Structural Effects of O-Glycosylation on a 15-Residue Peptide from the Mucin (MUC1) Core Protein†

Leo Kirnarsky,‡ Om Prakash,§ Shawn M. Vogen,‡,|| Mitsuharu Nomoto,‡,⊥ Michael A. Hollingsworth,‡ and Simon Sherman*,‡

Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska 68198-6805, and Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506

Received May 3, 2000; Revised Manuscript Received July 26, 2000

ABSTRACT: To study the effect of O-glycosylation on the conformational propensities of a peptide backbone, the 15-residue peptide PPAHGVTSAPDTRPA (PPA15) from the MUC1 protein core and its analogue PPA15(**T**7), glycosylated with α-*N*-acetylgalactosamine on Thr7, were prepared and investigated by NMR spectroscopy. The peptide contains both the GVTSAP sequence, which is an effective substrate for GalNAc-T1 and -T3 transferases, and the PDTRP fragment, which is a well-known immunodominant epitope recognized by several anti-MUC1 monoclonal antibodies. Useful structural results were obtained in water upon decreasing the temperature to 5−10 °C. The sugar attachment slightly affected the conformational equilibrium of the peptide backbone near the glycosylated Thr7 residue. The clustering of low-energy conformations for both PPA15 and PPA15(**T**7) within the GVTSAP and APDTRP fragments revealed structural similarities between glycosylated and nonglycosylated peptides. For the GVTSAP region, minor but distinct clusters formed by either PPA15 or PPA15(**T**7) conformers showed distinct structural propensities of the peptide backbone specific for either the nonglycosylated or the glycosylated peptide. The peptide backbone of the APDTRP fragment, which is a well-known immunodominant region, resembled an S-shaped bend. A similar structural motif was found in the GVTSAP fragment. The S-shaped structure of the peptide backbone is formed by consecutive inverse γ-turn conformations partially stabilized by hydrogen bonding. A comparison of the solution structure of the APDTRP fragment with a crystal structure of the MUC1 peptide antigen bound to the breast tumor-specific antibody SM3 demonstrated significant structural similarities in the general shape.

The tandem repeat of the MUC1[1] core protein is a major site of O-glycosylation by several polypeptide *N*-acetylgalactosaminyltransferases (GalNAc-transferases). These GalNAc-transferases catalyze the attachment of an *N*-acetylgalactosamine (GalNAc) to the side chains of specific Thr/Ser residues *(1−5)*. Extensive in vitro studies have shown that the coordinated action of different polypeptide GalNAc-transferases is required for full glycosylation of the MUC1 tandem repeat. Prior O-glycosylation of some sites facilitates glycosylation of other sites (*6*, *7*). The molecular details of interactions between the O-linked GalNAc residue and protein core, which promote subsequent glycosylation events, are poorly understood. Further elucidation of this process requires investigation of the influence of primary glycosylation on the conformational propensities of the peptide backbone near sites of glycosylation and at more distant sites.

Glycopeptides are widely used to study effects of O-glycosylation on the conformational propensities of the polypeptide backbone, because these allow one to reduce the structural effects of global protein folding and to focus on specific interactions between the glycan moiety and polypeptide. NMR studies of glycopeptides provide insight into the influence of O-glycosylation on the conformation of glycosylated Thr/Ser and flanking residues (*8*−*16*). Previous results indicated that O-glycosylation could affect peptide conformations near the glycosylation site (*8*, *9*) as well as the conformational propensities of the peptide backbone at distant sites (*10*, *11*).

Our studies are focused on molecular mechanisms that influence the specificity and kinetics of the O-glycosylation reaction. In this paper, we performed the structural analysis

† This work was partially supported by NIH Grants 1RO1 CA69234 and 1RO1 CA57362 to M.A.H. and 1RO1 CA84106 and NE DHHS LB506 #00-33 to S.S. The Molecular Modeling Core Facility of the UNMC Eppley Cancer Center used in this work is supported by Cancer Center Support Grant P30 CA36727.

* To whom correspondence should be addressed.

‡ University of Nebraska Medical Center.

§ Kansas State University.

|| Present address: Department of Pathology and Committee on Immunology, The University of Chicago, Chicago, IL 60637.

⊥ Present address: Second Department of Pathology, Kagoshima University Faculty of Medicine, Kagoshima, Japan.

[1] Abbreviations for amino acids conform to the recommendations of the IUPAC−IUB Joint Commission on Nomenclature (*35*). Other abbreviations: MUC1, human mucin; GalNAc, *N*-acetylgalactosamine; GalNAc-transferase, uridine diphosphate-*N*-acetyl-α-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase; Fmoc, 9-fluorenylmethoxycarbonyl; HPLC, high-performance liquid chromatography; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; ROESY, rotating frame nuclear Overhauser enhancement spectroscopy; TOCSY, total correlation spectroscopy; COSY, correlation spectroscopy; DQF-COSY, double quantum filtered COSY; DMSO-d_6, dimethyl sulfoxide labeled with deuterium; RMSD, root-mean-square deviation.

