Structure and Dynamics of Mucin-Like Glycopeptides. Examination of Peptide Chain Expansion and Peptide–Carbohydrate Interactions by Stochastic Dynamics Simulations†

Kenneth J. Butenhof and Thomas A. Gerken*

Department of Pediatrics and Biochemistry, W. A. Bernbaum Center for Cystic Fibrosis Research, Case Western Reserve University, Cleveland, Ohio 44106

Received August 10, 1992; Revised Manuscript Received December 18, 1992

ABSTRACT: Mucins and other highly O-linked glycoproteins have been found to exist in random-coil conformations with peptide chain dimensions about 3-fold more expanded than found for deglycosylated mucins or denaturated proteins. We have examined the origin of the peptide chain expansion in mucins by stochastic dynamics simulations which include a treatment of solvation energy effects based on solvent-accessible surface area and polarizability [GB/SA; Still, C. W., et al. (1990) *J. Am. Chem. Soc. 112*, 6127]. The glycopeptides studied contained pairs of threonine residues (flanked by alanine residues) which were O-glycosylated by the di- and monooligosaccharide side chains α-NeuNAc(2–6)α-GalNAc and α-GalNAc. These glycopeptides serve as simple models for native and asialo ovine submaxillary mucin. Computer stochastic dynamic simulations show a significant decrease in end-to-end distance and radius of gyration (32% and 33%, respectively) upon complete removal of carbohydrate from the glycopeptide AAA(NeuNAc-(2–6)GalNAc)-T(NeuNAc(2–6)GalNAc)-TAAA. These changes are consistent with the extrapolations of the mucin chain dimension data to glycopeptides of this size. The simulations have identified two potentially strong peptide–carbohydrate hydrogen bonds that can influence the orientation of O-linked GalNAc. With two contiguous glycosylated sites, the lowest energy conformation obtained is characterized by a GalNAc amide proton hydrogen bond to the carbonyl of the peptide residue C-terminal to the site of glycosylation. This conformation differs from the glycopeptide conformations predicted for glycopeptides with single or widely spaced glycosylation sites. The results suggest that the experimentally determined mucin peptide chain dimensions can be fully accounted for by short-range (±3 residue) intramolecular steric and hydrogen bond interactions resulting from the clustering of glycosylated residues.

Mucin glycoproteins are a class of highly O-glycosylated high molecular weight proteins (>10^6 Da) containing 30–50% serine and threonine hydroxy-amino acids. Mucins are the primary glycoprotein component of mucous secretions and are largely responsible for its viscoelastic properties. All mucins studied to date adopt extended or stiffened random coil conformations in aqueous solutions; as a result, the chain dimensions of mucins are found to be nearly 3-fold larger than for denatured globular proteins (Shogren et al., 1986, 1987, 1989). Studies of sequentially deglycosylated ovine submaxillary mucin (OSM)[1] indicate that the carbohydrate residues are solely responsible for the expanded conformation in solution (Shogren et al., 1989; Gerken et al., 1989). Studies of sequentially deglycosylated mucins show carbohydrate chain lengths of at least two residues appear to be required to achieve near-maximal mucin peptide core expansion. The expanded conformation and very high molecular weights are thought to be the primary determinants of the unique viscoelastic properties of mucin. Many other heavily O-glycosylated proteins contain Ser- and Thr-rich domains that are heavily O-glycosylated. These domains, which are expected to possess similar extended mucin-like conformations, are commonly found in the extracellular portions of membrane-anchored proteins tethering functional domains above the cell surface (Jentoft, 1990). The molecular interactions that are responsible for the unique extended conformation of highly O-glycosylated glycoproteins are not fully understood.

In order to elucidate the forces that produce the extended conformation in mucins and other extensively O-glycosylated glycoproteins, we have initiated a series of molecular dynamic simulations. Our goals in initiating the simulations described below are (1) to determine the influence of paired glycosylation sites (dyads) on peptide and carbohydrate solution conformation, (2) to determine the dependence of peptide conformation on the carbohydrate side chain length, and (3) to identify possible carbohydrate–peptide or carbohydrate–carbohydrate interactions which contribute to the formation of an extended conformation. It is hoped that the identification of the most probable low energy structures will prove useful in interpreting subsequent NMR studies on similar O-linked glycopeptides. This work is the first to attempt to investigate the origin of the chain extension observed in mucins by molecular dynamics methods.

METHODS

Rationale. We have chosen, initially, to model a set of short O-linked glycopeptides that contain a simplified peptide sequence motif that is common to mucins and the extensively

† This work was supported by N.I.H. Grants DK39918 and HL07415, a Cystic Fibrosis Foundation training fellowship to K.J.B., and a grant from the Ohio Supercomputer Center. This is a publication from the Protein Group at Case Western Reserve University.

[1] Abbreviations: OSM, ovine submaxillary mucin; GalNAc or *G*, *N*-acetylgalactosamine; NeuNAc or *N*,*N*-acetylneuraminic acid (sialic acid); Thr or T, threonine; Ser or S, serine; Ala or A, alanine; Thr_- and Thr_+ refer to the first and second Thr residue, respectively, in a Thr–Thr pair (dyad); RIS, rotational isomeric state; σ, standard deviation; $\langle R_g^2 \rangle^{1/2}$, root mean squared radius of gyration; $\langle R^2 \rangle^{1/2}$, root mean squared end-to-end distance. Torsion (dihedral) angle definitions: θ, peptide C′–N–C_α–C′; ψ, peptide N–C_α–C′–N; χ^1, side chain N–C_α–C_β–O; χ^2, side chain C_α–C_β–O–GalNAc-C_1; χ^3, side chain C_β–O–GalNAc-C_1–GalNAc-O_5 (IUPAC ϕ^g, 1970); χ^4, side chain GalNAc-C_4–GalNAc-C_5–GalNAc-C_6–O; χ^5, side chain GalNAc-C_5–GalNAc-C_6–O–NeuNAc-C_2; χ^6, side chain GalNAc–C_5–O–NeuNAc-C_2–NeuNAc-O_6.

0006-2960/93/0432-2650$04.00/0

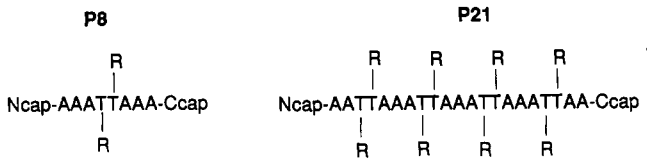

FIGURE 1: Peptide and O-linked glycopeptide structures used in simulations (A = alanine T = threonine, Ncap = -(C=O)CH_3, Ccap = -$NHCH_3$). For **P8**, R = -O-H; for **P8**-G_2, R = α-D-GalNAc-O-; for **P21**-G_8, R = α-D-GalNAc-O-; for **P8**-$(GN)_2$ R = α-D-NeuNAc(2–6)α-D-GalNAc-O-.

O-glycosylated domains of other proteins. The models chosen on the basis of the considerations discussed below are shown in Figure 1. Inspection of the available mucin peptide sequences (Hill et al., 1977; Timpte et al., 1988; Gum et al., 1989; Bhargava et al., 1990; Duperat et al., 1991; Porchet et al., 1991) reveals that significant numbers of potentially O-glycosylated hydroxy-amino acid residues occur in pairs and larger contiguous groupings. These groupings of hydroxy-amino acid residues are common and are present in higher concentrations than predicted for randomly ordered peptide residues with the same amino acid composition (unpublished calculations). Threonine was selected for both residues of the hydroxy-amino acid dyad. NMR studies indicate that the glycosidic linkages of glycosylated Thr residues and Ser are very similar, with linkages to Thr being somewhat more conformationally constrained than those to Ser (Gerken et al., 1989). Alanine was chosen because it is a major constituent amino acid in most mucins. In addition, although Ala is commonly found in helices, it does not bias peptide torsion angles toward any particular secondary structural conformation. We have chosen not to complicate the simulations by including the other amino acids that are common in mucins, such as Ser, Gly, Val, and Pro, in this initial glycopeptide conformational study.[2] The general peptide sequence motif chosen in this glycopeptide model, therefore, includes dyads of Thr flanked by Ala residues. The length of model **P8** (Figure 1) was chosen to provide long enough peptide segments above and below the two sites of glycosylation so that at maximum extension of the carbohydrate along the peptide chain in either the N- or C-terminal direction the carbohydrate would not extend beyond the peptide termini. A larger glycopeptide (model **P21**, Figure 1) has also been studied to identify the effects of potential interactions between clusters of glycosylated residues. Glycopeptide **P21** is also of interest because the central Thr dyads and nonglycosylated residues are not influenced by end-effects which modify peptide mobility and conformation at or near the peptide termini.

The O-linked carbohydrate side chains used in these simulations are the disaccharide α-NeuNAc(2–6)α-GalNAc and monosaccharide α-GalNAc present in native OSM and asialo-OSM, respectively. The α-NeuNAc(2–6)α-GalNAc side chain (*GN*, Figure 2) was chosen because this oligosaccharide is the shortest which produces the typical fully extended mucin conformation observed in mucins (Shogren et al., 1989; Gerken et al., 1989). The α-GalNAc side chain (*G*) was modeled because this carbohydrate is the most common denominator of mucin-like O-glycosidic linkages. Furthermore, mucins modified to contain only α-GalNAc monosaccharides show significantly expanded chain dimensions, although not as expanded as found for native mucin (Shogren et al., 1989; Gerken et al., 1989).

[2] We expect it will become feasible to consider the additional interactions resulting from specific amino acid side chains when nonbonding and charge interactions in the simplified models are understood and once more mucin peptide sequences and sequence–specific glycosylation information become available.

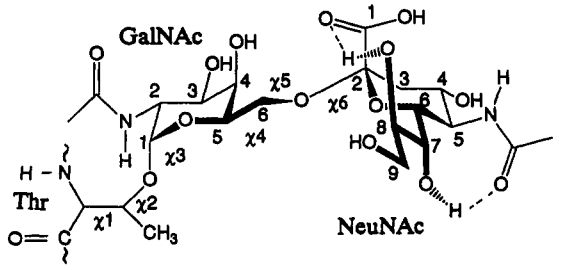

FIGURE 2: Carbohydrate side chain atomic numbering and torsion angle nomenclature. Experimentally indicated hydrogen bonds (Gerken & Dearborn, 1984; Gerken, 1986; Kooijman et al., 1990) for α-NeuNAc are shown. Side chain torsions (χ^n, n = 1–6) are defined in footnote 1.

The current modeling study has focused on fully glycosylated Thr dyads, rather than partially glycosylated dyads. The high percent of carbohydrate (Gerken & Dearborn, 1984; Gerken & Jentoft, 1987) in mucins numerically requires that dyads and larger groupings of Ser and Thr are predominantly fully glycosylated in vivo. In addition, it has been demonstrated that both threonines of a Thr dyad in synthetic peptides can be glycosylated by *N*-acetylgalactosamine transferase in vitro (Wang et al., 1992).

Construction of the Glycopeptide Models: Overview. The fully glycosylated **P8** glycopeptide structure, **P8**-$(GN)_2$, was constructed by adding energy-minimized disaccharide side chain units to the **P8** peptide positioned in an extended polyproline II conformation ($\phi = -80°, \psi = 160°$) as suggested from the RIS analysis of the chain dimensions of glycosylated OSM (Shogren et al., 1989). The **P8** models containing only GalNAc monosaccharides (**P8**-G_2) and no carbohydrate (**P8**) were constructed from energy-minimized **P8**-$(GN)_2$ by sequential removal of carbohydrate. **P21**-G_8 was constructed by joining together four dominant conformations of **P8**-G_2 obtained from molecular dynamics simulations. The carbohydrate conformations were taken from or derived from X-ray crystallographic data (Flippen et al., 1973; Darbon et al. 1984). The initial α-D-carbohydrate structures were refined by full semiempirical quantum mechanics energy minimization of the methyl or Thr glycosides using the AM1 field in GAUSSIAN 90 (Frisch, 1990) and molecular mechanics corrections to the amide bonds. The initial peptide and glycosidic torsion angles for the stochastic dynamics simulations were determined by preliminary molecular dynamics simulations, grid search, and molecular mechanics simulations with attention to the torsion angle values proposed on the basis of proton NMR experiments by Jennings and Bhattacharjee (1977) and Gerken (1986). The initial glycosidic torsions used for **P8**-$(GN)_2$ were $\chi^1 = 40°$, $\chi^2 = 150°$, $\chi^3 = 80°$, $\chi^4 = 160°$, $\chi^5 = -120°$, and $\chi^6 = 70°$ (defined in footnote 1, see also Figure 2) which corresponded to the lowest energy conformation of the disaccharide that was appropriate for peptide binding. Other higher energy carbohydrate side chain conformations were also investigated. Further details of the construction and equilibration of these models are given in the microfilm supplement.

Structure Optimization. Structure optimization was achieved using the OPLS forcefield (Jorgensen & Tirado-Rives, 1988) as implemented in MACROMODEL3.1 (Still et al., 1990). The criteria for convergence was generally an energy gradient of less than 0.001 kJ/(Å·mol) using the full matrix Newton Rapson method. Unless specified to the contrary, an "implicit" solvation treatment, the GB/SA method (Still et al., 1990, and references therein), was included in both molecular mechanics and molecular dynamics energy calculations. The advantage of implicit methods is primarily

that they are considerably faster than methods utilizing explicit solvent molecules. In the GB/SA approach, the solvent is treated as a statistical continuum where the energy of solvation is represented by solvent–solvent (G_{cav}), solute–solute van der Waals (G_{vdw}), and solute–solvent polarization (G_{pol}) energy terms. In the GB/SA method, the energy contribution from $G_{cav} + G_{vdw}$ is taken to be proportional to the solvent-accessible surface area of each atom, with each atom type assigned an empirical weighting constant. The contribution resulting from polarization of solute atoms in solution is obtained from a simplification of the generalized Born equation (Still et al., 1990).

Molecular Dynamics. Models **P8**-$(GN)_2$, **P8**-G_2, and **P8** were allowed to evolve for 140 ps or longer (data not shown) at 300 K, depending on the internal flexibility of the model, using leap frog stochastic dynamics (Van Gunsteren & Berendsen, 1988) and the OPLS potential energy field. Temperature stabilization is an intrinsic feature of the stochastic dynamics method. The stochastic dynamics algorithm attempts to incorporate the damping effects of solvent by using Langevin oscillators and frictional damping coefficients (2.5 ps^{-1} in these simulations). This approach generally allows greater exploration of torsion space at a given temperature and is less prone to produce simple oscillation about a local minimum than is true of traditional molecular dynamics methods. Electrostatic cutoffs were set at 15–20 Å during the equilibration phase with updating of the nonbonded atom list every 0.1 ps. These conditions were required because of the high degree of mobility present, especially at the equilibration stage of the stochastic dynamics runs. Nonbonding energy cutoff switching functions were not used since they are a potential source of artifacts, especially when using constant dielectric molecular dynamics forcefields such as OPLS (Loncharich & Brooks, 1989). The molecular dynamics equilibration stage typically involved several cycles of molecular dynamics simulations alternating with energy minimization in 20- or 40-ps stages. This process was repeated until a state of equilibrium was reached, characterized by the absence of systematic and usually concerted increases or decreases of torsion angles accompanied by temperature increases. The uses of a constant temperature bath (required in stochastic dynamics simulations) may obscure the temperature indicator, thus, we rely on both temperature and torsion angle monitors to evaluate the progress toward a conformational equilibrium. The equilibrated system, including structure and velocity files, was then used to initiate the final stochastic dynamic simulations. All final molecular dynamics calculations were performed with a 0.0015-ps step size using SHAKE (Ryckaert, 1985) on all bonds to hydrogens and an electrostatic cutoff at 20 Å.

The reliability of the molecular dynamics method used in this study was evaluated by preforming additional conformational studies of α-NeuNAc methyl glycoside methyl ester and of the glycopeptide Ac-T(*G*)AA-OMe for which significantly greater quality experimental structural data are available. These simulations followed the procedures outlined above.

Miscellaneous Calculations. The molecular dynamic structures were stored every 0.1 ps throughout the final molecular dynamic simulations, giving 800–1000 individual structures per simulation. These structures were used to determine mean molecular dimensions, donor–acceptor distance tables, and mean torsion angles. The mean squared end-to-end-distance, $\langle R^2 \rangle^{1/2}$, and the mean squared radius of gyration, $\langle R_g^2 \rangle^{1/2}$, were calculated using the relationships

$$\langle R_g^2 \rangle^{1/2} = \sqrt{\sum_{i=1}^{na}\sum_{j=1}^{nf} r_{icm,j}^2 [(nf)(na)]^{-1}} \quad (1)$$

$$\langle R^2 \rangle^{1/2} = \sqrt{\sum_{j=1}^{nf} r_{ete,j}^2 \, nf^{-1}} \quad (2)$$

where r_{icm}^2 is the distance from atom i to the center of mass, r_{ete} is the distance between the N- and C-terminal caps, nf is the number of structures, and na is the number of atoms in the structure.

Hydrogen bonds were identified on the basis of the distance between donor and acceptor atoms. The linearity of the angle of a putative hydrogen bond was not evaluated. To simplify the analysis, hydrogen bonds were sorted into three classes, strong, medium, and weak, with the following cutoff definitions: $S \leq 2.5$, $2.5 < M \leq 3.5$, and $3.5 < W \leq 4.0$ Å.

Histograms of selected internal coordinates (in particular torsion angles) were determined from data automatically collected in 0.0015-ps steps (approximately 150 000 sets per simulation). The mean, standard deviation, skew, and kurtosis of the internal coordinates were determined by calculation from the cartesian coordinates stored at 0.1-ps increments.

Instrumentation. All molecular dynamics and molecular mechanics energy minimizations were calculated on a VAX 785 with an Evans and Sutherland PS-390 terminal and Stereographics stereo generator for graphics display using MACROMODEL3.1 (Still et al., 1990). MOPAC 5.0 (Stewart, 1989), GAUSSIAN 90 (Frisch et al., 1990), and CHELPG (Breneman, 1990) calculations were performed at the Ohio Super Computer Center on a CRAY Y-MP.

RESULTS

The results of conformational and dynamical molecular simulations using implicit solvation are presented below for several short glycopeptides and an α-NeuNAc derivative. For these calculations, inclusion of GB/SA solvation was found to yield several benefits for both molecular mechanics optimization and molecular dynamic simulations. Equilibrated structures obtained from simulations in which GB/SA solvation energies were omitted produced more compact structures which were characterized by greater rigidity and a greater number of strong intramolecular hydrogen bonds. Hydrogen bonding interactions were somewhat less important in determining the conformation when using GB/SA implicit solvation was used, yet they still had a very significant influence on the molecular dynamic trajectories. Full molecular dynamics equilibration of **P8**, **P8**-G_2, and **P8**-$(GN)_2$ without solvation energy terms resulted in a mean contraction in peptide dimensions of ca. 15% relative to the mean structure using GB/SA solvation, with the exact value depending on the glycopeptide model. In addition, the total computation time required to equilibrate the GB/SA simulations was considerably less than that required for simulations without GB/SA solvation, despite the increased computational overhead incurred by its inclusion. This is most likely due to a reduced incidence of conformation trapping in intermediate local minima when GB/SA solvation is used. In the subsequent discussions, all simulations include GB/SA solvation unless stated to the contrary.

Overview of the Conformations of ***P8***-$(GN)_2$ *Following Equilibration.* The picture of the bis(disaccharide) side chains of **P8**-$(GN)_2$ that emerges following conformational equilibration (OPLS forcefield, GB/SA solvation) shows the disaccharide side chains loosely wrapped about the peptide

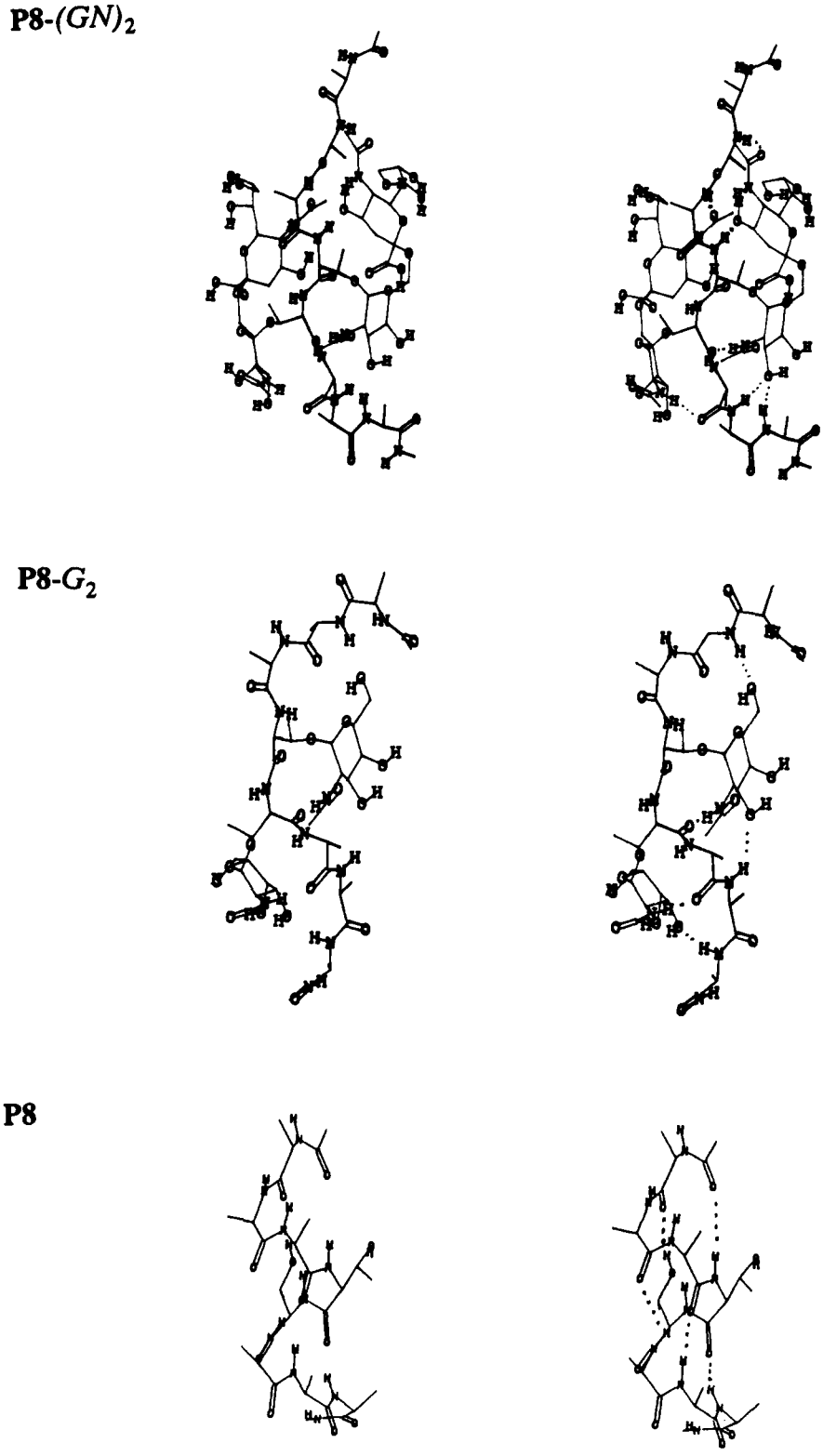

FIGURE 3: Minimized and equilibrated **P8**, **P8**-G_2, and **P8**-$(GN)_2$ conformations (stereo pairs). These conformations (along with equilibrium velocity vectors) defined the starting point of the reported stochastic dynamics simulations (GB/SA solvation and the OPLS field). Major hydrogen bonds are indicated by dotted lines. Each structure is plotted at approximately the same scale with the peptide N-termini at the top.

with the 2–6-linked disaccharide chain extending toward the N-terminus of the peptide as shown in Figure 3 (top). This conformational preference is strongly influenced by an O_{i+1}→GalNAc-HN hydrogen bond[3] for each GalNAc residue. Both hydrogen bonds are discernible in Figure 3. In the preferred conformation, the carboxylic acid group of α-NeuNAc tends to face outward relative to the peptide core. In this conformation, the GalNAc-O3 and GalNAc-O4 oxygens are fairly accessible for glycosylation. No strong peptide secondary structure is indicated and internal flexibility is high relative to a globular protein. However, the peptide ψ torsion angles are significantly expanded relative to the nonglycosylated **P8** (see below). **P8**-$(GN)_2$ average nonterminal peptide φ and ψ torsion angles are 90° and 130°, respectively. The lack of strong α-helical or β-sheet secondary structure and the degree of flexibility is consistent with the NMR and CD studies for mucins (Gerken & Dearborn, 1984; Gerken, 1986, Shogren et al., 1989). Extensive simulations starting from

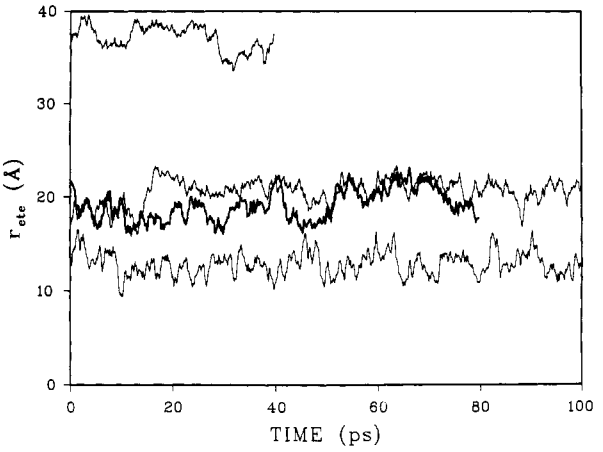

FIGURE 4: Time dependence of the end-to-end distance (r_{ete}) determined from stochastic dynamics calculations using GB/SA implicit solvation for **P8**, **P8**-G_2, **P8**-$(GN)_2$, and **P21**-$(GN)_8$. **P8**, bottom trace; **P8**-$(GN)_2$, bold trace; **P8**-G_2, center trace; and **21**-G_8, top trace.

other possible carbohydrate conformations, including the simulations discussed in the Alternate Peptide–GalNAc Conformation Section (below) have not identified any lower energy conformations of **P8**-$(GN)_2$ that are significantly different from the (static) structure of **P8**-$(GN)_2$ in Figure 3 (top).

*Overview of **P8**-G_2 and **P21**-G_8 Conformations Following Equilibration.* The equilibrium peptide and GalNAc conformations of **P8**-G_2 (Figure 3, center) and **P21**-G_8 (not shown) are similar to that found for **P8**-$(GN)_2$. In particular, a conformationally important O_{i+1}→GalNAc-HN hydrogen bond (see below) is favored over the other possible mutually exclusive hydrogen bonds involving the *N*-acetyl group of GalNAc. The O6 hydroxyl groups in **P8**-G_2 do not strongly influence its global structure.

*Carbohydrate Free Peptide (**P8**) Conformation Following Equilibration.* In **P8**, regions of nascent α-helical structure ($\phi \approx -60°$, $\psi \approx -50°$) were formed during the molecular dynamics and molecular mechanics equilibration. The structure resulting from equilibration that was used to initiate subsequent stochastic dynamic simulations is shown in Figure 3 (bottom). When GB/SA solvation is omitted, the peptide collapses further to form a more ridged α-helix.

Chain Dimension Estimates. Values of root-mean-squared radius of gyration, $\langle R_g^2 \rangle^{1/2}$, and root-mean-squared end-to-end distance, $\langle R^2 \rangle^{1/2}$, were calculated from each equilibrium molecular dynamic simulation as described in the Methods section. The time dependence of the end-to-end-distances (r_{ete}) obtained from the stochastic dynamics simulations following equilibration are shown in Figure 4. The figure shows that during the postequilibration simulation there is no trend toward further expansion or contraction for any of the simulations. However, significant variation in the end-to-end distance is present in all of the simulations. These data were used to calculate the $\langle R_g^2 \rangle^{1/2}$ and $\langle R^2 \rangle^{1/2}$ MD values in Table I. Estimates of these chain dimension parameters derived from dynamic light scattering and rotational isomeric state (RIS) calculations, EXP/RIS, are also included in Table I. These experimental/RIS data were calculated using RIS rotational matrices constructed to reproduce the experimental chain dimensions of native and sequentially deglycosylated OSM glycopeptides over a wide range of peptide lengths (Shogren et al., 1989). The glycopeptide chain dimensions obtained from the dynamic simulations as a whole indicate that the peptide cores of models **P8**-$(GN)_2$, **P8**-G_2, and **P21**-G_8 are stiffened and extended relative to the nonglycosylated peptides by steric and peptide–carbohydrate hydrogen bond interactions. The $\langle R_g^2 \rangle^{1/2}$ and $\langle R^2 \rangle^{1/2}$ values calculated from molecular dynamics data follow trends similar to those of the experimental/RIS peptide chain dimensions. The mean squared end-to-end distances, $\langle R^2 \rangle^{1/2}$, decrease upon complete deglycosylation by 32% for **P8**-$(GN)_2$ and by 37% for **P8**-G_2. The experimental/RIS chain contractions predicted from the RIS calculations (Shogren et al., 1989) are 20% for both **P8**-$(GN)_2$ and **P8**-G_2. The decrease in radius of gyration, $\langle R_g^2 \rangle^{1/2}$, due to deglycosylation (33% and 26% for **P8**-$(GN)_2$ and **P8**-G_2, respectively) also corresponds well with the RIS calculations (22% for both **P8**-$(GN)_2$ and **P8**-G_2). The contraction resulting from deglycosylation in these simulations appears to be strongly influenced by the loss of two peptide–carbohydrate hydrogen bonds: O_{i+1}→GalNAc-HN and GalNAc-O3→HN_{i+3}. These hydrogen bonding interactions and concomitant peptide–carbohydrate and peptide–peptide steric interactions stabilize ψ torsions in the range 60°–170° and destabilize ψ torsions in the range 0° to –60°, which in the deglycosylated peptide allows a nascent helical conformation to dominate. The absolute chain lengths and radius of gyrations are somewhat less than expected on the basis of the experimental/RIS data. This may be due to a remaining overemphasis of intramolecular hydrogen bonding despite the treatment of solvation, particularly for the nonglycosylated peptide.

Table I: Estimate of Chain Dimensions[a]

	$\langle R_g^2 \rangle^{1/2}$		$\langle R^2 \rangle^{1/2}$	
model	MD[b] (Å)[d]	EXP/RIS[c] (Å)[d]	MD[b] (Å)[d]	EXP/RIS[c] (Å)[d]
P8-$(GN)_2$	6.9 (4.5)	9 (1)	19.2 (3.2)	25 (2)
P8-G_2	6.2 (4.8)	9 (1)	20.7 (2.8)	25 (2)
P8	4.6 (3.2)	7 (1)	13.0 (2.6)	20 (2)
P21-$(GN)_8$	10.8 (7.2)	18 (2)	30.6 (2.2)	54 (7)

[a] $\langle R_g^2 \rangle^{1/2}$, root mean square radius of gyration; $\langle R^2 \rangle^{1/2}$ root mean square end-to-end distance. [b] Calculated from the fully equilibrated stochastic dynamics simulations originating from the low-energy conformation using OPLS and GB/SA solvation at 300 K. [c] Calculated from RIS matrices calibrated from dynamic light scattering data from native and sequentially deglycosylated OSM (Shogren et al., 1989) (see text). [d] Values are angstroms, and standard deviations, 2σ, are given in parentheses.

Torsion Angle Dynamics. The evolution of peptide and carbohydrate torsion angles as a function of time were recorded in order to monitor the conformational equilibration of all molecular dynamics simulations. Jumps between discrete conformations are observed for many torsion angles during the molecular dynamics equilibrations but were rare in the simulations after equilibration. Mean peptide torsion angles are given in Figure 5A,B as a function of residue position (numbered relative to the Thr dyad) for **P8** at all glycosylation levels and **P21**-G_8. Note the dramatic change in the peptide ψ tortion angles between glycosylated and nonglycosylated peptides while no systematic differences in the ϕ torsions are observed. The glycosylated peptides possess ψ torsion values indicating an expanded conformation, while those in the

[3] This notation describes a hydrogen bond by the donor→acceptor pair. Subscripts indicate the residue position relative to the site of glycosylation. O is the peptide carbonyl, O3, O4, and O6 are carohydrate hydroxyl oxygens, O_{Ac} is the *N*-acetyl carbonyl oxygen, and HN is either a peptide or *N*-acetyl amide proton. When explicit residues are being referred to, the amino acid designation is used as a prefix as in A6O→T5GalNAc-HN. Because hydrogen bonds often stabilize a conformational segment, this hydrogen bond nomenclature is used in the text to refer to or describe conformations.

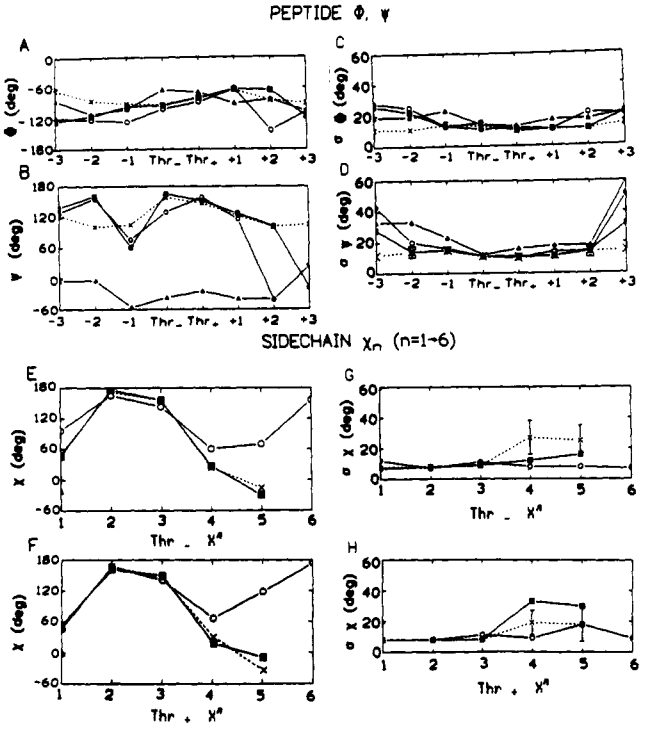

FIGURE 5: Torsion angle versus residue position obtained from stochastic dynamics simulations using GB/SA implicit solvation for **P8**-$(GN)_2$ (○), **P8**-G_2 (■), **P8** (▲), and **P21**-G_8 data (x, with dashed lines). Thr_ refers to the N-terminal Thr in a Thr dyad (Thr4 in **P8**; Thr3, -8, -13, -18 in **P21**). Thr+ refers to the C-terminal Thr in a Thr dyad (Thr5 in **P8**; Thr4, -9, -14, -19 in **P21**). (A) Peptide ϕ torsion angles; (B) peptide ψ torsion angles; (C and D) standard deviations of ϕ and ψ, respectively; (E) Thr4 χ^n torsion angle; (F) Thr5 χ^n torsion angles; (G and H) standard deviation of Thr4 χ^n and Thr5 χ^n, respectively. **P21**-G_8 data are averaged using all four threonine dyads omitting the most outlying torsion angles involving residues A1, A11, and A21.

nonglycosylated **P8** peptide lie within or near the torsional space indicative of α- or 3_{10} helices.[4] Such helical conformations are commonly observed for short peptides in water (Dyson & Wright, 1991). Figure 5C,D also shows the standard deviations from the mean of the torsional fluctuations. The values of the standard deviations indicate increased ϕ and ψ torsion mobility for those residues closest to the chain termini. The mean torsion angles and flexibility of the Thr dyads and adjacent Ala residues in **P21**-G_8 are very similar to those for **P8**-G_2 except for the higher flexibility of residues near the termini of **P8**-G_2. The nonterminal side chain standard deviations (Figure 5G,H) are also similar to the standard deviations found for the nonterminal peptide torsions (Figure 5C,D). The side chain χ^1, χ^2, and χ^3 torsion angles (Figure 5E,F) and mobilities (Figure 5G,H) for Thr_*G* (Thr4 in **P8** and Thr3, -8, -13, -18 in **P21**) and Thr+*G* (Thr5 in **P8** and Thr4, -9, -14, -19 in **P21**) are not significantly affected by sialation, with the exception of Thr4 χ^1 in **P8**-$(GN)_2$. This exception, resulting from a mutually exclusive hydrogen bonding pattern, will be discussed in the Peptide–Carbohydrate Hydrogen Bonding Interactions section (below).

A more complete picture of the range of torsion angle values explored during the simulations can be obtained from plots of peptide ϕ versus ψ for peptide angles and χ^n versus χ^{n+1} for the side chain torsion angles given in Figure 6. These plots display the shifts in torsion angle ensemble due to changes in glycosylation and record the torsional excursions that occasionally occur.

The peptide ϕ,ψ torsional space covered by **P8**-$(GN)_2$ and **P8**-G_2 are very similar to nearly identical for most residues. The corresponding ϕ,ψ distributions for nonglycosylated **P8** are broader and less symmetric about the mean. The distributions for Ala3 in Figure 6 (row A) are representative. The glycosylated peptides have peptide torsion angle distributions which are generally symmetric about the mean and are well described by the standard deviations (Figure 5C,D,G,H). The Ala torsion angles in the **P8** have a broader and less symmetric distribution than in the glycosylated peptides, which results in the larger standard deviations (Figure 5C,D). The Thr torsion angle distributions in **P8**, for example Thr4 (Figure 6, row B at right), are narrow and fairly symmetric, more like that found in the glycosylated peptides due primarily to Thr hydroxyl hydrogen bonds.

The only major difference between **P8**-G_2 and **P8**-$(GN)_2$ torsion angles occurs for residue Ala7 (Figure 5A,B; residue +2 relative to the Thr dyad). For **P8**-$(GN)_2$, ψAla 7 is near −40°, instead of the more extended value of about 120° that is typical of the other residues in both **P8**-G_2 and **P8**-$(GN)_2$. This difference is due to an alternate hydrogen bond involving Ala7 in **P8**-$(GN)_2$ which is not typically observed in glycopeptides containing closely spaced dyads. This is discussed in the Peptide–Carbohydrate Hydrogen Bonding Interactions section (below).

The symmetrical pattern of the side chain χ^n versus χ^{n+1} torsion angle distributions for **P8**-$(GN)_2$ and **P8**-G_2 are generally similar to that shown in row D, column 1, Figure 6. There are, however, two notable differences between plots of sequential side chain torsions for **P8**-$(GN)_2$, **P8**-G_2, and **P8**. Both involve terminal torsion angles (χ^5 for **P8**-G_2 and χ^2 for **P8**). The **P8**-G_2 plot of χ^4 versus χ^5 (Figure 6, row D, column 2) shows two clusters which result from two interconverting low energy conformations. The corresponding plot for **P8**-$(GN)_2$ (Figure 6D, column 1) shows a single symmetric distribution of torsion angles. The side chain torsion angles that differ between columns 1 and 2 in Figure 6D represent torsions in **P8**-G_2 where Thr4*G* O6 is hydrogen bonded to Ala2 NH. Figure 6D, column 3, shows the distribution of χ^1 versus χ^2 for **P8** in which the hydroxyl oxygen of Thr4 is undergoing nearly free rotation in the nonglycosylated peptide. The distribution of χ^2 for the glycosylated peptides, in contrast, is restricted to about 166° (Figure 5E,F) with a standard deviation of 8° (Figure 5G,H). Complete side chain plots of Thr4 χ^n versus χ^{n+1} for **P8** models are included in the supplement.

Peptide–Carbohydrate Hydrogen Bonding Interactions. In all simulations, regardless of inclusion of GB/SA solvation energies, several long-lived relatively strong peptide–carbohydrate hydrogen bonds were present which appear to have important structural implications. The significant hydrogen bonds observed during the simulation initiated from the lowest energy conformation of each model are listed in Tables II and III for the **P8** and **P21**-G_8 models, respectively. The hydrogen bond data are presented as the percent of time in which a given interaction is observed during the simulation. The most dominant hydrogen bonds observed using the OPLS forcefield are illustrated in Figure 7. These structures are substructures taken from the stochastic dynamics simulations. (The hydrogen bond types in Figure 7 will henceforth be referred to by the letter used in the figure to describe the interaction.) The type 7b hydrogen bond, O_{i+1}→GalNAc-HN, is present in all glycosylated **P8** and **P21** models. In **P21**-G_8, seven out

[4] The theoretical optimal ϕ,ψ torsion angles for an α-helix are −57° and −47°, respectively; the optimal values for a 3_{10} helix are −49° and −26°, respectively. The observed ϕ,ψ average for residues A3 to A7 in nonglycosylated **P8** are −85° and −35°, respectively.

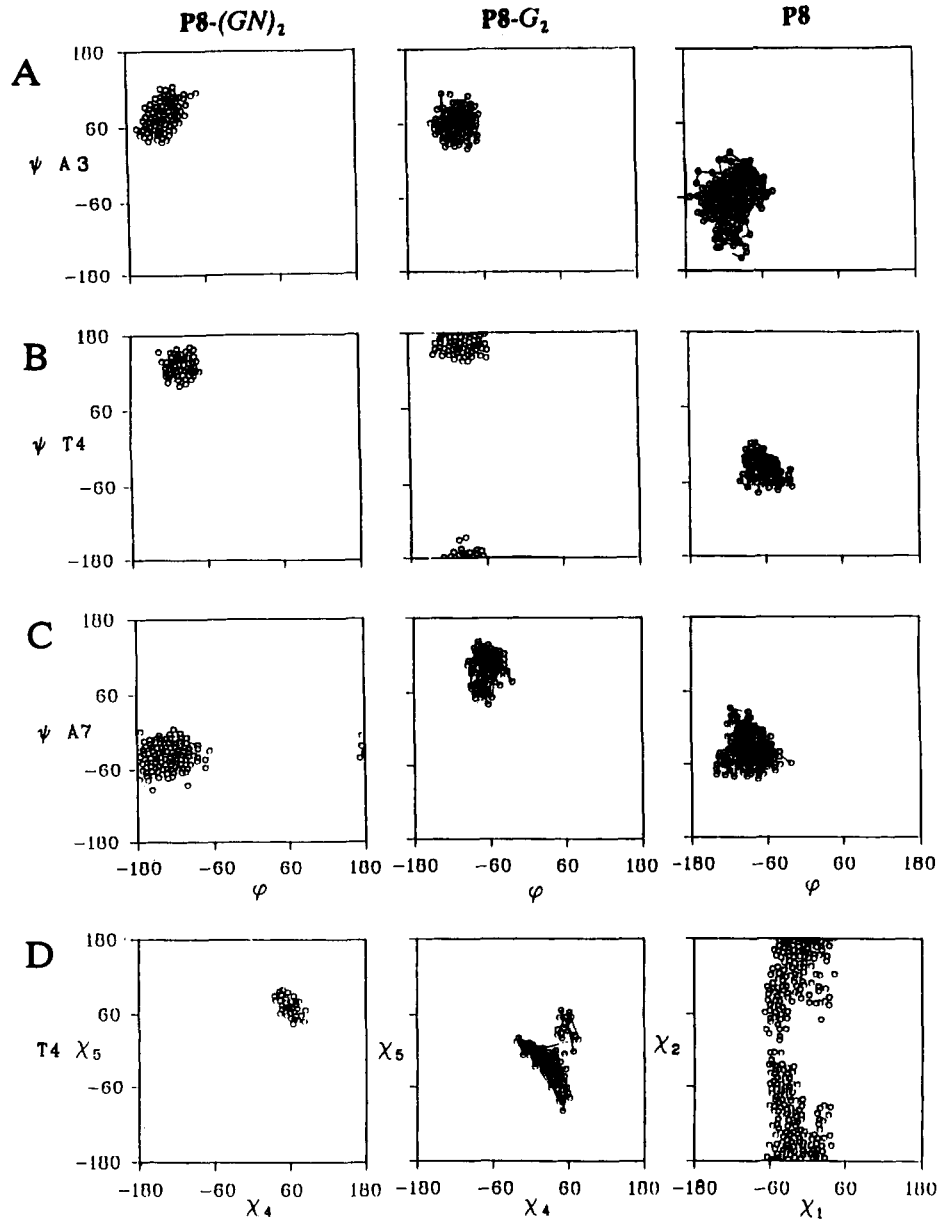

FIGURE 6: Selected peptide ϕ, ψ, and side chain χ^n, χ^{n+1} dihedral angle plots showing the range of torsion angles explored during the stochastic dynamics simulations using GB/SA implicit solvation. Row A: Ala3 ϕ vs ψ for **P8**-$(GN)_2$, **P8**-G_2, and **P8**. Row B: Thr4 ϕ vs ψ for **P8**-$(GN)_2$, **P8**-G_2, and **P8**. Row C: Ala 7 ϕ vs ψ for **P8**-$(GN)_2$, **P8**-G_2, and **P8**. Row D: Thr4 χ^4 vs χ^5 for **P8**-$(GN)_2$, χ^4 vs χ^5 for **P8**-G_2, and χ^1 vs χ^2 for **P8**. Points which are closer than 4° in both dimensions to a point already plotted have been omitted for clarity. Additional plots are included in the microfilm supplement.

of eight GalNAc residues display this interaction[5] (Table III). The type 7b hydrogen bond formed from the initial extended conformations (in which no peptide–carbohydrate hydrogen bonds were initially present) and formed repeatedly in preference to other possible mutually exclusive hydrogen bonding interactions. The effect of the type 7b interaction is to restrict Thr χ^1 torsions to mean values of 50–100° (Figure 5) for **P8**-$(GN)_2$, **P8**-G_2, and **P21**-G_8. However, the presence of χ^1 torsions in this range is not diagnostic of a type 7b interaction (refer to the Alternate Peptide–GalNAc Conformation section, below). The type 7c interaction (GalNAc-O3→HN_{i+3}) stabilizes the type 7b interaction (O_{i+1}→GalNAc-HN) in these simulations by constraining the motion of the GalNAc ring. The type 7c interaction is present as a strong hydrogen bond in **P8**-G_2 on both GalNAc side chains. In **P21**-G_8, it is found in seven out of eight GalNAc chains, and in **P8**-$(GN)_2$ it is only present for T4GalNAc. In **P8**-$(GN)_2$, the Thr5 type 7c hydrogen bond (T5GalNAc-O3→A8HN) does not occur because the required peptide amide proton (of Ala8) is oriented to hydrogen bond with T4GalNAc-O_{Ac} (a Thr4 type 7d interaction). Simulations of **P8**-G_2 have shown no Thr4 type 7d interactions (both T4GalNAc and T5GalNAc display primarily the type 7c interaction), though a Thr5 type 7d interaction is present (Table IIB). An explanation for these differences, in part, may be that the Thr_- type 7d and Thr_+ type 7c hydrogen bonds are mutually exclusive. The different hydrogen bonding pattern of the second Thr residue (Thr_+) in the Thr dyad accounts for most of the small (10%) difference in peptide chain dimensions observed in the simulations of **P8**-$(GN)_2$ and **P8**-G_2. Simulations in which the Thr4 type 7d was introduced in **P8**-G_2 and the Thr5 type 7c interaction was introduced in **P8**-$(GN)_2$ indicate that the Ala7 ϕ,ψ torsion angle required to form these interactions is equally favorable for both model **P8**-$(GN)_2$ and model **P8**-G_2, though the interconversion energy is high. However, in **P21**-

[5] T9GalNAc is prevented from possessing this hydrogen bond by a peptide turn that is stabilized by interdyad hydrogen bonding interaction between carbohydrate residues (see Table III).

Table II: Characterization of **P8** Series Donor–Acceptor Hydrogen Bond Interactions from Stochastic Dynamics Simulations[a]

(A) Peptide Donors

interaction[b]	P8 S	P8 M	P8 W	P8-G_2 S	P8-G_2 M	P8-G_2 W	P8-$(GN)_2$ S	P8-$(GN)_2$ M	P8-$(GN)_2$ W	interaction type[c]
capC=O→A2 HN	12	63	24	0	11	18	5	18	19	
napC=O→A3 HN	43	32	9	0	0	0	0	0	0	
napC=O→T4 HN	30	30	18	0	0	0	0	0	0	
ncapC=O→T4 HO	34	23	17							
A1C=O→A1 HN	0	5	17	58	40	1	43	47	10	
A1C=O→A3 HN	0	38	37	0	0	6	1	3	11	
A1C=O→T4 HO	0	3	5							
A1C=O→T5 HO	77	15	4							
A2C=O→A2 HN	0	8	52	41	59	0	65	34	1	
A2C=O→T4 HN	0	5	17	64	36	0	0	46	51	
A2C=O→A6 HN	12	27	18	0	0	0	0	0	0	
A2C=O→T5 HO	1	12	21							
A2C=O→T4*G*HO6	3	51	39							
A2C=O→T4*N*HO4							95	5	0	7f
A2C=O→T4*N*HN							0	17	50	
A3C=O→A3 HN	0	11	42	0	36	62	0	96	3	
A3C=O→T5 HN	0	51	42	0	0	2	0	25	55	
A3C=O→A6 HN	69	29	1	0	0	0	0	0	0	
A3C=O→A7 HN	75	20	3	0	0	0	0	0	0	
A3C=O→T5*N*HO4							11	14	19	7f
A3C=O→T5*N*HO7							0	17	29	
T4C=O→T4 HN	0	0	1	11	89	0	15	85	0	
T4C=O→A6 HN	3	73	24	0	4	38	0	1	10	
T4C=O→A7 HN	26	65	9	0	0	0	0	0	0	
T4C=O→A8 HN	94	6	0	0	0	0	0	0	0	
T4C=O→ccapHN	33	11	10	0	0	0	0	0	0	
T4OH→A3 HN	1	19	23							
T4OH→T4 HN	94	6	0							
T4OH→T5 HN	1	33	38							
T5C=O→T5 HN	0	0	2	6	94	0	17	83	0	
T5C=O→A7 HN	0	16	33	0	66	31	6	85	9	
T5C=O→A8 HN	2	18	23	0	0	0	0	0	12	
T5C=O→ccapHN	12	24	29	0	0	0	0	0	0	
T5C=O→T4*G*HN				100	0	0	99	1	0	7b
T5C=O→T5*G*HN				0	12	62	0	31	58	
T5OH→A3 HN	6	33	18							
T5OH→T4 HN	0	18	53							
T5OH→T5 HN	99	1	0							
T5OH→A8 HN	1	58	34							
A6C=O→A6 HN	0	0	20	0	87	13	0	54	45	
A6C=O→A8 HN	0	18	38	24	75	1	0	0	3	
A6C=O→ccapHN	9	22	15	0	2	8	0	2	2	
A6C=O→T5*G*HN				100	0	0	72	19	5	7b
A6C=O→T5*G*HO3			0	0	1	6	32	18		
A7C=O→A7 HN	12	28	18	0	19	78	0	12	81	
A7C=O→ccapHN	0	0	0	0	13	29	11	44	27	
A8C=O→A8 HN	2	22	46	0	6	55	1	27	39	
A8C=O→T4*G*HO3				0	0	0	6	30	16	
A8C=O→T4*G*HO4				0	0	0	5	16	14	

(B) GalNAc Donors

interaction[b]	P8-G_2 S	P8-G_2 M	P8-G_2 W	P8-$(GN)_2$ S	P8-$(GN)_2$ M	P8-$(GN)_2$ W	interaction type[c]
T4*G* O_{Ac}→A8 HN	0	0	0	0	7	18	
T4*G*O3→A7 HN	71	19	4	97	3	0	7c
T4*G*O3→A8 HN	0	0	0	91	9	0	7d
T4*G*O3→ccapHN	0	0	0	28	32	14	
T4*G*O3→T4*G*HN	1	99	0	0	99	1	
T4*G*O3→T4*G*HO4	1	98	1	2	97	1	
T4*G*O4→A7 HN	0	0	13	0	0	38	
T4*G*O4→A8 HN	0	0	0	0	17	56	
T4*G*O4→ccapHN	0	0	0	3	27	14	
T4*G*O4→T4*G*HO6	0	28	59				
T4*G*O6→A2 HN	99	1	0				7a
T4*G*O6→T4*G*HO4	0	30	62				
T5*G*O3→A8 HN	92	7	1	0	0	0	7c
T5*G*O3→ccap HN	62	31	52	0	0	0	7d
T5*G*O3→T5*G*HO4	1	99	0	77	22	1	
T5*G*O3→T5*G*HN	2	98	0	3	92	4	

Table II: (Continued)

interaction[b]	(B) GalNAc Donors P8-G_2 S	M	W	P8-$(GN)_2$ S	M	W	interaction type[c]
T5*G*O4→ccapHN	0	4	28	0	0	0	
T5*G*O4→T5*G*HO3	100	0	0	23	74	3	
T5*G*O4→T5*G*HO6	0	23	32				
T5*G*O6→A3 HN	0	0	0				7a
T5*G*O6→T5*G*HO4	15	20	41				

interaction[b]	(C) NeuNAc Donors P8-$(GN)_2$ S	M	W	interaction type[c]
T4*N*O4→A2 HN	0	1	17	
T4*N*O4→A3 HN	0	0	13	
T4*N*O4→T4 HN	99	1	0	7g
T4NO_{Ac}→A2 HN	55	24	9	7e
T4*N*COO→T4*N*HO7	0	72	27	
T5*N*O4→T5 HN	0	1	8	7g
T5NO_{Ac}→A3 HN	98	2	0	7e
T5*N*COO→T5*N*HO7	0	63	35	

[a] Values given represent the duration of the interaction as a percent of simulation time. Strong (S) medium (M) and weak (W) hydrogen bond interactions are based on the following distance constraints: S < 2.5 Å ≤ M < 3.5 Å ≤ W < 4.5 Å. [b] Abbreviations: ncap, -(C=O)CH_3; ccap, -$NHCH_3$; *G*, GalNAc; *N*, NeuNAc; A, Ala; T, Thr; O_{Ac}, C=O of acetyl group; COO, C=O of carboxylate group. [c] Interactions are as listed in Figure 7.

Table III: Characterization of **P21**-G_8 Donor–Acceptor Hydrogen Bond Interactions from Stochastic Dynamic Simulations

selected donors[a] interaction[b]	S	M	W	type[c]
A4C=O→T3*G*HN	100	0	0	7b
A4C=O→T4*G*HN	0	22	52	
A5C=O→T4*G*HN	100	0	0	7b
T9C=O→T8*G*HN	100	0	0	7b
A12C=O→T9*G*HN	98	2	0	
T14C=O→T13*G*HN	100	0	0	7b
A15C=O→T14*G*HN	100	0	0	7b
A16C=O→T18*G*HN	16	27	18	
T3*G*O3→A6 HN	56	22	7	7c
T4*G*O_{Ac}→T9*G*HO6	14	68	15	
T4*G*O3→A7 HN	41	47	9	7c
T8*G*O3→A11 HN	96	4	0	7c
T9*G*O3→T14 HN	84	15	1	
T9*G*O4→T14*G*HO6	2	82	16	
T9*G*O6→T4*G*HO3	0	62	36	
T13*G*O3→A16 HN	88	12	0	7c
T14*G*O3→A17 HN	99	1	0	7c
T14*G*O6→T9*G*HO3	0	50	48	
T14*G*O6→T9*G*O4H	99	1	0	

[a] See footnote *a* of Table II. Only interactions four residues or more from the peptide termini are given. [b] Abbreviations are as given in footnote *b* of Table II. [c] Interactions are as listed in Figure 7.

G_8, the Thr_ type 7d hydrogen bond (GalNAc-O3→HN_{i+4}) is considerably less favorable and is not observed due to destabilizing steric interactions between carbohydrate residues of adjacent dyads (for example, between Thr+*G* and Thr_*G* of the next dyad). This result suggests that type 7d hydrogen bonds may not be important for glycopeptides in which the separation between glycosylation sites is small (less than four residues) as is the case for most mucin sequences.

The peptide–NeuNAc interactions in **P8**-$(GN)_2$ for Thr4 are different from those for Thr5 (Table II). For the NeuNAc of Thr4 (Thr4*G*N), there are three strong hydrogen bonds involving α-NeuNAc and peptide (types 7e, 7f, and 7g) while the NeuNAc C2–C1 bond points toward the C-terminus of the peptide. In contrast, the α-NeuNAc at Thr5 (Thr5*G*N) is more flexible, with fewer H-bonds, and its C2–C1 bond points away from the peptide chain. Although Thr5*G*N participates in several strong hydrogen bonds (Table IIC), they persist for a shorter duration than in Thr4*G*N. The reduced stability of the T5*G*N H-bonds is due to steric interaction with peptide induced by the T4*G*N type 7f and 7g interactions. High temperature (500 K) simulations followed by simulations at 300 K suggest that either disaccharide residue (T4*G*N or T5*G*N) may adopt a conformation similar to that shown for T4*G*N in model **P8**-$(GN)_2$, but it is energetically unfavorable for both residues to adopt this conformation at the same time. It is likely the NeuNAc ring's would be found in both orientations in solution.

Alternate Peptide–GalNAc Conformations. The interactions that most strongly affect the orientation of the carbohydrate with respect to the peptide core are interactions involving peptide and the *N*-acetyl group of GalNAc. The peptide–carbohydrate conformations resulting from the full stochastic dynamic equilibration of **P8**-$(GN)_2$ and **P8**-G_2 (Figure 3) are characterized in part by the presence of O_{i+1}→GalNAc-HN (type 7b) hydrogen bonds which persist for more than 200 ps of stochastic dynamics (including equilibration time). Other less stable peptide–GalNAc conformations and peptide–GalNAc hydrogen bonds were observed in the initial full equilibrations and from equilibrations of altered initial conformations. For **P8**-GN_2, the most stable alternate hydrogen bonding interactions between peptide and the *N*-acetyl group of GalNAc (i.e., excluding interaction 7b) are shown in Figure 8. The interactions that characterize the conformations shown in the figure are O_i→GalNAc-HN (Figure 8A), O_{i-2}→GalNAc-HN (Figure 8B), and GalNAc-O_{Ac}→HN_i (Figure 8C) in order of decreasing stability.

The stability of **P8**-$(GN)_2$ conformations containing the alternate peptide–GalNAc conformations shown in Figure 8 was examined by molecular mechanics and stochastic dynamics simulations. Interconversions from one hydrogen bonding peptide–GalNAc conformation to another were monitored by recording the key electron donor and acceptor pair distances throughout the simulations. Figure 9 shows distances as a function of simulation time for several important peptide to GalNAc *N*-acetyl group interactions for simulations starting with two of the initial conformations. Figure 9A shows the interaction traces for simulations on **P8**-$(GN)_2$ initiated with the GalNAc-O_{Ac}→HN_i interaction (shown in Figure 8C). A rapid conformational change involving Thr4*G* allows formation of the A2O→T4GalNAc-HN interaction (Figure 8B) while Thr5*G* rapidly converts to the T5O→T5GalNAc-HN interaction (Figure 8A). Figure 9B shows the interaction trace of

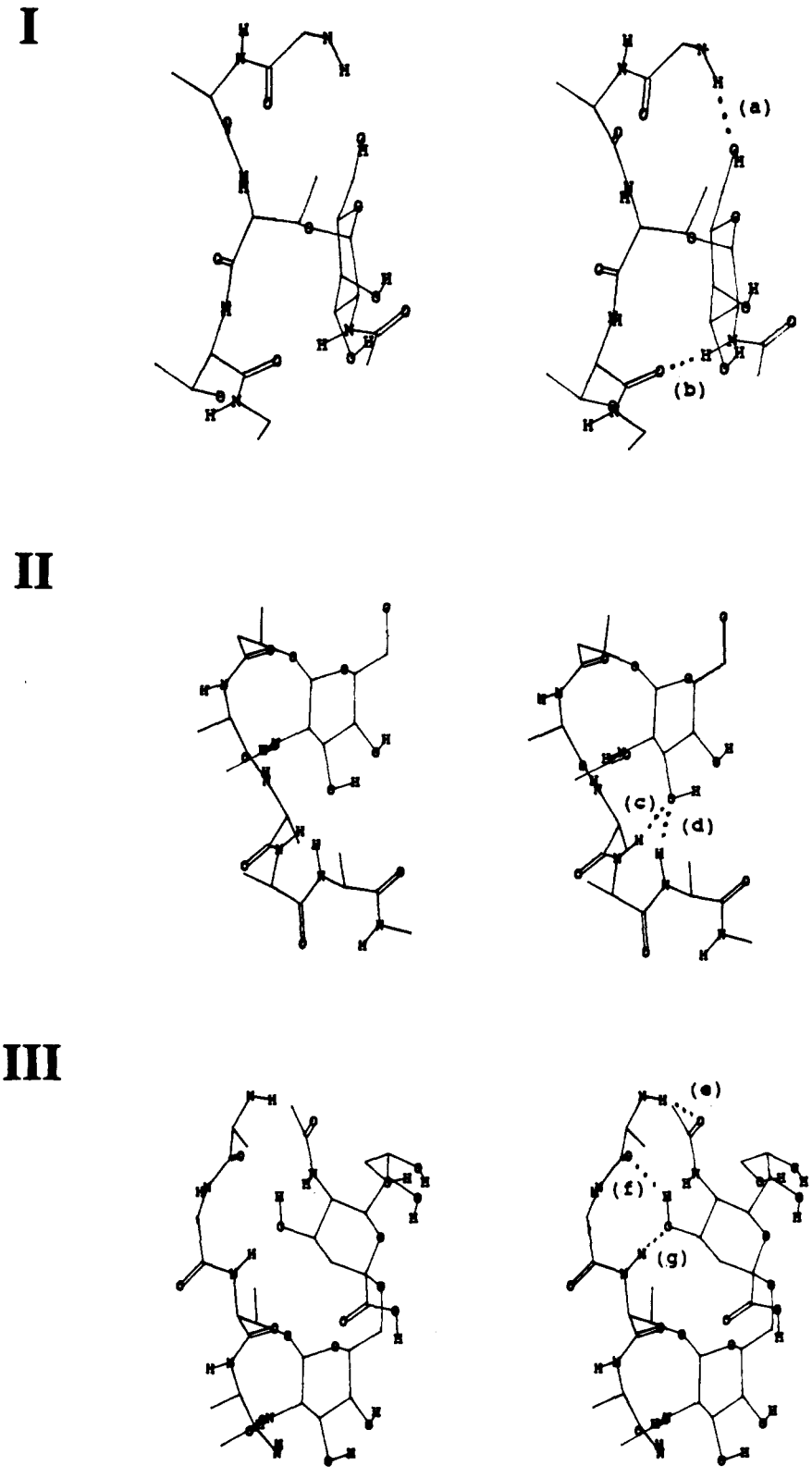

FIGURE 7: Primary peptide–carbohydrate hydrogen bonding interactions observed from the stochastic dynamics simulations using GB/SA implicit solvation. Stereo pairs shown are substructures of **P8**-$(GN)_2$ and **P8**-G_2 (peptide N-termini at top). Panel I: (a) GalNAc-O6→HN_{i-2}, (b) O_{i+1}→GalNAc-HN. Panel II: (c) GalNAc-O3→HN_{i+3}, (d) GalNAc-O3→HN_{i+4}; Panel III: (e) NeuNAc-O_{Ac}→HN_{i-2}, (f) O_{i-2}→NeuNAc-HO4, (g) NeuNAc-O4→HN_i.

the second simulation of **P8**-$(GN)_2$ originating with the T5O→T5GalNAc-HN interaction shown in Figure 8A, in which a rapid series of periodic excursions occurs between two primary conformations (Figure 8A,B). In Figure 9B, the O_i→GalNAc-HN hydrogen bond (Figure 8A) is dominant for both Thr4*G* and Thr5*G*. The lowest energy conformation for the peptide–GalNAc linkage, O_{i+1}→GalNAc-HN (the type 7b interaction in Figure 7, panel Ib) was not achieved in either of the 40-ps stochastic dynamic simulations. This is not surprising given the short simulation time, as compared to the original full equilibration process. In addition, the full equilibration process involved a combination of stochastic dynamics, molecular mechanics, and grid search of key side chain torsion angles. The total difference in energy between the **P8**-G_2 type 7b conformer obtained by full energy equilibration and the next lowest energy conformer identified

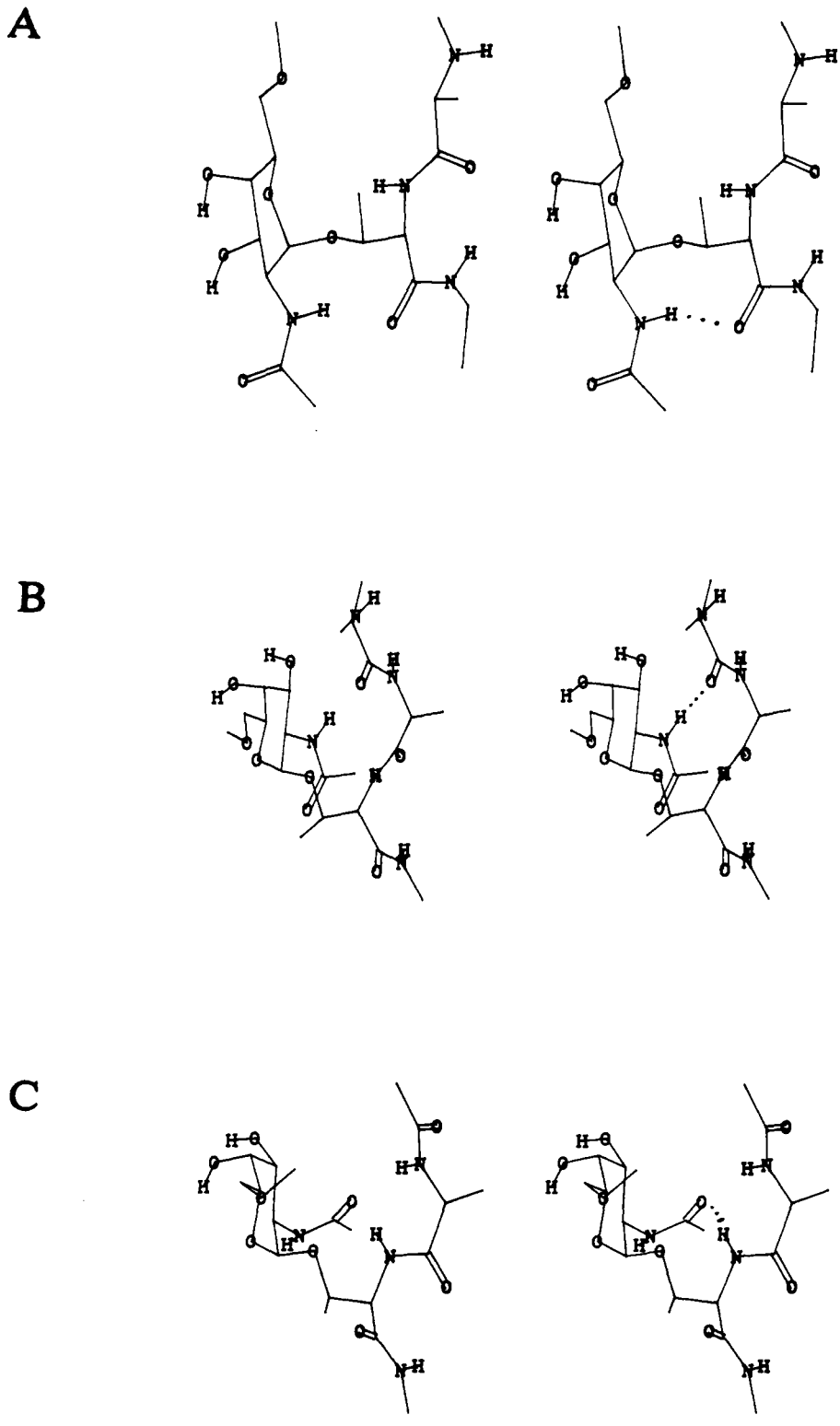

FIGURE 8: Alternate peptide–GalNAc hydrogen bonding patterns for **P8**-$(GN)_2$. Clipped stereo pairs were selected from altered initial conformation simulations as described in the text, and are shown with peptide N-termini at top. (A) $O_i \rightarrow$GalNAc-HN interaction; (B) $O_{i-2} \rightarrow$GalNAc-HN interaction; and (C) GalNAc-$O_{Ac} \rightarrow HN_i$ interaction. For comparison, the lowest energy conformation for the Thr-GalNAc linkage of **P8**-$(GN)_2$ is shown in Figure 7, panel I.

(shown in Figure 8A) after 2000 iterations of energy minimization) is ~200 kJ/mol.[6] In addition to information concerning relative stability, the simulations traced in Figure 9 demonstrate the ability of the GB/SA stochastic dynamics method to allow interconversion of locally stable conformers each of which have significant hydrogen bond interaction energies.

Test of the Molecular Dynamics Field for Glycoproteins and Carbohydrates. Additional conformational tests on molecules closely related to the structures shown in Figure 1 have been conducted in order to evaluate the stochastic dynamics method used in this work. For these molecules, methyl ester methyl glycoside of NeuNAc and Ac-T(*G*)AA-OMe, high quality X-ray and NMR structural information is available that is not yet available for the **P8** and **P21** model compounds.

Elements of the structure of Ac-T(*G*)AA-OMe, including the hydrogen bonding pattern, have been inferred from ^1H

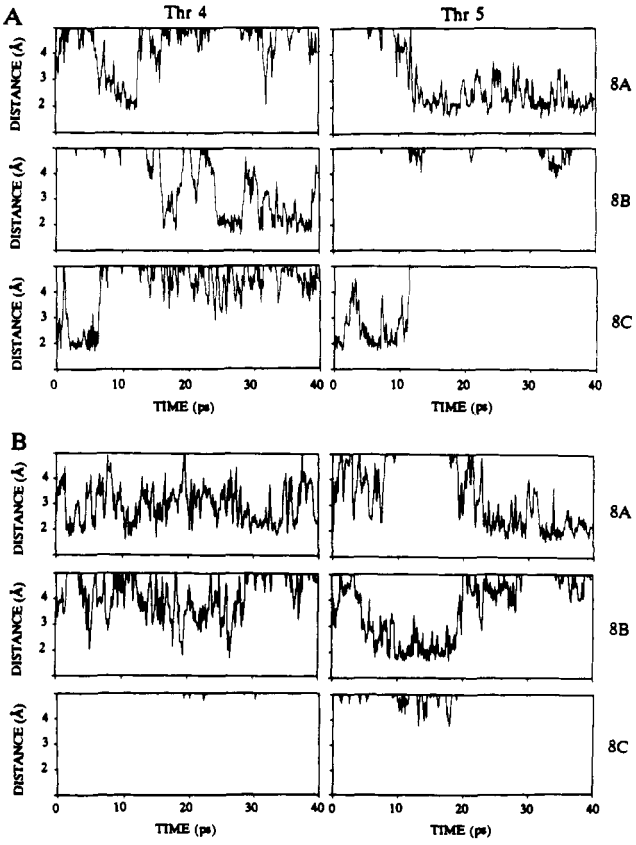

FIGURE 9: Plots showing the formation and dissociation of selected GalNAc–peptide hydrogen bonds in nonequilibrated simulations of **P8**-$(GN)_2$. Interatomic distances for selected possible hydrogen bonds are shown versus simulation time at 500 K for Thr4*G* (left column) and Thr5*G* (right column). Two different simulations (A and B) with different initial GalNAc–peptide conformations (see text) are shown. (A) Simulation originating with a strong GalNAc-O_{Ac}→HN_i hydrogen bond (Figure 8C) showing interatomic distance versus time for the following interactions: O_i→GalNAc-HN (8A), O_{i-2}→GalNAc-HN (8B), GalNAc-O_{Ac}→HN_i (8C). (B) Simulation originating with a strong O_i→GalNAc-HN hydrogen bond (Figure 8A) showing interatomic distances as described above. Interaction distances of less than 2.5 Å indicate a strong hydrogen bond.

NMR in DMSO and water (Mimura et al., 1989). The NMR results of Mimura et al. (1989) imply that for this model the fundamental peptide–GalNAc *N*-acetyl hydrogen bond interaction (in H_2O) is T1O→T1GalNAc-HN (Figure 8A) and not interaction 7b. Our simulations of Ac-T(*G*)AA-OMe using the OPLS forcefield agree. The energy difference between the T1O→T1GalNAc-HN (Figure 8A) conformation and the A2O→T1GalNAc-HN (Figure 7b) conformation (after energy minimization following the procedure used for model **P8**) is about 2.5 kJ/mol in favor of T1O→T1GalNAc-NH (Figure 8A). The lowest energy peptide–GalNAc conformation obtained by molecular mechanics methods (Figure 8A) can be simply converted to the conformation shown in Figure 8C with negligible shifts in Thr1 peptide ϕ or Thr1 side chain χ^1 torsion angles, as proposed by Mimura et al. (1989, and references therein). This conformation is not stable in the simulations and was not observed experimentally by Maeji et al. (1987) unless the N-terminal cap is significantly larger than an acetyl group (for example, (benzyloxy)carbonyl). Thus, the OPLS field appears to successfully predict the experimentally observed conformation and hydrogen bond pattern for this glycopeptide. From our simulations, we conclude that the preferred conformation of the single GalNAc side chain in Ac-T(*G*)AA-OMe (Figure 8A) is different from the GalNAc side chain conformation in the Thr dyad containing **P8** and **P21** glycopeptide models (Figure 7, panel I).

The atomic coordinates of an α-NeuNAc derivative (the methyl ester of the α-NeuNAc methyl glycoside) are known from its crystal structure (Kooijman et al., 1990). Molecular mechanics calculations were performed using the OPLS field to evaluate the use of this field for simulations involving carbohydrates. The energy of the OPLS energy-minimized crystal structure was −98.24 kJ/mol with a final energy gradient of 0.0006 kJ/(Å·mol), indicating excellent convergence to an energy minima. The only significant difference between the crystal structure and the energy-minimized structure was the conformation about C9–O9. This difference arises from intermolecular hydrogen bonds in the crystal. Omitting C9 and associated atoms, the rms deviation between the OPLS optimized structure and the crystal structure is satisfactory (0.21 Å). The crystal structure's intramolecular hydrogen bonding pattern is preserved in this energy-minimized structure and most torsion angles lie within 3° of the crystal structure.[7] However, during the molecular dynamic simulation of α-NeuNAc the C7, C8, and C9 side chain conformation changes from that in the crystal. This altered conformation, which contains hydrogen bonds between the vicinal hydroxyls of C7, C8, and C9, appears to be the global minimum using the OPLS field, E = −169.9 kJ/mol (final rms deviation = 0.0005 kJ/(Å·mol)). It is significantly more stable (by 72 kJ/mol) than the energy-minimized conformation derived from the crystal structure. For comparison, the same energy minimizations were performed using the Macromodel implementations of the (unified) AMBER (Weiner et al., 1984) and all-atom MM2 forcefields (Allinger et al., 1977), both of which use explicit hydrogen bond potentials, unlike OPLS. The AMBER field very slightly favored the α-NeuNAc crystal structure's side chain conformation (by 0.9 kJ/(Å·mol)) and MM2 clearly favored the α-NeuNAc crystal structure's side chain conformation (by 20.3 kJ/(Å·mol)). The effect on the current simulations of the OPLS forcefield's apparent bias for hydrogen bond formation between exocyclic vicinal hydroxyl groups is that the *N*-acetyl group of NeuNAc is more free to interact with other charged groups than would otherwise be the case. In addition, potential interactions of the NeuNAc C7 and C8 hydroxyls with other hydrogen bonding groups are not explored using OPLS because they are so strongly hydrogen bonded within the NeuNAc C7–C9 chain. The ramifications of this finding on the OPLS simulations were explored more quantitatively by generation of a stochastic dynamics simulation of **P8**-$(GN)_2$ using AMBER with GB/SA solvation. The starting conformation was that given in Figure 3, except that the NeuNAc hydrogen bonding conformation was that of the crystal structure (Figure 2). After 10 ps of simulation in which the NeuNAc hydrogen bonds were constrained (to allow peptide equilibration), 60 ps of unrestrained stochastic dynamics were recorded. In the simulation the NeuNAc-O_{Ac}→NeuNAc-HO7 hydrogen bond was predominantly replaced in both residues with a type 7e (NeuNAc-O_{Ac}→HN_{i-2}) hydrogen bond between peptide and carbohydrate. This preference for a type 7e carbohydrate–

[6] The conformation shown in Figure 8A was energy minimized for 2000 iterations. The energy at the energy minima of **P8**-G_2 in this hydrogen bond conformation was −2292 kJ/mol. The energy for the **P8**-G_2 lowest energy conformation (Figure 3) after 2000 iterations of energy minimization was −2479 kJ/mol. The energy difference for this conformational change expressed for a single GalNAc is therefore about 93.5kJ/mol.

[7] Torsion angle comparisons of particular interest are (crystal vs simulation) the following: C–O–C2–C1 = −179° vs 179°, C–O–C2–O6 = 61° vs 63°, and C–C2–O6–C6 = 166 vs 168°.

peptide interaction over a NeuNAc-O_{Ac}→NeuNAc-HO7 internal interaction was also found in the OPLS simulations of **P8**-$(GN)_2$. The NeuNAc-O1→NeuNAc-HO8 interaction known to be present experimentally (Gerken & Dearborn, 1984) was detected in the AMBER GB/SA simulations for **P8**-$(GN)_2$ although weak and transient (3% strong, 15% medium). Finally, the NeuNAc C7 and C9 hydroxyl groups show an increase in hydrogen bond formation with peptide at the expense of the C4 hydroxyl group.[8] There was no significant net increase or decrease in NeuNAc hydroxyl–peptide hydrogen bonding or carbohydrate–peptide steric interactions compared with the OPLS field results. On this basis, we believe that the preference for hydrogen bond formation between vicinal hydroxyl in the OPLS field has a negligible effect on the conformation and mobility of NeuNAc in these glycopeptide models and no effect on overall peptide chain dimensions.

DISCUSSION

These studies were begun to identify the interactions responsible for the chain expansion of multiply O-glycosylated peptides. In the simulations reported, a simple glycopeptide model containing clustered glycosylation sites was examined. The high occurrence of clustered glycosylation sites that is evident from the amino acid sequences of mucins and other O-linked glycopeptides suggests a possible conformational role for this motif.

In this initial study of multiply glycosylated O-linked glycopeptides, we used stochastic dynamics, the OPLS forcefield, and implicit solvation (GB/SA) to increase torsional flexibility and reduce surface hydrogen bond formation. The results of this study, with the probable exception of the conformation of the α-NeuNAc (sialic acid) C7, C8, and C9 hydroxyl groups, are in good agreement with the available data from crystal structures, NMR data, and chain dimensions derived from light scattering studies. The most significant deviations of these simulations from the experiment are attributable to a remaining general overemphasis of hydrogen bond energy in these highly solvent-exposed polar molecules. This deviation is largest for the nonglycosylated model **P8**, possibly due to the decreased significance of van der Waals interactions and thus an increased importance of electrostatic factors in determining the allowed conformers relative to the glycosylated peptides. The overemphasis of hydrogen bond strength is much more severe when solvation is not taken into account. Simulations of O-linked glycopeptide conformation using more sophisticated custom partial charge and forcefield parameters, including the carbohydrate field parameters of Ha et al. (1988) and Homans (1990), are in progress.

The stochastic dynamic simulations, using OPLS and GB/SA solvation, clearly indicate that the peptide models undergo a significant expansion in chain dimension due to glycosylation. This chain expansion is necessarily the result of changes in the intramolecular steric and hydrogen bonding interactions in the immediate vicinity of the glycosylation sites. More remote (±>3 residue) intramolecular or intermolecular interactions are not required to reproduce the experimentally indicated conformations. The increase in chain expansion due to removal of carbohydrate determined from the OPLS GB/SA simulations is ~30% (relative to the mean dimension of nonglycosylated **P8**). The mean peptide chain dimensions of the bis(disaccharide) model, **P8**-$(GN)_2$, is very similar to the chain dimension of the **bis**(monosaccharide) glycopeptide model, **P8**-G_2. These values are in excellent agreement with the values extrapolated from OSM and asialated-OSM glycopeptides of equal peptide residue length, given the presence of sequence heterogeneity in the fractionated OSM fragments used for the chain dimension determination (Shogren et al., 1990).

Mimura et al. (1989) have studied a variety of short synthetic GalNAc-containing glycopeptides. They have proposed the existence of two mutually exclusive hydrogen bonds between GalNAc and peptide, shown in Figure 8A,C (O_i→GalNAc-HN and GalNAc-O_{Ac}→HN_i). The ratio of the occurrence of these "mutually exclusive" hydrogen bonds was found to vary depending on the bulk of the N-terminal capping residue. With an acetyl group, only the GalNAc-O_{Ac}→HN_i interaction (Figure 8A) is observed. Our simulations of Ac-T(*G*)AA-OMe corroborate that the conformer shown in Figure 8A is the minimum energy structure. Interestingly, the GalNAc-O_{Ac}→HN_i conformer (Figure 8A) is not the preferred conformer in simulations of the multiply glycosylated peptides. In all models containing the threonine dyad, the O_{i+1}→GalNAc-HN conformer (7b) is strongly preferred over the conformers shown in Figure 8A,C (by about 40–90 kJ/mol of GalNAc). For Ac-T(*G*)AA-OMe, however, the 8A conformation is more stable than the type 7b conformer (by about 2 kJ/mol). Thus, the preferred hydrogen bonding conformation of GalNAc may depend on the glycosylation site density of the peptide.

Carbon-13 NMR T_1 relaxation times at 67.5 MHz (Gerken & Dearborn, 1984) indicate substantial flexibility for NeuNAc due to rotation about the C5–C6 bond of GalNAc (χ^4). In contrast, our simulations show the NeuNAc–GalNAc glycosidic linkage to be rotationally restricted as a result of hydrogen bonding interactions of NeuNAc with peptide (the standard deviation of torsion angles is ~10°). However, since ^{13}C relaxation is sensitive to substantially lower frequency motions than are accessible to molecular dynamics simulations, the carbon-13 NMR T_1 data do not necessarily conflict with our molecular dynamic results. In fact, carbon-13 NOE, T_1, and T_2 data collected at 100 MHz, 67.5 MHz, and 45 MHz (K. J. Butenhof and T. A. Gerken, unpublished results) places the NeuNAc ring carbons' correlation time at about 300 ps, which is not inconsistent with our simulation results. The primary interactions responsible for the restricted rotation of NeuNAc in **P8**-$(GN)_2$ are hydrogen bonds between the *N*-acetyl oxygen at C5 and the hydroxy group at C4 of NeuNAc with peptide (type 7e, 7f, and 7g hydrogen bonds). Due to their large size and apparent sequence heterogeneity, these interactions are difficult to identify in native mucins by NMR.

The internal mobility of the nonglycosylated peptide, **P8**, is found to be significantly greater than that of the glycosylated peptides due primarily to larger fluctuations of the ψ torsion angles (average standard deviation of 20° versus 14° for the glycosylated peptides, excluding C-terminal residue ψ values). The flexibility of terminal residues is generally higher for all models due to simple end-effects. When end-effects are eliminated, as in the central residues of **P21**-G_8, it is apparent that there is no difference in the torsion angle flexibility between glycosylated and nonglycosylated peptide residues in the glycosylated models studied. Thus, the stiffening effect that results from glycosylation is uniformly delocalized when the glycosylated dyads are separated by at least three (nonglycosylated) residues. This finding is consistent with estimates based on RIS analysis (Shogren et al., 1989; Tanpipat

[8] The greatest differences in the AMBER versus the OPLS strong hydrogen bond network resulting from simulations of **P8**-$(GN)_2$ are the following: A2O→NeuNAc-HO4 (15% vs 95%), A2O→T5-NeuNAc-HO7 (20% vs 0%), A3O→T5-NeuNAc-HO4 (2% vs 11%), A3O→T5-NeuNAc-HO9 (10% vs 0%).

& Mattice, 1990) in which it was estimated that conformational changes spanning approximately seven residues for each glycosylation site are required to produce the experimental chain extension in high molecular weight mucins.

The proposed conformation of the carbohydrate in these O-linked mucin-like peptides and glycopeptides suggest the interesting possibility that the accessibility of Thr_+ (the more C-terminal threonine of the dyad) in a nonglycosylated threonine cluster will be somewhat greater than Thr_- (the more C-terminal threonine residue in the dyad). Furthermore, on the basis of the structures of mono- and bis(glycosylated) derivatives, the accessibility of Thr_- for glycosylation appears to be severely reduced by the glycosylation of Thr_+ while the accessibility of Thr_+ is considerably less affected by glycosylation of Thr_-. Indeed, this trend is at least partially corroborated by the recent studies of Wang et al. (1992) using *N*-acetylgalactosaminyltransferase to glycosylate peptides containing Thr dyads. Also of interest is that in our simulations the GalNac-O3 site appears to be as accessible to glycosylation in the bisGalNac glycopeptide as it is in the bisNeuNAc(2–6)GalNAc glycopeptide, suggesting that the termination of carbohydrate chain growth caused by (2–6) sialation is not simply due to steric blocking of the GalNAc C3 hydroxyl group.

CONCLUSIONS

Molecular dynamics simulations have identified two strong hydrogen bonds between peptide and GalNAc (O_{i+1}→GalNAc-HN and GalNAc-O3→HN_{i+3}) which appear to direct the orientation of O-linked GalNAc in glycopeptides containing glycosylated Thr dyads. This orientation differs from that predicted for peptides with widely spaced glycosylation sites. Using implicit solvation and stochastic dynamics simulations, it has been possible to satisfactorily account for the expansion of the peptide chain dimensions due to O-glycosylation. These findings imply that short-range (±3 residue) intramolecular steric and hydrogen bond interactions involving glycosylated dyads can sufficiently account for the expanded conformation of heavily O-glycosylated mucin-type glycoproteins.

ADDED IN PROOF

Our calculations are further supported by the recent report of Mimura et al. (1992), indicating that the GalNAc-O_{Ac}→HN hydrogen bond proposed earlier (Mimura et al., 1989) is instead the O_f→GalNAcNH interaction.

ACKNOWLEDGMENT

We acknowledge Randal Shogren's assistance in the rotational isomeric state extrapolations of $\langle R_g^2\rangle^{1/2}$ and $\langle R^2\rangle^{1/2}$ values.

SUPPLEMENTARY MATERIAL AVAILABLE

Details of the construction and determination of initial conformation of NeuNAc(2–6)GalNAc, **P8**-$(GN)_2$, **P21**-$(GN)_8$, **P8**-G_2, and **P8** models, figures showing torsion angles versus time during the stochastic dynamics simulations for peptide ϕ and ψ angles and carbohydrate Thr4 χ^1 and Thr5 χ^1 torsion angles, and figures showing the torsional flexibility of the stochastic dynamics simulations (peptide ϕ, ψ plots and carbohydrate χ^n, χ^{n+1} plots) (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Allinger, N. L. (1977) *J. Am. Chem. Soc. 99*, 8127.
Bhargava, A. K., Woitach, J. T., Davidson, E. A., & Bhavanandan, V. P. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6798–6802.
Breneman, C. (1990) *J. Comput. Chem. 11*, 361–373.
Darbon, N., Oddon, Y., Lacombe, J., & Pavia, A. A. (1984) *Carbohydr. Res. 130*, 55–64.
Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A., & Aubert, J. P. (1991) *Biochem. Biophys. Res. Commun. 175*, 414–422.
Dyson, H. J., & Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519–538.
Flippen, J. L. (1973) *Acta Crystallogr. B29*, 1881–1886.
Frisch, M. J., Head-Gordon, M., Trucks, G. W., Foresman, J. B., Schlegel, H. B., Raghavachari, K., Robb, M. A., Binkey, J. S., Gonzalez, C., Defrees, D. J., Fox, D. J., Whiteside, R. A., Seeger, R., Melius, C. F., Baker, J., Martin, R. L., Kahn, L. R., Stewart, J. J. P., Topiol, S., & Pople, J. A. (1990) GAUSSIAN, Inc., Pittsburgh, PA.
Gerken, T. A., & Dearborn, D. G. (1984) *Biochemistry 23*, 1485–1497.
Gerken, T. A. (1986) *Arch. Biochem. Biophys. 247*, 239–253.
Gerken, T. A., Butenhof, K. J., & Shogren, R. (1989) *Biochemistry 28*, 5536–43.
Gerken, T. A., & Jentoft, N. (1987) *Biochemistry 26*, 4689–4699.
Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A., & Kim, Y. S. (1989) *J. Biol. Chem. 264*, 6480–6487.
Ha, S. N., Giammona, A., Field, M., & Brady, J. W. (1988) *Carbohydr. Res. 180*, 207–221.
Hill, H. D., Schwyzer, M., Steinman, H. M., & Hill, R. L. (1977) *J. Biol. Chem. 252*, 3799–3804.
Homans, S. W. (1990) *Biochemistry 29*, 9110–9118.
Jentoft, N. (1990) *Trends Biochem. Sci. 15*, 291–294.
Jennings, H. J., & Bhattacharjee, A. K. (1977) *Carbohydr. Res. 55*, 105–112.
Jorgensen, W. L., & Tirado-Rives, J. J. (1988) *Am. Chem. Soc. 110*, 1657–1666.
Kooijman, H., Kroon-Batenburg, L. M. J., & Kroon, J. (1990) *Acta Crystallogr. C46*, 407–410.
Loncharich, R. J., & Brooks, B. R. (1989) *Proteins: Struct., Funct., Genet. 6*, 32–45.
Maeji, N. J., Inoue, Y., & Chûjô, R. (1987) *Biopolymers 26*, 1753–1767.
Mimura, Y., Inoue, Y., Maeji, N., & Chûjô, R. (1989) *Biopolymers 34*, 363–368.
Mimura, Y., Yasuhiko, Y., Inou, Y., & Chujo, R. (1992) *Int. J. Biol. Macromol. 14*, 242–248.
Porchet, N., VanCong, N., Dufosse, J., Audie, J. P., & Guyonnet-Rasmussen, K. (1982) *Acta Chem. Scand. A 36*, 323–327.
Ryckaert, J. P. (1984) *Mol. Phys. 55*, 549–556.
Shogren, R. L., Jamieson, A. M., Blackwell, J., & Jentoft, N. (1986) *Biopolymers 25*, 1505–1517.
Shogren, R. L., Jentoft, N., Gerken, T. A., Jamieson, A. M., & Blackwell, J. (1987) *Carbohydr. Res. 160*, 317–327.
Shogren, R. L., Gerken, T. A., & Jentoft, N. (1989) *Biochemistry 28*, 5525–5536.
Still, W. C., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990) *J. Am. Chem. Soc. 112*, 6127–6129.
Stewart, J. J. P. (1989) *MOPAC V5.0 (Quantum Chemistry Program Exchange 455)*, Frank J. Seiler Research Laboratory, USAF Academy, Colorado Springs, CO.
Tanpipat, N., & Mattice, W. E. (1990) *Biopolymers 29*, 377–383.
Timpte, C. S., Eckhardt, A. E., Abernethy, J. L., & Hill, R. L. (1988) *J. Biol. Chem. 263*, 1081–1088.
Wang, Y., Abernethy, J. L., Abernethy, J. L., Eckhardt, A. E., & Hill, R. L. (1992) *J. Biol. Chem. 267*, 12709–12716.
Weiner, S. J., Kollman, P. A., Case, P. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784.
Van Gunsteren, W. F., & Berendsen, H. J. C. (1988) *Mol. Simul. 1*, 173.