 751.
(11) Imperiali, B.; Rickert, K. W. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 97.
(12) Hanson, S. R.; Culyba, E. K.; Hsu, T. L.; Wong, C. H.; Kelly, J. W.; Powers, E. T. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3131.
(13) Imperiali, B.; O'Connor, S. E. *Curr. Opin. Chem. Biol.* **1999**, *3*, 643.
(14) O'Connor, S. E.; Imperiali, B. *Chem. Biol.* **1996**, *3*, 803.
(15) Imperiali, B.; Shannon, K. L. *Biochemistry* **1991**, *30*, 4374.
(16) Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.* **2004**, *126*, 8421.
(17) Ellis, C. R.; Noid, W. G. *J. Phys. Chem. B* **2014**, *118*, 11462.
(18) Lee, H. S.; Qi, Y.; Im, W. *Sci. Rep.* **2015**, *5*, 8926.
(19) Wormald, M. R.; Dwek, R. A. *Structure* **1999**, *7*, R155.
(20) Wang, X. Y.; Ji, C. G.; Zhang, J. Z. *J. Phys. Chem. B* **2012**, *116*, 11570.
(21) Wyss, D. F.; Choi, J. S.; Li, J.; Knoppers, M. H.; Willis, K. J.; Arulanandam, A. R.; Smolyar, A.; Reinherz, E. L.; Wagner, G. *Science* **1995**, *269*, 1273.
(22) Price, J. L.; Culyba, E. K.; Chen, W.; Murray, A. N.; Hanson, S. R.; Wong, C.-H.; Powers, E. T.; Kelly, J. W. *Biopolymers* **2012**, *98*, 195.
(23) Gavrilov, Y.; Shental-Bechor, D.; Greenblatt, H. M.; Levy, Y. *J. Phys. Chem. Lett.* **2015**, *6*, 3572.
(24) Shental-Bechor, D.; Levy, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 8256.
(25) Lu, D.; Yang, C.; Liu, Z. *J. Phys. Chem. B* **2012**, *116*, 390.
(26) Ellis, C. R.; Maiti, B.; Noid, W. G. *J. Am. Chem. Soc.* **2012**, *134*, 8184.
(27) Culyba, E. K.; Price, J. L.; Hanson, S. R.; Dhar, A.; Wong, C. H.; Gruebele, M.; Powers, E. T.; Kelly, J. W. *Science* **2011**, *331*, 571.
(28) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14127.

(29) Dalziel, M.; Crispin, M.; Scanlan, C. N.; Zitzmann, N.; Dwek, R. A. *Science* **2014**, *343*, 1235681.
(30) Sola, R. J.; Griebenow, K. *J. Pharm. Sci.* **2009**, *98*, 1223.
(31) Walsh, G.; Jefferis, R. *Nat. Biotechnol.* **2006**, *24*, 1241.
(32) Chen, W.; Enck, S.; Price, J. L.; Powers, D. L.; Powers, E. T.; Wong, C.-H.; Dyson, H. J.; Kelly, J. W. *J. Am. Chem. Soc.* **2013**, *135*, 9877.
(33) Petrescu, A. J.; Milac, A. L.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R. *Glycobiology* **2004**, *14*, 103.
(34) Surleac, M. D.; Spiridon, L. N.; Tacutu, R.; Milac, A. L.; Petrescu, S. M.; Petrescu, A.-J. In *Glycosylation*; Petrescu, S., Ed. 2012.
(35) Zielinska, D. F.; Gnad, F.; Schropp, K.; Wisniewski, J. R.; Mann, M. *Mol. Cells* **2012**, *46*, 542.
(36) Zielinska, D. F.; Gnad, F.; Wisniewski, J. R.; Mann, M. *Cell* **2010**, *141*, 897.
(37) Chen, M. M.; Bartlett, A. I.; Nerenberg, P. S.; Friel, C. T.; Hackenberger, C. P.; Stultz, C. M.; Radford, S. E.; Imperiali, B. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 22528.
(38) Otvos, L., Jr.; Thurin, J.; Kollat, E.; Urge, L.; Mantsch, H. H.; Hollosi, M. *Int. J. Pept. Protein Res.* **1991**, *38*, 476.
(39) Otvos, L., Jr.; Urge, L.; Xiang, Z. Q.; Krivulka, G. R.; Nagy, L.; Szendrei, G. I.; Ertl, H. C. *Biochim. Biophys. Acta* **1994**, *1224*, 68.
(40) Tagashira, M.; Tanaka, A.; Hisatani, K.; Isogai, Y.; Hori, M.; Takamatsu, S.; Fujibayashi, Y.; Yamamoto, K.; Haneda, K.; Inazu, T.; Toma, K. *Glycoconjugate J.* **2001**, *18*, 449.
(41) Hashimoto, Y.; Toma, K.; Nishikido, J.; Yamamoto, K.; Haneda, K.; Inazu, T.; Valentine, K. G.; Opella, S. J. *Biochemistry* **1999**, *38*, 8377.
(42) Palian, M. M.; Jacobsen, N. E.; Polt, R. *J. Pept. Res.* **2001**, *58*, 180.
(43) Urge, L.; Jackson, D. C.; Gorbics, L.; Wroblewski, K.; Graczyk, G.; Otvos Jr, L. *Tetrahedron* **1994**, *50*, 2373.
(44) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605.
(45) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622.
(46) Dunbrack Jr, R. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431.
(47) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 400.
(48) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins* **2000**, *40*, 389.
(49) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. *J. Mol. Biol.* **1978**, *125*, 357.
(50) Imberty, A.; Perez, S. *Protein Eng.* **1995**, *8*, 699.
(51) Davis, J. T.; Hirani, S.; Bartlett, C.; Reid, B. R. *J. Biol. Chem.* **1994**, *269*, 3331.
(52) UniProtConsortium *Nucleic Acids Res.* **2015**, *43*, D204.
(53) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *3*, 535.

(54) Lütteke, T.; von der Lieth, C.-W. *BMC Bioinform.* **2004**, *5*, 69.
(55) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(56) Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566.
(57) Mu, Y.; Nguyen, P. H.; Stock, G. *Proteins* **2005**, *58*, 45.
(58) Sittel, F.; Jain, A.; Stock, G. *J. Chem. Phys.* **2014**, *141*, 014111.
(59) Rodriguez, A.; Laio, A. *Science* **2014**, *344*, 1492.
(60) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.
(61) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47.
(62) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 3257.
(63) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.
(64) Casey, A.; Walsh, G. *Bioresource Technol.* **2003**, *86*, 183.
(65) Fonseca-Maldonado, R.; Maller, A.; Bonneil, E.; Thibault, P.; Botelho-Machado, C.; Ward, R. J.; Polizeli Mde, L. *Protein Expr. Purif.* **2014**, *99*, 43.
(66) Guo, M.; Hang, H.; Zhu, T.; Zhuang, Y.; Chu, J.; Zhang, S. *Enzyme Microb. Technol.* **2008**, *42*, 340.
(67) Han, Y.; Lei, X. G. *Arch. Biochem. Biophys.* **1999**, *364*, 83.
(68) Repo, H.; Kuokkanen, E.; Oksanen, E.; Goldman, A.; Heikinheimo, P. *Proteins* **2014**, *82*, 300.
(69) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. *Mol. Syst. Biol.* **2011**, *7*.
(70) Dierks, T.; Dickmanns, A.; Preusser-Kunze, A.; Schmidt, B.; Mariappan, M.; von Figura, K.; Ficner, R.; Rudolph, M. G. *Cell* **2005**, *121*, 541.
(71) Roeser, D.; Preusser-Kunze, A.; Schmidt, B.; Gasow, K.; Wittmann, J. G.; Dierks, T.; von Figura, K.; Rudolph, M. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 81.
(72) Schlotawa, L.; Radhakrishnan, K.; Baumgartner, M.; Schmid, R.; Schmidt, B.; Dierks, T.; Gartner, J. *Eur. J. Hum. Genet.* **2013**, *21*, 1020.
(73) Anthony, C. S.; Corradi, H. R.; Schwager, S. L. U.; Redelinghuys, P.; Georgiadis, D.; Dive, V.; Acharya, K. R.; Sturrock, E. D. *J. Biol. Chem.* **2010**, *285*, 35685.
(74) Corradi, H. R.; Schwager, S. L. U.; Nchinda, A. T.; Sturrock, E. D.; Acharya, K. R. *J. Mol. Biol.* **2006**, *357*, 964.
(75) Roberts, S. J.; Petropavlovskaja, M.; Chung, K. N.; Knight, C. B.; Elwood, P. C. *Arch. Biochem. Biophys.* **1998**, *351*, 227.
(76) Shen, F.; Wang, H. Q.; Zheng, X.; Ratnam, M. *Biochem. J.* **1997**, *327*, 759.
(77) Broccolini, A.; Gidaro, T.; De Cristofaro, R.; Morosetti, R.; Gliubizzi, C.; Ricci, E.; Tonali, P. A.; Mirabella, M. *J. Neurochem.* **2008**, *105*, 971.
(78) Lafrance, M. H.; Vezina, C.; Wang, Q.; Boileau, G.; Crine, P.; Lemay, G. *Biochem. J.* **1994**, *302*, 451.
(79) Kwong, P. D.; Wyatt, R.; Desjardins, E.; Robinson, J.; Culp, J. S.; Hellmig, B. D.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *J. Biol. Chem.* **1999**, *274*, 4115.

(80) Fusco, D.; Barnum, T. J.; Bruno, A. E.; Luft, J. R.; Snell, E. H.; Mukherjee, S.; Charbonneau, P. *PLoS ONE* **2014**, *9*, e101123.
(81) Price, W. N., 2nd; Chen, Y.; Handelman, S. K.; Neely, H.; Manor, P.; Karlin, R.; Nair, R.; Liu, J.; Baran, M.; Everett, J.; Tong, S. N.; Forouhar, F.; Swaminathan, S. S.; Acton, T.; Xiao, R.; Luft, J. R.; Lauricella, A.; DeTitta, G. T.; Rost, B.; Montelione, G. T.; Hunt, J. F. *Nat. Biotechnol.* **2009**, *27*, 51.
(82) Derewenda, Z. S. *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67*, 243.
(83) Juers, D. H.; Matthews, B. W. *J. Mol. Biol.* **2001**, *311*, 851.
(84) Lehninger, A.; Nelson, D.; Cox, M. *Lehninger Principles of Biochemistry*; W. H. Freeman, 2008.
(85) Yang, J.; Zhao, K.; Gong, Y.; Vologodskii, A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1998**, *120*, 10646.
(86) Louis-Jeune, C.; Andrade-Navarro, M. A.; Perez-Iratxeta, C. *Proteins* **2012**, *80*, 374.
(87) Shental-Bechor, D.; Levy, Y. *Curr. Opin. Struct. Biol.* **2009**, *19*, 524.
(88) Olson, C. A.; Shi, Z.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2001**, *123*, 6451.
(89) Olson, C. A.; Spek, E. J.; Shi, Z.; Vologodskii, A.; Kallenbach, N. R. *Proteins* **2001**, *44*, 123.
(90) Shi, Z.; Olson, C. A.; Bell, A. J., Jr.; Kallenbach, N. R. *Biophys. Chem.* **2002**, *101-102*, 267.
(91) Zhou, G.; Voelz, V. A. *J. Phys. Chem. B* **2016**, *120*, 926.
(92) Pawson, T.; Nash, P. *Genes Dev.* **2000**, *14*, 1027.
(93) Ryan, D. P.; Matthews, J. M. *Curr. Opin. Struct. Biol.* **2005**, *15*, 441.
(94) Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001.
(95) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13.
(96) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177.
(97) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. *Nat. Biotechnol.* **2007**, *25*, 71.
(98) Smith, R. D.; Hu, L.; Falkner, J. A.; Benson, M. L.; Nerothin, J. P.; Carlson, H. A. *J. Mol. Graphics Model.* **2006**, *24*, 414.
(99) Bhat, A.; Roberts, L. R.; Dwyer, J. J. *Eur. J. Med. Chem.* **2015**, *94*, 471.
(100) Bock, J. E.; Gavenonis, J.; Kritzer, J. A. *ACS Chem. Biol.* **2013**, *8*, 488.
(101) Giordanetto, F.; Kihlberg, J. *J. Med. Chem.* **2014**, *57*, 278.
(102) Katsara, M.; Tselios, T.; Deraos, S.; Deraos, G.; Matsoukas, M. T.; Lazoura, E.; Matsoukas, J.; Apostolopoulos, V. *Curr. Med. Chem.* **2006**, *13*, 2221.
(103) Tapeinou, A.; Matsoukas, M. T.; Simal, C.; Tselios, T. *Biopolymers* **2015**, *104*, 453.
(104) Thapa, P.; Espiritu, M.; Cabalteja, C.; Bingham, J.-P. *Int. J. Pep. Res. Ther.* **2014**, *20*, 545.
(105) Yudin, A. K. *Chem. Sci.* **2015**, *6*, 30.
(106) Kling, A.; Lukat, P.; Almeida, D. V.; Bauer, A.; Fontaine, E.; Sordello, S.; Zaburannyi, N.; Herrmann, J.; Wenzel, S. C.; Konig, C.; Ammerman, N. C.; Barrio, M. B.; Borchers, K.; Bordon-Pallier, F.; Bronstrup, M.; Courtemanche, G.; Gerlitz, M.; Geslin, M.; Hammann, P.; Heinz, D. W.; Hoffmann, H.; Klieber,

S.; Kohlmann, M.; Kurz, M.; Lair, C.; Matter, H.; Nuermberger, E.; Tyagi, S.; Fraisse, L.; Grosset, J. H.; Lagrange, S.; Muller, R. *Science* **2015**, *348*, 1106.
(107) Loll, P. J.; Axelsen, P. H. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 265.
(108) Demmer, O.; Frank, A. O.; Hagn, F.; Schottelius, M.; Marinelli, L.; Cosconati, S.; Brack-Werner, R.; Kremb, S.; Wester, H.-J.; Kessler, H. *Angew. Chem., Int. Ed.* **2012**, *51*, 8110.
(109) Miranda, E.; Nordgren, I. K.; Male, A. L.; Lawrence, C. E.; Hoakwie, F.; Cuda, F.; Court, W.; Fox, K. R.; Townsend, P. A.; Packham, G. K.; Eccles, S. A.; Tavassoli, A. *J. Am. Chem. Soc.* **2013**, *135*, 10418.
(110) Morse, R. P.; Willett, J. L. E.; Johnson, P. M.; Zheng, J.; Credali, A.; Iniguez, A.; Nowick, J. S.; Hayes, C. S.; Goulding, C. W. *J. Mol. Biol.* **2015**, *427*, 3766.
(111) Xiong, J. P.; Stehle, T.; Zhang, R. G.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A. *Science* **2002**, *296*, 151.
(112) Gierasch, L. M.; Karle, I. L.; Rockwell, A. L.; Yenal, K. *J. Am. Chem. Soc.* **1985**, *107*, 3321.
(113) Ramakrishnan, C.; Paul, P. K. C.; Ramnarayan, K. *Proc. Int. Symp. Biomol. Struct, Interactions, Suppl. J. Biosci.* **1985**, *8*, 239.
(114) Matter, H.; Kessler, H. *J. Am. Chem. Soc.* **1995**, *117*, 3347.
(115) Bean, J. W.; Kopple, K. D.; Peishoff, C. E. *J. Am. Chem. Soc.* **1992**, *114*, 5328.
(116) Kessler, H.; Gratias, R.; Hessler, G.; Gurrath, M.; Muller, G. *Pure Appl. Chem* **1996**, *68*, 1201.
(117) Kopple, K. D.; Bean, J. W.; Bhandary, K. K.; Briand, J.; Dambrosio, C. A.; Peishoff, C. E. *Biopolymers* **1993**, *33*, 1093.
(118) Kopple, K. D.; Savrda, J.; Logan, R. H.; Go, A. *J. Am. Chem. Soc.* **1972**, *94*, 973.
(119) Wu, L.; Lu, Y.; Zheng, Q.-T.; Tan, N.-H.; Li, C.-M.; Zhou, J. *J. Mol. Struct.* **2007**, *827*, 145.
(120) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.
(121) Dama, J. F.; Parrinello, M.; Voth, G. A. *Phys. Rev. Lett.* **2014**, *112*, 240602.
(122) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553.
(123) Damas, J. M.; Filipe, L. C. S.; Campos, S. R. R.; Lousa, D.; Victor, B. L.; Baptista, A. M.; Soares, C. M. *J. Chem. Theory Comput.* **2013**, *9*, 5148.
(124) Paissoni, C.; Ghitti, M.; Belvisi, L.; Spitaleri, A.; Musco, G. *Chem.--Eur. J.* **2015**, *21*, 14165.
(125) Spitaleri, A.; Ghitti, M.; Mari, S.; Alberici, L.; Traversari, C.; Rizzardi, G.-P.; Musco, G. *Angew. Chem., Int. Ed.* **2011**, *50*, 1832.
(126) Yu, H.; Lin, Y.-S. *Phys. Chem. Chem. Phys.* **2015**, *17*, 4210.
(127) Jiang, F.; Zhou, C. Y.; Wu, Y. D. *J. Phys. Chem. B* **2014**, *118*, 6983.
(128) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665.
(129) Karle, I. L.; Gibson, J. W.; Karle, J. *J. Am. Chem. Soc.* **1970**, *92*, 3755.

(130) Kartha, G.; Bhandary, K. K.; Kopple, K. D.; Go, A.; Zhu, P.-P. *J. Am. Chem. Soc.* **1984**, *106*, 3844.
(131) Yang, C.-H.; Brown, J. N.; Kopple, K. D. *J. Am. Chem. Soc.* **1981**, *103*, 1715.
(132) Bovey, F. A.; Brewster, A. I.; Patel, D. J.; Tonelli, A. E.; Torchia, D. A. *Acc. Chem. Res.* **1972**, *5*, 193.
(133) Torchia, D. A.; Wong, S. C. K.; Deber, C. M.; Blout, E. R. *J. Am. Chem. Soc.* **1972**, *94*, 616.
(134) Blout, E. R. *Biopolymers* **1981**, *20*, 1901.
(135) Kopple, K. D.; Go, A.; Schamper, T. J. *J. Am. Chem. Soc.* **1978**, *100*, 4289.
(136) Kostansek, E. C.; Lipscomb, W. N.; Thiessen, W. E. *J. Am. Chem. Soc.* **1979**, *101*, 834.
(137) Kostansek, E. C.; Thiessen, W. E.; Schomburg, D.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1979**, *101*, 5811.
(138) Nielsen, D. S.; Lohman, R.-J.; Hoang, H. N.; Hill, T. A.; Jones, A.; Lucke, A. J.; Fairlie, D. P. *ChemBioChem* **2015**, *16*, 2289.
(139) Torchia, D. A.; Wong, S. C. K.; Blout, E. R.; Corato, A. D.; Deber, C. M. *J. Am. Chem. Soc.* **1972**, *94*, 609.
(140) Hossain, M. B.; Van der Helm, D. *J. Am. Chem. Soc.* **1978**, *100*, 5191.
(141) McHugh, S. M.; Rogers, J. R.; Yu, H.; Lin, Y.-S. *J. Chem. Theory Comput.* **2016**.
(142) Kamens, A. J.; Eisert, R. J.; Corlin, T.; Baleja, J. D.; Kritzer, J. A. *Biochemistry* **2014**, *53*, 4758.
(143) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(144) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(145) MacKerell, A. D., Jr.; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257.
(146) Portella, G.; Hub, J. S.; Vesper, M. D.; de Groot, B. L. *Biophys. J.* **2008**, *95*, 2275.
(147) Capelli, R.; Villemot, F.; Moroni, E.; Tiana, G.; van der Vaart, A.; Colombo, G. *J. Phys. Chem. Lett.* **2016**, *7*, 126.
(148) Cecchini, M.; Krivov, S. V.; Spichty, M.; Karplus, M. *J. Phys. Chem. B* **2009**, *113*, 9728.
(149) Esque, J.; Cecchini, M. *J. Phys. Chem. B* **2015**, *119*, 5194.
(150) Tyka, M. D.; Clarke, A. R.; Sessions, R. B. *J. Phys. Chem. B* **2006**, *110*, 17212.
(151) McCarty, J.; Valsson, O.; Tiwary, P.; Parrinello, M. *Phys. Rev. Lett.* **2015**, *115*.
(152) Shaffer, P.; Valsson, O.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 1150.
(153) Valsson, O.; Parrinello, M. *Phys. Rev. Lett.* **2014**, *113*.
(154) Valsson, O.; Parrinello, M. *J. Chem. Theory Comput.* **2015**, *11*, 1996.