10.1021/bi0010120 CCC: $19.00
Published on Web 09/02/2000

of a 15-residue MUC1 peptide, PPAHGV**T**SAPDTRPA (PPA15), and its glycosylated counterpart, PPA15(**T**7), by NMR and computational methods. PPA15(**T**7) was prepared with a GalNAc attached by an α linkage to Thr at position 7 (boldface). These peptides include the GVTSAP sequence, which is a highly effective substrate for the GalNAc-T1 and -T3 transferases. It has been shown that glycosylation of threonine at GVTSAP by the polypeptide transferases GalNAc-T1, -T2, or -T3 is a prerequisite for subsequent glycosylation of the serine at GVTSAP by GalNAc-T4 (*2*). In addition, both peptides included a sequence of the well-known immunodominant region, PDTRP, which is recognized by several anti-MUC1 monoclonal antibodies of important immunodiagnostic significance (*17*−*19*). Structural features of peptides with sequence comprising the immunodominant region of MUC1 were intensively studied to better understand the influence of glycosylation on the structure of tumor-associated epitopes, and to facilitate the design of immunogenic peptides for structurally based synthetic cancer vaccines (*12*, *20*−*22*). Our NMR studies were directed toward determining preferred conformations of PPA15 and PPA15(**T**7) in aqueous solution and identifying structural determinants that contribute to substrate−enzyme interactions during O-glycosylation reactions. NMR-derived structural data on the PDTRP fragment were also compared with the recently published crystal structure of the peptide epitope complexed with the tumor-specific antibody SM3 (*22*).

MATERIALS AND METHODS

Peptide Synthesis. Nonglycosylated peptides were synthesized by standard Fmoc solid-phase methodologies on an Applied Biosystems model 430A synthesizer. Side-chain deprotection and cleavage from the resin were achieved in a standard single-step acidolysis reaction. Peptides were purified and characterized as previously described (*23*).

Glycopeptides were synthesized using similar methodology. Tri-O-acetyl-GalNAc-α-Fmoc-threonine was obtained from Oxford Glycosciences (Bedford, MA). The residue was used in the standard Fmoc synthesis procedure but required an additional deprotection step, using a strong base, to obtain the final desired peptide. Throughout the synthesis, the coupling and deprotection cycles were monitored using the ninhydrin reaction.

Deprotection of the glycosylated peptide was accomplished in two steps. First, side-chain deprotection of standard protecting groups and cleavage from the resin were achieved in a single-step acidolysis reaction. This step generated the glycosylated peptide with acetyl groups protecting the hydroxyl moieties of the GalNAc sugar moiety. This peptide was purified using analytical and preparative HPLC on columns packed with C_{18}-bonded silica. The final step in the deprotection scheme involved dissolving 50 mg of purified acetylated GalNAc-peptide in 8 mL of 0.1 M NaOH for 5 min. The fully deprotected peptide was immediately purified by preparative HPLC. The removal of the acetyl-protecting groups was monitored by analytical HPLC. The GalNAc-peptide sequence was verified by amino acid composition analysis and mass spectrometry.

NMR Spectroscopy. High-resolution 1D and 2D ^{1}H NMR experiments were performed on a Varian UNITY plus 500 spectrometer (500 MHz). All NMR spectra were recorded in H_2O/D_2O (90:10) at 5, 10, 20, and 30 °C. NMR data were collected in a hypercomplex, phase-sensitive mode. Water peak suppression was obtained by low-power irradiation during relaxation delay (1.2 s). Standard pulse sequences were used for 2D TOCSY, NOESY, ROESY, and DQF-COSY experiments. Sequence-specific proton resonance assignments were established by a comparison of cross-peaks in a NOESY spectrum with those of a TOCSY spectrum acquired for the peptide under similar experimental conditions. Mixing times of 100, 200, 300, and 400 ms were employed for the NOESY experiments and 300 ms for the ROESY spectra in H_2O/D_2O (90:10). The temperature coefficients of amide protons were studied from 1D spectra collected in 5 °C intervals between 5 and 35 °C. The backbone proton−proton coupling constants $^3J_{N\alpha}$ were measured from DQF-COSY spectra and from 1D spectra at 5 °C.

Initial data processing utilized VNMR 6.1A software (Varian, Inc.) on a Silicon Graphics Indigo2 XZ workstation. Before processing, the t1 dimension of data sets from all experiments was zero-filled to 2K. When necessary, spectral resolution was enhanced by Lorenzian−Gaussian apodization. Proton chemical shifts and NOE cross-peak assignments were determined by standard procedures (*24*). Qualitative evaluations of NOE intensities were performed by visual inspection of the contour levels collected with a mixing time of 200 ms. The NOE peaks were divided into four groups with upper limits of 2.5, 3.0, 3.5, and 4.0 Å.

Structure Determination Protocol. The three-dimensional (3D) structures of the nonglycosylated and glycosylated mucin peptides were generated from the NMR data sets (NOE distance constraints and dihedaral angle constraints) using the structure determination protocol described previously (*25*). The COMBINE procedure (*26*), utilizing the FiSiNOE-3 (*27*) and HABAS computer programs (*28*), was used to narrow the allowed ranges for the ϕ, ψ, and χ_1 angles and to make stereospecific assignments. These torsion angles and the stereospecific assignments together with the intra-residue and sequential NOE distance constraints were used as input data for the DYANA program (*29*) implemented in the SYBYL 6.4 software package (TRIPOS Associates, Inc., St. Louis, MO) to generate 3D structures. Standard minimization levels and parameters for the DYANA program within SYBYL were used to generate 300 structures for each peptide, consistent with the NMR data set.

At the next stage, the 100 best structures of each peptide derived from DYANA were energy-minimized using the AMBER force field with the penalty function on distance constraint violations implemented in SYBYL. After energy minimization, all structures within energy intervals of 15 kcal/mol and with maximum violation of upper limits less than 0.25 Å were clustered using the root-mean-square deviation criterion (RMSD < 1.0 Å) for backbone heavy atoms, starting from the lowest-energy structure. For each cluster, the tightness of conformers was characterized by the mean pairwise RMSD for the backbone heavy atoms of the conformers and the corresponding average structure of the structural family (*25*). The assignment of possible hydrogen bonds was based on the distance and geometry of the probable donor/acceptor groups (120° < C═O...H angle < 240° and a CO...HN distance < 2.80 Å).

Table 1: Differences in ^{1}H Chemical Shifts (in ppm) between Nonglycosylated and Glycosylated Peptides in H_2O/D_2O (pH 4.5) at 5 °C

residue	HN	Hα	Hβ	Hγ
V6	nd[a]	0.11	nd	nd
T7	0.46	0.30	0.16	−
S8	0.28	nd	nd	−
A9	0.15	−0.20	nd	−

[a] nd: no differences detected.

RESULTS

NMR Analysis. NMR spectra for both peptides were collected under several different conditions. Useful structural information in water was obtained only at lower temperatures (5 and 10 °C) probably because of the lower sampling of conformational substates of these relatively small linear peptides (*10*). Chemical shifts of the ^{1}H resonances for PPA15 and PPA15(**T**7) peptides at 5 °C in H_2O are listed in Supporting Information Tables 1 and 2. The attachment of the carbohydrate moiety to the Thr7 residue mainly affected chemical shifts for Val6, Thr7, Ser8, and Ala9 (see Table 1).

Backbone proton−proton coupling constants $^3J_{N\alpha}$ were measured from one-dimensional spectra and ranged from 5 to 9 Hz, that might indicate an absence of ordered portions of a secondary structure.

Due to the overlap in amide resonances, temperature coefficients ($-\Delta\delta/\Delta T$) for the chemical shifts of amide protons were only obtained for some residues in each peptide. The values of the temperature coefficients varied from 5.3 to 8.3 ppb/K, indicating that no stable hydrogen bonds or solvent-shielded amide protons were present under the experimental conditions. However, lower values for temperature coefficients of the Val6 and Thr12 residues (~5.3 ppb/K) may suggest the partial involvement of these residues in the formation of more defined structural elements, including hydrogen bonds. Strong $d_{\alpha\delta}(i, i+1)$ connectivities for Ala9-Pro10 and Arg13-Pro14 were diagnostic for a trans conformation of the peptide backbone. The trans orientation of the peptide backbone for Pro1-Pro2 was accepted based on the absence of the $d_{\alpha\alpha}(i, i+1)$ contacts. The chemical shift index (CSI), calculated as a deviation of αH proton chemical shifts from "random-coil" values, was used to qualitatively represent a propensity for secondary structures (*30*). Sequential NOE connectivities and CSIs for PPA15 and PPA15(**T**7) are summarized in Figure 1.

The NOE connectivities and CSIs were very similar for both peptides except for the VTSA region. The αH chemical shifts are close to the values of a random-coil structure in water. In aqueous solution, neither peptide showed distinct evidence of a stabilized secondary structure. A large number of strong consecutive $d_{\alpha N}(i, i+1)$ connectivities indicated a predominance of extended backbone conformations. These NMR data are characteristic for linear peptides in solution that show a rapidly fluctuating set of conformers in preferably random (extended) conformations (*31*). Sequential interactions, $d_{NN}(i, i+1)$, were observed in Gly5-Val6-Thr7 and Asp11-Thr12-Arg13 fragments for both glycosylated and nonglycosylated peptides.

Upon sugar attachment, small but distinct changes in the NOE connectivities were observed near the glycosylation site.

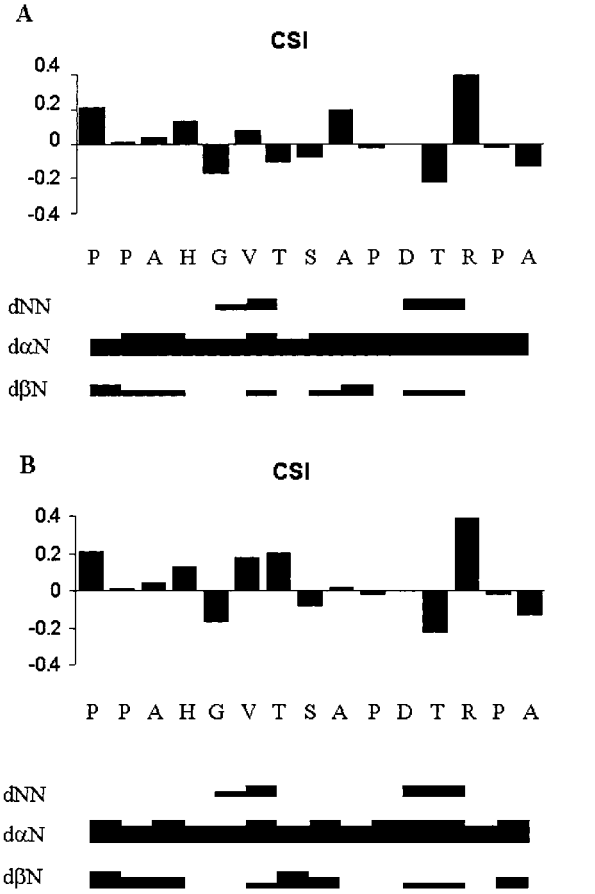

FIGURE 1: Summary of the Cα chemical shift indices (CSI) and sequential ($i, i+1$) NOE connectivities for nonglycosylated (A) and glycosylated (B) in H_2O/D_2O (pH 4.5) at 5 °C. Line thickness for the NOEs corresponds to the NOE intensity. In the case of proline, NH refers to δH protons.

The glycosylated peptide, PPA15(**T**7), revealed several NOE cross-peaks related to sugar−peptide interactions. The methyl group of galactosamine was involved in weak contacts with the amide proton and the side chain of the Ala9. Also, a weak NOE cross-peak was observed between the H5 proton of the sugar ring and the methyl group of the side chain of Thr7. One strong contact was observed in water between the amide proton of the glycosylated Thr7 residue and the *N*-acetyl group of galactosamine under all temperature conditions (Figure 2).

The limited set of NMR-derived constraints for these relatively small, linear peptides produced a wide variation in the conformers generated by DYANA. For each peptide, the 100 best conformers from the DYANA calculations were chosen for energy minimizations with the NOE-derived constraints. After energy minimization, a final set of 62 low-energy structures (<15 kcal/mol) was selected for the nonglycosylated PPA15 peptide. The glycosylated peptide, PPA15(**T**7), yielded a set of 52 structures. Both glycosylated and nonglycosylated structures (114 conformers) were grouped in structural clusters using RMSD criteria for polypeptide backbone heavy atoms. Two overlapping peptide fragments, GVTSAP and APDTRP, which represent glycosylated Thr7 with flanking residues and the more distant immunodominant region of MUC1, were selected for cluster analysis to determine the structural effects of glycosylation.

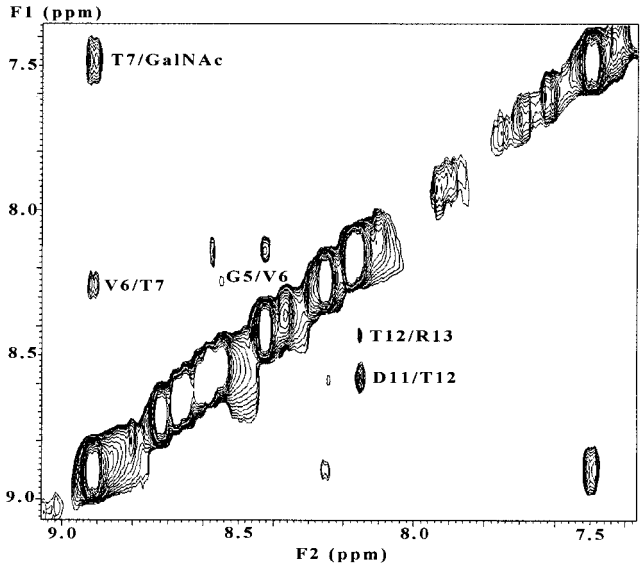

FIGURE 2: NH−NH region of the NOESY spectrum in H_2O/D_2O (90:10) at 5 °C (200 ms mixing time) for the glycosylated peptide PPA15(**T**7). Relevant peaks are labeled. T7/GalNAc indicates the cross-peak between the amide protons of Thr7 and the *N*-acetyl group of the *N*-acetylgalactosamine unit.

For the GVTSAP segment, four distinct clusters were obtained using the RMSD criterion of 1.0 Å. The most populated cluster included 28 conformers of the nonglycosylated peptide and 23 conformers of the glycosylated peptide. The mean pairwise RMSD for the backbone heavy atoms of these conformers and the corresponding average structure was equal to 0.46 ± 0.10 Å. The second cluster of 28 mixed structures, nonglycosylated and glycosylated, contained the lowest-energy conformer of the glycosylated peptide. The mean pairwise RMSD for the backbone heavy atoms of the second cluster was 0.64 ± 0.12 Å. The lowest-energy structure of the nonglycosylated peptide fell within the third distinct cluster of 18 conformers of the nonglycosylated peptide with RMSD equal to 0.62 ± 0.12 Å. The fourth cluster with RMSD equal to 0.42 ± 0.13 Å included 17 conformers of the glycosylated peptide. Thus, 69% of all low-energy structures for the GVTSAP fragment fell in two clusters that included conformers of both nonglycosylated and glycosylated peptides, while 31% of structures were assigned to the separate clusters with conformers of either nonglycosylated or glycosylated peptide.

The existence of separate structural clusters composed of conformers of either nonglycosylated or glycosylated peptide suggests that these clusters represent distinct conformational propensities. To evaluate conformational features specific to each cluster, the mean values and standard deviations of the backbone conformational angles ϕ and ψ were calculated and compared for each peptide residue in a cluster. A distinguishable difference between the two most populated clusters with mixed content was found for the Thr7 residue. The mean values and standard deviations of the ϕ angle of Thr7 were −98 ± 27° for the first cluster and −148 ± 22° for the second cluster. Distinguishable variations in the conformational angles between two distinct clusters with either nonglycosylated or glycosylated conformers were detected for the Thr7 and Ser8 residues. The mean values and standard deviations of the ψ angle of Thr7 for the nonglycosylated and glycosylated peptide were 67 ± 11° and 159 ± 8°, respectively, while the ϕ angle of Ser8 was equal to −82 ± 4° for the nonglycosylated peptide and −50 ± 8° for the glycosylated peptide. These variations of the conformational angles resulted in different relative orientation of the side chains of the Thr7 and Ser8 residues for the structures within the clusters of either nonglycosylated or glycosylated peptide. The relative orientation of side chains was evaluated by the virtual torsion angle $C^{\beta}-C^{\alpha}$ (T7)...$C^{\alpha}-C^{\beta}$ (S8) defined by rotating vectors of $C^{\alpha}-C^{\beta}$ bonds of Thr7 and Ser8 around the virtual bond between their C^{α} atoms. The mean value of the virtual torsional angles for the conformers within the cluster of either nonglycosylated or glycosylated peptide was equal to 164° ± 14 and −84° ± 15, correspondingly.

Four structural forms were observed for the GVTSAP fragment. All of these forms were composed of mostly extended structures that included inverse γ-turn-like motifs formed by triplets of either His4-Gly5-Val6, Gly5-Val6-Thr7, or Val6-Thr7-Ser8. The relative position and the number of inverse γ-turns varied from none to three consecutive turns. Conformations with three consecutive inverse γ-turns were observed only for the nonglycosylated peptide. Glycosylation enables hydrogen bonding between the carbohydrate and CO groups of Gly5 or Thr7 and reduces the propensity to form the consecutive inverse γ-turns in this region.

The distinct structural features of this inverse γ-turn-like conformation include the proximity of the CO group of residue i and the NH group of residue $i + 2$ and/or conformational angles ϕ and ψ for the $i + 1$ residue that are typical for an inverse γ-turn (*32*). While a conventional reversed γ-turn is assumed to fold the peptide chain back (*32*), the GVTSAP fragment demonstrated a mostly extended structural shape. Thus, the term "γ-turn-like" designates this specific conformation assumed by the central $i + 1$ residue in the above triplets and includes the possibility of hydrogen bonding between the i and $i + 2$ residues.

For the APDTRP fragment, the cluster analysis also resulted in four clusters. The largest cluster included 36 nonglycosylated and 28 glycosylated low-energy structures with the mean pairwise RMSD of 0.65 ± 0.13 Å. The second cluster of 37 conformers also consisted of a mix of nonglycosylated and glycosylated structures with an RMSD equal to 0.68 ± 0.17 Å. In addition, two small clusters consisting of eihther nonglycosylated or glycosylated structures were obtained with RMSDs of 0.61 ± 0.16 and 0.51 ± 0.13 Å, respectively. Most structures (88%) fell into two clusters of mixed (glycosylated and nonglycosylated) conformers. This suggests that GalNAc attachment to the Thr7 residue only slightly (if at all) affected the conformational propensities of the structural ensemble of the peptide conformers at the PDTR site. The finding of distinct structural clusters of either nonglycosylated or glycosylated peptide conformers may indicate that there are unique structural features specific for each peptide. However, in the absence of valid NMR evidences of structural changes for the PDTR fragment, the meaningfulness of these small structural clusters is questionable. The comparison of two clusters composed of mixed conformers showed that only one conformational angle, ϕ of the Thr12 residue, was altered significantly (more than 30°). Conformations of the Asp11 residue in all clusters were very similar, $\phi \sim -80°$and $\psi \sim +65°$, which are typical for the central residues of inverse

γ-turns (*32*). The same inverse γ-turn-like conformation was found for the Thr12 residue in both mixed clusters. For these clusters, conformations of Arg13 fell within the B area of the Ramachandran plot (left-upper quadrant) with the ϕ angles within the $-95° < \phi < -120°$ range and the ψ angles within the $100° < \psi < 140°$ range. The conformational angle ψ of Pro10 was about 110°. Thus, the structural features of the APDTRP fragment included mostly extended conformations combined with inverse γ-turn-like motifs of the PDT portion. The peptide backbone of the APDTRP fragment resembled an S-shaped bend where Asp and Thr residues formed the most curved part. Similar bending can be seen on the GVTSAP fragment.

Structural proximity and spatial orientation of potential hydrogen-bond donors and acceptors were analyzed to define possible hydrogen bonding. Within the GVTSAP fragment, CO groups of His4, Gly5, and Val6 may form hydrogen bonds with amide protons of Val6, Thr7, and Ser8, respectively. In the APDTRP fragment, CO groups of Ala9, Pro10, and Asp11 may serve as hydrogen-bond acceptors, while potential donors are the NH groups of Asp11, Thr12, and Arg13. Both sets of energy-minimized structures, nonglycosylated and glycosylated, demonstrated the possibility for hydrogen bonding within each fragment formed by these donor/acceptor groups; however, the number and position of these hydrogen bonds varied.

In addition, many conformers showed the possibility of hydrogen bonding that involved the side-chain carboxyl group of Asp11 and the guanidinium group of Arg13. Possible donors of hydrogen bonds with COO^- of Asp11 included the amide proton and the guanidinium group of the Arg13 residue. Also, the guanidinium group of Arg13 may be involved in hydrogen bonding with either of the backbone carbonyls of Thr12 or Pro14, or the C-terminal carboxyl. Different combinations of these hydrogen bonds stabilized structures of the APDTRP structural ensemble.

DISCUSSION

This study pursued two major goals. The first goal was to determine structural features of the APDTRP fragment, a well-known, cancer-specific peptide antigen (*17*−*19*). The second goal was to understand how the attachment of a GalNAc residue to Thr affects peptide backbone conformation within the GVTSAP fragment and the more distant APDTRP fragment.

The structure of the APDTRP fragment, a tumor marker and potential candidate for developing cancer vaccines against several adenocarcinomas, has been studied by NMR spectroscopy (*11, 20, 21, 25*). Two structural models have been proposed for this immunodominant region of MUC1. The first model postulated a type I reverse β-turn formed by the PDTR (*20*). The model was based on NMR data collected in DMSO-d_6 for 11- and 20-residue peptides. The same type I β-turn structure was assumed for a 16-residue peptide in the mixture of methanol and water (*12*). A second model consisting of a type II β-turn formed by residues APDT was proposed as a key element of a knob-like structure within the immunodominant region (*21*). This model was suggested for the 60-residue fragment (3 consecutive tandem repeats) in an aqueous solution. In our previous NMR study of a series of 9-residue peptides, which included the PDTR fragment at different positions, we considered structural features of the reactive Thr and several flanking residues (*25*). The PDTR fragment was shown to form internally self-stabilized substructures, which were independent of positioning within the peptide.

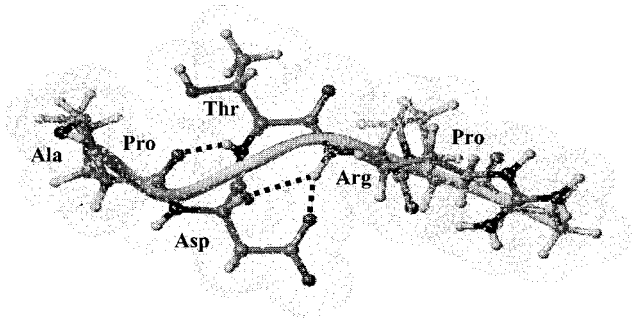

FIGURE 3: Average structure of the most populated cluster from the ensemble of the APDTR structures. Possible hydrogen bonds between the CO group of Pro and the NH group of Thr, the CO group of Asp and the NH of Arg, and the side-chain carboxyl group of Asp and the NH of Arg are shown by broken lines. The S-shaped bend of the peptide backbone (shown by a tube) is formed by two overlapping inverse γ-turns (double γ-turn) in positions PDT and DTR.

NMR data on the APDTR fragment presented here are very similar to those described previously (*12, 20, 21, 25*). However, the interpretation of these data and the assignment of structures are somewhat different from the suggested β-turn structures of type I or II. Indeed, the data presented here as well as NMR data reported previously (*12, 20, 21*) do not provide direct NMR evidences supporting the existence of either type I β-turn or type II β-turn. For a type I β-turn, the NMR data should demonstrate characteristic $d_{\alpha N}(i, i + 3)$ and $d_{NN}(i, i + 2)$ NOE connectivities (*24*). For a type II β-turn, $d_{NN}(i, i + 2)$ connectivities should be observed (*24*). Our structural modeling was based on NMR-derived constraints and did not presume any particular type of secondary structure. The NMR data showed that in an aqueous solution the APDTR fragment adopted a compact structure that was stabilized by backbone−backbone and side-chain−backbone interactions. The backbone structure of this fragment resembled an S-like bend, in which the Asp and Thr residues adopted mainly half-twisted conformations and formed the curved component of the bend. One can envision a structural equilibrium between this S-shaped bend and structures in which inverse γ-turns formed by the PDT and/or DTR fragments are the major components. Some conformers from the APDTR structural ensemble simultaneously had two inverse γ-turns, overlapping at position DT (a double inverse γ-turn). Different combinations of hydrogen bonds stabilized the structures of the APDTR structural ensemble. The average structure of the most populated cluster of the APDTR fragment is shown in Figure 3.

A similar S-shaped, multiple-turn conformation was found in the crystal structure of the peptide epitope from the HIV-1 V3 loop bound to the neutralizing antibody (*33, 34*). The same S-shaped motif may be seen in the 'wave-type' conformation of the peptide backbone newly described in a hexaglycosylated decapeptide from human glycophorin A (*16*).

Recently, the crystal structure of the breast tumor-specific antibody SM3 complexed with a 13-residue MUC1 peptide

Table 2: Comparison of Published NMR-Derived Conformations of the APDTRP Fragment and the Average Structure from the Most Populated Cluster of PPA15 and PPA15(**T**7) Peptides with the Crystal Structure[a]

residue	torsion angle	Scanlon et al. (*20*)	Liu et al. (*12*)	Fontenot et al. (*21*)	PPA15 and PPA15(**T**7)	X-ray (*22*)
Ala	ϕ	−	−	−159	−126	−92
	ψ	−	−	148	139	160
Pro	ϕ	−	−107	−66	−75	−81
	ψ	−	−13	161	114	195
Asp	ϕ	−44	−65	57	−82	−78
	ψ	−54	−5	40	64	101
Thr	ϕ	−118	−106	−110	−122	−97
	ψ	18	−3	35	68	16
Arg	ϕ	−	−127	43	−103	−71
	ψ	−	119	67	120	143
Pro	ϕ	−	−	−64	−72	−65
	ψ	−	−	163	152	135

[a] Values of the ϕ, ψ torsion angles are given in degrees.

antigen was resolved by X-ray methods (*22*). The peptide antigen, TSAPDTRPAPGST, included the APDTRP fragment and a portion of the GVTSAP sequence. It is likely that the bound conformation of this peptide antigen complexed with the SM3 antibody reflects some biologically important structural features of the MUC1 immunodominant region. The conformation of the peptide antigen was found to be extended, with a "kinked" backbone conformation between Asp5 and Arg7 (*22*). This kinked backbone conformation was defined by two hydrogen bonds between 1) the side chain of Asp5 and the amide proton of Arg7, and 2) the CO group of Pro4 and the NH group of Thr6. A comparison of the crystal structure with the previously described knob-like solution structure (*21*) revealed significant differences in overall shape, which were interpreted as an apparent conformational change between the bound and unbound MUC1 peptide epitope (*22*). In contrast, a comparison of the average structure from the most populated cluster of the PPA15 and PPA15(**T**7) conformers with the MUC1 epitope bound to SM3 (*22*) demonstrated significant structural similarities in the general shape of the APDTRP fragment in solution and crystal. The published solution conformations of the APDTRP fragment and the average structure for the most populated cluster of the PPA15 and PPA15(**T**7) peptides are compared with the crystal structure in Table 2.

The type I β-turn structure (*12*, *20*) is not consistent with the crystal structure for the conformations of the Pro-Asp residues. The type II β-turn structure (*21*) is inconsistent with the Asp and Arg conformations observed in the crystal structure. As can be seen from the table, the average structure for the most populated cluster of the PPA15 and PPA15(**T**7) conformers presented here was well-correlated with the crystal structure. The suggested S-shaped bend for the PDTR fragment resembles a conformation that is defined in the crystal structure, rather than the knob-like motif (*21*).

The O-glycosylation of Thr7 caused changes in the chemical shifts of a few protons near the glycosylation site and slightly affected structural features of the conformational ensemble of the GV**T**SAP fragment. The extended backbone conformations that included inverse γ-turn-like structural motifs were predominant among the structural families. The molecular modeling results indicated that the monosaccharide was positioned almost perpendicular to the peptide backbone. Such positioning is consistent with our observation of the strong contact between the *N*-acetyl NH proton of the GalNAc and the amide proton of the glycosylated Thr7 residue and, simultaneously, with the NOE cross-peaks observed between the methyl group of galactosamine and the Ala9 residue.

Previously, we proposed a structural model for an effective acceptor substrate for the polypeptide GalNAc-transferases that could explain some structural aspects of a substrate specificity (*25*). This model was based on the NMR data and in vitro kinetic studies of the substrate specificities for two purified recombinant transferases, GalNAc-T1 and -T3. We concluded that an inverse γ-turn conformation of the residue to be glycosylated and an extended conformation of the flanking residues was an important structural motif for effective substrates. The role of hydrogen bonding and the involvement of distant sites in substrate−enzyme interactions were also discussed. Our NMR studies of PPA15 and PPA15(**T**7) showed very similar structural motifs in both GVTSAP and APDTRP fragments. However, O-glycosylation of these fragments is known to be catalyzed in a different fashion by the orchestrated action of the distinct GalNAc-transferases (*6*, *7*). To date, six members of the mammalian GalNAc-transferase family have been reported (*1*−*5*), and additional members are expected to be cloned and purified. It is now clear that full O-glycosylation of the MUC1 tandem repeat requires the combined action of several GalNAc-transferases that act in an ordered manner.

As was shown recently, the activity of the GalNAc-T4 transferase, which utilized two unique sites in the MUC1-like peptides not used by GalNAc-T1, -T2, or -T3, requires prior glycosylation of the peptide substrate by these transferases (*2*). For example, glycosylation of the Ser residue in VTSA by GalNAc-T4 required a prior glycosylation of the preceding Thr. It is possible that prior glycosylation of threonine induces specific conformational changes in the substrate peptide backbone that are favorable for GalNAc-T4 activation. Our NMR data showed that glycosylation caused mild but distinct effects on conformational propensities of the polypeptide backbone of the GVTSA fragment. Two distinct structural clusters for glycosylated and nonglycosylated peptide differed in the degree of an extended conformation for the glycosylated threonine and in the relative orientation of the side chains of Thr7 and Ser8. The flanking Ser residue, which is glycosylated by GalNAc-T4, demonstrated a conformation ($\phi \sim -50°$ and $\psi \sim +160°$) that was unique among the four observed structural clusters. This specific conformation combined with the extended conformation of the preceding glycosylated threonine may be responsible for the activation of the substrate for glycosylation that is catalyzed by GalNAc-T4. The lack of an analogous structural motif for the Thr residue within the DTR may indicate that prior O-glycosylation within a distant -GSTA- of the tandem repeat may be required to induce proper conformational changes of the peptide backbone for the PDTRPAPGSTA.

SUPPORTING INFORMATION AVAILABLE

Supporting information is available on the chemical shifts for nonglycosylated and glycosylated peptides (Tables 1 and 2) (2 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Clausen, H., and Bennett, E. P. (1996) *Glycobiology 6*, 635−646.
2. Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitrou, J., Hollingsworth, M. A., Merkx, G., van Kessel, A. G., Eiberg, H., Steffensen, R., and Clausen, H. (1998) *J. Biol. Chem. 273*, 30472−30481.
3. Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merkx, G., van Kessel, A. G., Olofsson, S., and Clausen, H. (1999) *J. Biol. Chem. 274*, 25362−25370.
4. Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., VanWuyckhuyse, B., and Tabak, L. A. (1998) *J. Biol. Chem. 273*, 27749−27754.
5. Ten Hagen, K. G., Tetaert, D., Hagen, F. K., Richet, C., Beres, T. M., Gagnon, J., Balys, M. M., VanWuyckhuyse, B., Bedi, G. S., Degand, P., and Tabak, L. A. (1999) *J. Biol. Chem. 274*, 27867−27874.
6. Müller, S., Goletz, S., