*Pure & Appl. Chem.*, Vol. 70, No. 1, pp. 33–40, 1998.
Printed in Great Britain.


# The conformational basis of asparagine-linked glycosylation

Barbara Imperiali* and Sarah E. O'Connor

*Division of Chemistry and Chemical Engineering 164-30*
*California Institute of Technology,*
*Pasadena, CA 91125, USA*

*Abstract:*
The structural and functional integrity of many proteins is highly dependent upon enzyme catalyzed covalent modification reactions. The timing of these transformations may be co-translational, occurring as a protein is being biosynthesized or post-translational, acting on folded protein substrates. Our research focuses on understanding the conformational basis of specificity and selectivity in the process of co-translational protein glycosylation. This objective presents a significant challenge because the opportunities for competing transformations on the densely functionalized protein substrate in this reaction are innumerable. Since enzyme-catalyzed glycosylation is an essential eukaryotic process, an understanding of the origin of specificity is of utmost importance both to fundamental biochemistry and to a consideration of the mechanisms of homeostatic control.

## Introduction

Protein glycosylation impacts both the functional capacity and structural framework of all glycoproteins (1). The carbohydrate modifications of proteins fall into three general categories: *N*-linked modification of asparagine (2, 3, 4), *O*-linked modification of serine or threonine (5), and glycosylphosphatidyl inositol derivatization of the C-terminus carboxyl group (6) (Fig. 1). Each of these transformations is catalyzed by one or more enzymes which demonstrate different peptide sequence requirements and reaction specificities.

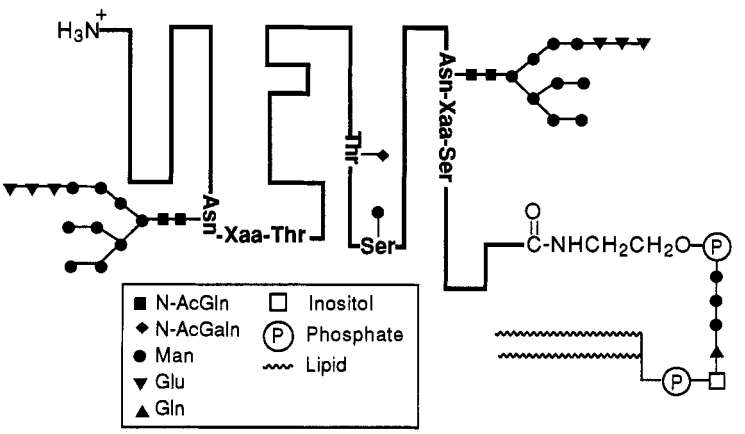



Fig. 1 Schematic illustrating the various forms protein glycosylation

*N*-linked glycosylation is catalyzed by a single enzyme, oligosaccharyl transferase (OT), and involves the co-translational transfer of a lipid-linked tetradecasaccharide ($GlcNAc_2$-$Man_9$-$Glc_3$) to an asparagine side chain (in the consensus sequence Asn-Xaa-Ser/Thr) within a nascent polypeptide. The subsequent diversification of these conjugates arises from enzyme catalyzed processing steps that occur in the endoplasmic reticulum (ER) and Golgi apparatus *after* the addition of the first triantennary oligosaccharide complex. In contrast, *O*-linked glycosylation proceeds through the direct, post-translational transfer of single monosaccharides to threonine and/or serine residues within folded polypeptides. In general, the peptide sequence specificities of the *O*-linked glycosylation enzymes are not well understood and activity is attributed to several different enzymes (5). In addition, proteins can be modified with a complex glycosyl phosphatidyl inositol (GPI) construct at the polypeptide carboxyl terminus; this derivatization confers new properties by anchoring the protein to membrane bilayers (6).

This paper will focus on *N*-linked glycosylation, which is the most common of the eukaryotic glycosylation reactions (2). The reaction is illustrated in Fig. 2. This co-translational event occurs in the ER while the polypeptide is being biosynthesized on membrane associated ribosomes. Approximately 14 residues of the nascent peptide must clear the luminal surface of the ER membrane before OT mediated glycosylation can occur, thereby implying that the active site of OT resides in the soluble domain of the enzyme (7).

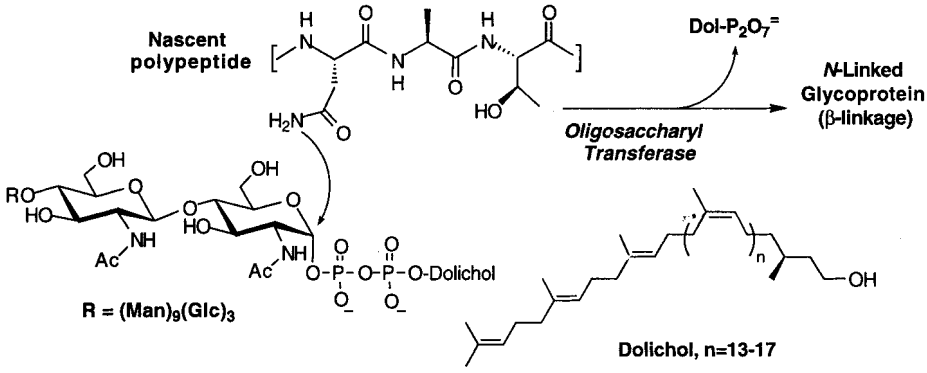



Fig. 2 Reaction catalyzed by oligosaccharyl transferase

The function of the carbohydrate in glycoproteins is unclear in many cases (8). Study of *N*-linked glycoproteins is complicated for a variety of reasons, including the need for eukaryotic expression systems. Furthermore, both the mobility and the heterogeneity of carbohydrates hamper crystallization and complicate NMR structural analysis. Despite these difficulties, some specific structural roles of the glycan have been illuminated. Glycosylation has been shown to maintain protein structure and stability through long range hydrogen bonding and hydrophobic interactions between the oligosaccharide and the protein in several cases. Moreover, many immature proteins misfold in the absence of glycosylation, indicating that this co-translational event may assist in the direction of correct folding pathways (9).

## Conformational Requirements for Asparagine-Linked Glycosylation

OT glycosylates at the tripeptide recognition sequence Asn-Xaa-Thr/Ser, where Xaa represents any of the encoded amino acids except proline (10). Because *N*-linked glycosylation occurs before the nascent polypeptide has completely folded, the global protein structure presumably does not play a role in the recognition events that lead to this modification. However, since tripeptides can act as substrates for OT, the local secondary structure of the recognition sequence and adjacent peptide may well enable recognition by

the enzyme and modulate the reactivity of the amide nitrogen that acts as a nucleophile to form the protein-carbohydrate linkage.

The ability of OT to recognize and glycosylate short peptidyl substrates has facilitated investigation of the structural requirements for glycosylation. Specifically, peptide analogs which contain the recognition elements of the consensus sequence, but which are constrained to specific conformations have been designed and synthesized. The parallel structural and kinetic analysis of these compounds has provided an understanding of the role of conformation in asparagine-linked glycosylation. This new knowledge has formed a foundation for the development of a mechanistic model for the enzyme-catalyzed modification of the asparagine amide.

Since tripeptides are accepted as substrates, the length of the consensus triad primarily limits the types of available hydrogen-bonded motifs to simple turns, since other structural features such as α-helices or β-sheets tend to require longer peptides for complete formation. The tripeptide substrates would be compatible with two types of turns: the β-turn and the Asx-turn (11). These two turns are illustrated in Fig. 3. The β-turn forms a complete chain reversal, and is characterized by a hydrogen bond between the threonine amide and the carbonyl immediately preceding the asparagine. The Asx-turn involves a non-covalent interaction similar to that of a β-turn, however, in this case, the hydrogen bond acceptor is the carbonyl oxygen of the asparagine side chain. The polypeptide backbone has a more extended conformation and does not manifest a complete chain reversal. This hydrogen-bonding motif is rather common; approximately 18% of all asparagine and aspartic acid side chains appear to be involved in Asx-turns in proteins (12). It is noteworthy that the homologous residue glutamine is never glycosylated and this apparent contradiction in reactivity may be explained by the distinct conformational preferences of this residue. Specifically, the carboxamide side chain of glutamine is seldom involved in short range hydrogen-bonding interactions (12). The possibility that the Asx-turn or the β-turn is the unique secondary structure preferentially recognized by OT has been considered as an explanation for the both specificity of the enzyme and the enhanced nucleophilicity of the peptidyl substrate, particularly in model mechanisms which rely on secondary structure to position the hydroxyl moiety of the essential serine or threonine within direct proximity to the asparagine side chain.

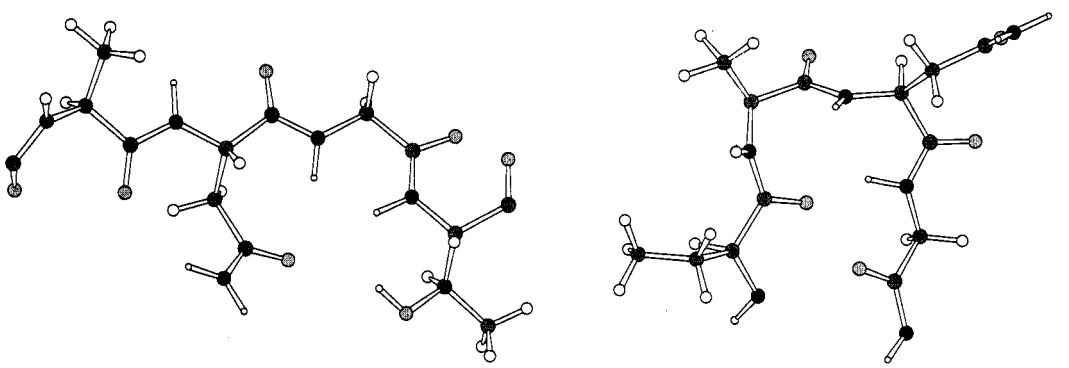
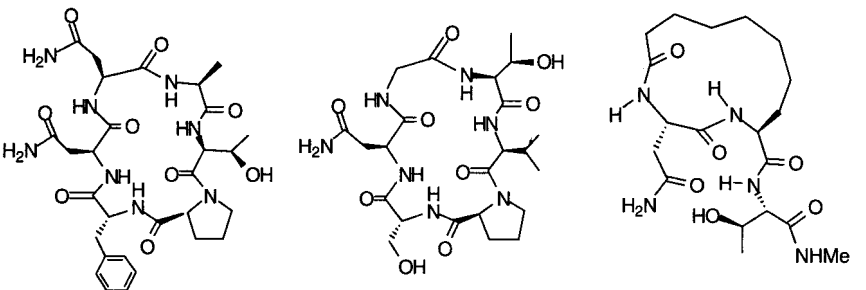



Fig. 3 Comparison of the structures of a Type I β-turn and an Asx-turn. Carbon is shown is gray, hydrogen in white, nitrogen in black and oxygen in speckled.

To rigorously distinguish between the Asx-turn and the β-turn, a study was carried out which placed the reactive tripeptide sequence with asparagine in either the (i) or (i + 1) position of a constrained type I β-turn that had been built into the architecture of a cyclic

hexapeptide (**1** and **2**) (13). The type I β-turn was fixed by the incorporation of a prolyl-D-amino acid dipeptide at the non-reactive end of the cyclic compound. This dipeptide sequence imposes a constrained type II β-turn in the cyclic peptide and therefore limits the conformations accessible to the structure. One and two dimensional NMR studies were utilized to verify the turn conformations of compounds **1** and **2**. These two compounds were assayed for substrate activity with OT and compared to the uncyclized analogs. With the acceptor asparagine conformationally constrained into a β-turn (as in compounds **1** and **2**), the resultant peptides failed to exhibit any detectable substrate behavior with OT.




In complementary studies, the Asn-Xaa-Thr was constrained, through a side chain to main chain lactam cyclization to an Asx-turn in the peptide c[Asn-Add]-Thr-NHMe, **3** (14). The unnatural amino acid, (*S*)-2-aminodecanedioic acid (Add), was utilized in the central position to provide functionality for cyclization with the amino terminus. The presence of an Asx-turn was confirmed through NMR and CD studies in aqueous media. When **3** was analyzed for substrate activity with OT and compared to the linear analog, $N^{\alpha}$-butanoyl-Asn-Leu-Thr-NHMe, it was established that introduction of the cyclic constraint enhanced the $K_M$ ten-fold, from 800 mM to 78 mM. These studies provided clear evidence that the preorganization provided by the Asx-turn enhanced enzyme/substrate affinity. The relative maximal velocities of the two compounds remained similar. The enhancement of enzyme affinity for the constrained peptide firmly demonstrates the importance of the Asx-turn over the β-turn as the recognition motif for OT.

A remarkable correlation between the propensity of a given peptide to form an Asx-turn and substrate behavior is observed. Substitution of glutamine for asparagine disrupts both glycosyl acceptor ability as well as the propensity to form an Asx-turn. Additionally, proline in the central position of the recognition sequence disrupts both the Asx-turn and the substrate properties of the peptide. More importantly, the conformational studies have prompted the development of a mechanistic proposal for OT-mediated catalysis which explains the enhanced asparagine nucleophilicity (15).

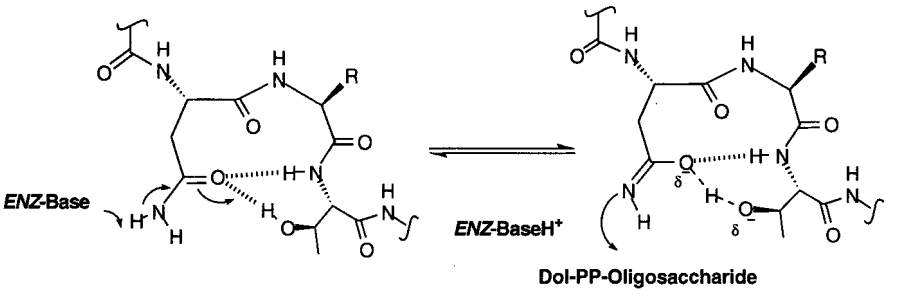



Fig. 4 A proposal for the mechanism of action of oligosaccharyl transferase

In this proposal, the unique hydrogen-bonding array provided by the Asx-turn is suggested to facilitate protonation of the carbonyl of the asparagine side chain while simultaneous, enzyme-mediated deprotonation at the nitrogen would effect the tautomerization of the carboxamide to an imidol species. Thus, tautomerization would afford a neutral, nucleophilic species, which could then react with the electrophilic lipid-linked oligosaccharide (Fig. 4). This mechanism incorporates both the issues of specificity and reactivity as well as the absolute requirement for a hydroxy amino acid by integrating substrate structural requirements with participation of enzyme active site residues. Thus, the likelihood of an Asn-Xaa-Thr/Ser sequence to undergo glycosylation would be governed by the ability of each potential substrate to adopt an Asx-turn conformation within the active site of the enzyme.

## Conformational consequences of asparagine-linked glycosylation

The co-translational timing of the glycosylation event may influence the conformational dynamics of the unfolded peptide proximal to the glycosylation site. Such a structural modulation of the nascent polypeptide may well direct future folding pathways. In fact, many site-directed mutagenesis studies have demonstrated the frequent requirement of *N*-linked glycosylation for proper protein folding (16). It is unsurprising that the large, hydrophilic carbohydrate moiety has a profound impact on the backbone structure of a polypeptide; numerous examples of glycosylation-mediated conformational change for both proteins and small peptides are available. However, the structural or functional reasons for this glycosylation-induced structural modulation are frequently unclear. Although a number of studies have beautifully illustrated how the glycan can stabilize a mature protein, it is difficult to understand how glycosylation can assist protein folding in the nascent polypeptide.

Understanding the significance of *N*-linked glycosylation in the context of protein folding requires a rigorous spectroscopic characterization of glycosylation-mediated conformational change. Short peptides lack the long range interactions found in fully folded proteins and therefore provide a more realistic model system of the flexible, unfolded polypeptide. Moreover, a careful choice of peptide sequence is critical if a correlation between glycosylation-induced conformational change and biological function is to be established. All peptide sequences were taken from natural glycosylation sites that have been demonstrated to significantly impact the folding of the protein. Additionally, all peptides were studied in aqueous media; glycopeptides were derivatized with the disaccharide $GlcNAc_2$, a truncated version of the tetradecasaccharide transferred *in vivo*.

Several short peptides, taken from the hemagglutinin protein and the Fab fragment of IgA, were chosen for spectroscopic analysis (17). The glycosylation sites taken from the hemagglutinin protein were found in β-turn conformations in the native protein. The β-turn is a motif of considerable interest both because it has been identified as a common feature of glycosylation sites (18), and because it has been implicated as a structural nucleation element in protein folding (19). A glycosylation site taken from a β-sheet region of IgA was also examined. The peptides were characterized using time-resolved fluorescence resonance energy transfer (FRET), a technique that is able to measure a single interprobe distance on the same nanosecond timescale over which conformational fluctuations occur. This technique is therefore particularly powerful in the study of short, flexible peptides that exhibit extensive conformational dynamics. After glycosylation, the peptide sequences taken from β-turn regions were found to adopt a more compact conformation, as evidenced by the decreased interfluorophore distance (a decrease from 14.8 Å to 9.5 Å and 12.4 Å to 7.7 Å). In contrast, the peptide taken from the β-sheet of IgA exhibited little change in interfluorophore distance after glycosylation, indicating that its conformation remained largely extended. This structural data is quite significant in light of the ultimate structure adopted by each sequence in the glycosylated protein. Glycosylation may facilitate protein folding by assisting the formation of native structural elements.

The structural characterization of these glycopeptides revealed no specific interactions between the peptide and saccharide thereby suggesting that the conformational changes observed after peptide glycosylation may largely be driven by the nature of the surrounding medium. In support of this hypothesis, FRET studies in 1:1 acetonitrile:water reveal that glycosylation has a negligible impact on peptide conformation in a less polar solvent system. Glycosylation may therefore exert its structural influence by modulating the microenvironment of the peptide, perhaps by altering the local solvation. Additionally, this observation emphasizes the importance of performing such structural studies in aqueous media.

The FRET analysis also revealed that these peptides largely existed in a single conformation, rendering these peptides suitable for study on a slower time scale. NMR was chosen to provide a much more detailed structural picture than the single interfluorophore distance determined by the FRET analysis. NMR studies of the A281-A288 region of the hemagglutinin protein gave results consistent with those obtained from the FRET analysis (Fig. 5).

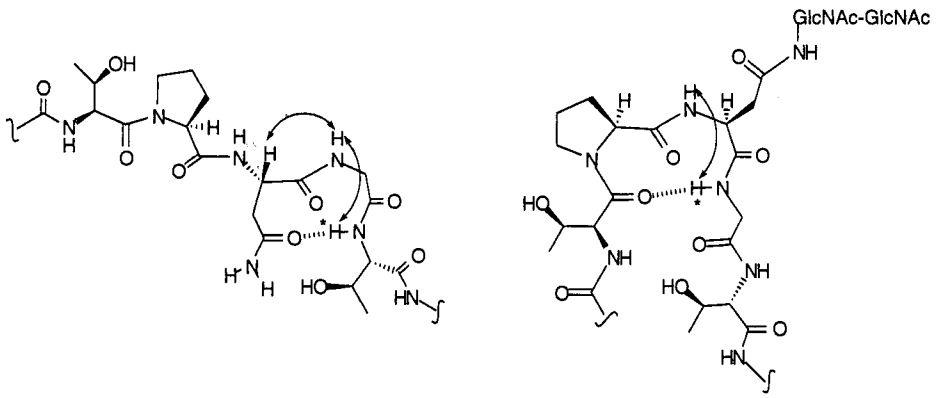



Fig. 5 Several strong key NOEs distinguish between an Asx and β-turn structure. Starred amide protons have a low variable temperature coefficient, suggesting the presence of intermolecular hydrogen bonds.

The extended structure of the non-glycosylated peptide was found to resemble the Asx turn proposed to be required for OT recognition. The more compact structure observed in the FRET studies of the glycosylated derivative strongly resembled the type I β-turn found in the native protein at this sequence. NOE data were incorporated as distance restraints into a simulated annealing protocol to produce the illustrated structures (Fig. 6).

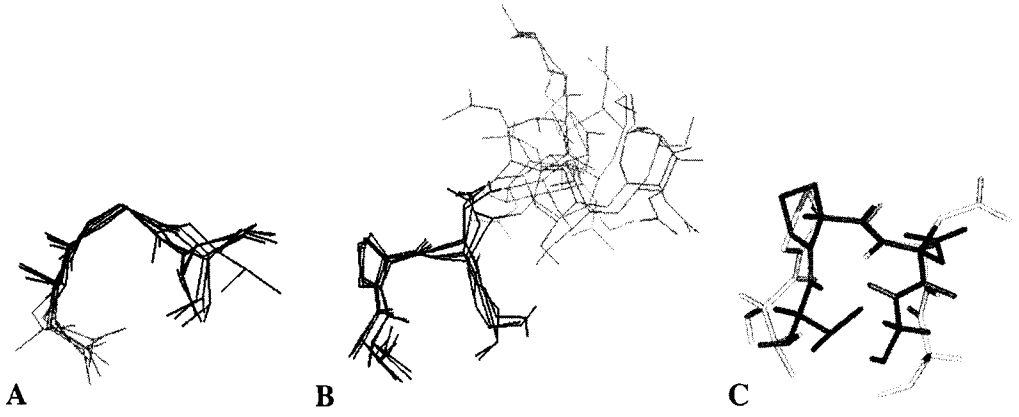



Fig. 6 Structures of A. non-glycosylated and B. glycosylated peptides derived from a simulated annealing procedure that incorporated all NOE data. C. An overlay of a low energy simulated annealing structure of the glycosylated peptide with the crystal structure of the hemagglutinin protein.

An additional peptide model derived from the nicotinic acetylcholine receptor (nAChR) was also examined (20). Each of the five subunits of this oligomeric protein contains a highly conserved extracellular loop region that is thought to be critical for mediating the intersubunit interactions required for protein assembly. The loop sequence of the mature α-1 subunit of the neuromuscular nAChR from *Torpedo californica* comprises residues 128-143. The conserved residues include a glycosylation site (Asn141), a disulfide bond (Cys142 and Cys128), and a proline (Pro136). Glycosylation has been shown to be critical for the assembly and stability of this pentameric complex, but the mechanism by which glycosylation assists the formation of the mature protein is unknown (21). An NMR analysis of the glycosylated and non-glycosylated loop peptides of this α subunit was undertaken to determine if glycosylation affects the conformation of this critical region. The sequence of the peptide is shown in Fig. 7.

Ac-Tyr-Cys-Glu-Ile-Ile-Val-Thr-His-Phe-Pro-Phe-Asp-Gln-Gln-Asn-Cys-Thr-$NH_2$

Fig. 7 Structure of nAChR loop peptide

Kinetic studies of the disulfide bond formation demonstrated that the glycosylated peptide thermodynamically prefers the oxidized state by $\Delta\Delta G$ of 1.4 kJ $mol^{-1}$. Glycosylation may therefore modulate the conformation of the reduced peptide by bringing the cysteines closer together to facilitate the formation of a disulfide bond. Because disulfide bond formation is one of the rate-limiting steps in protein folding, this role of glycosylation may greatly facilitate the formation of the mature protein structure.

An NMR analysis of the oxidized, non-glycosylated peptide revealed that the peptide was present as two slowly exchanging conformers in a ratio of approximately 1 to 1. These conformations were attributed to the *cis* and *trans* amide isomers of the conserved proline residue at position 136. In contrast, examination of the two-dimensional NMR spectra for the oxidized, glycosylated peptide revealed that the amount of *cis* isomer dropped to 30%. Proline *cis-trans* isomerization, along with disulfide-bond formation, is considered to be one of the slow steps in the folding process. In fact, peptide prolyl isomerases exist to catalyze this potentially rate-limiting isomerization. The α-1 subunit has no such enzymatic requirement, suggesting that glycosylation is capable of counteracting the inherent tendency of this sequence to form the *cis* amide isomer, thereby ensuring correct folding of the subunit. A solution structure of the oxidized, glycosylated variant was obtained from the NMR data (Fig. 8). A similar analysis of the non-glycosylated peptide was not possible because of its low solubility and the spectroscopic ambiguity that resulted from the high content of *cis* isomer.

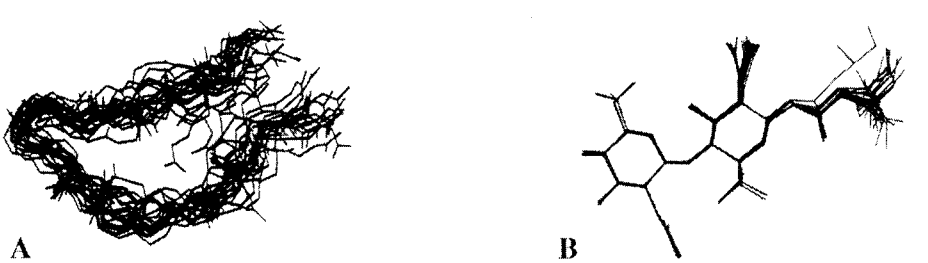



Fig. 8 Simulated annealing structures of A. the nAChR loop peptide B. the carbohydrate moiety.

While the significance of glycosylation is not fully understood, some of the structural implications of the glycan are now being elucidated. Without co-translational glycosylation, some immature proteins misfold, aggregate, and are degraded before leaving the ER.

Essentially, glycosylation appears to initiate proper folding in many glycoproteins. The conformational behavior of small peptides and glycopeptides provides insight into the structural implications of *N*-linked glycosylation and provides insight into how glycosylation may affect the structure of nascent polypeptides and therefore influence early events in protein folding.

## Conclusions

Asparagine-linked glycosylation is one of the most complex enzyme-catalyzed protein modification reactions. While remarkable progress has been made towards understanding the molecular players in the glycosylation reaction, there remain many obstacles to overcome before a clear mechanistic picture of the enzyme emerges. In the next phase of research the complementary application of state-of-the-art bioorganic and biophysical methods, together with the tools of contemporary molecular biology will contribute to a more satisfactory description of this fascinating enzyme along with its many biological implications.

## Acknowledgment

This research was supported by the National Institutes of Health (GM 39334).

## References

1. A. Varki. *Glycobiology* **3**, 97-130 (1993).
2. R. Kornfield and S. Kornfield. *Ann. Rev. Biochem.* **54**, 631-664 (1985).
3. S. Silberstein and R. Gilmore. *FASEB J.* **10**, 849-858 (1996).
4. B. Imperiali and T. L. Hendrickson. *Bioorg. Med. Chem.* **3**, 1565-1578 (1995).
5. I. B. H. Wilson, Y. Gavel and G. Von Heijne. *Biochem. J.* **275**, 529-534 (1991).
6. P. T. Englund. *Ann. Rev. Biochem.* **62**, 121-138 (1993).
7. I. Nilsson and G. von Heijne. *J. Biol. Chem.* **268**, 5798-5801 (1993).
8. J. C. Paulson. *Trends in Biochem. Sci.* **14**, 272-276 (1989).
9. A. Helenius. *Molecular Biol. Cell* **5**, 253-265 (1994).
10. R. D. Marshall. *Ann. Rev. Biochem.* **41**, 673 (1972).
11. A. Abbadi, M. Mcharfi, A. Aubry, S. Premilat, G. Boussard, M. Marraud. *J. Am. Chem. Soc.* **113**, 2729-2735 (1991).
12. E. N. Baker and R. E. Hubbard. *Prog. Biophys. Molec. Biology* **44**, 97 (1984).
13. B. Imperiali, K. L. Shannon and K. W. Rickert. *J. Am. Chem. Soc.* **114**, 7942-7944 (1992).
14. B. Imperiali, J. R. Spencer and M. D. Struthers. *J. Am. Chem. Soc.* **116**, 8424-8425 (1994).
15. B. Imperiali, K. L. Shannon, M. Unno and K. W. Rickert. *J. Am. Chem. Soc.* **114**, 7944-7945 (1992).
16. S. E. O'Connor and B. Imperiali. *Chemistry and Biology* **3**, (1996).
17. B. Imperiali and K. W. Rickert. *Proc. Natl. Acad. Sci. USA* **92**, 97-101 (1995).
18. J. J. Beintema. *Bioscience Reports* **6**, 709-714 (1986).
19. P. E. Wright, H. J. Dyson and R. A. Lerner. *Biochemistry* **27**, 7167-7175 (1988).
20. K. W. Rickert and B. Imperiali. *Chemistry and Biology* **2**, 751-759 (1995).
21. M. Mishina, T. Takamasa, K. Imoto, K. Tanaka, Y. Fujita, K. Fukuda, M. Kurasaki, H. Takahashi, Y. Morimoto, T. Hirose, S. Inayama, T. Takahashi, M. Kuno and S. Numa. *Nature* **313**, 364-369 (1985).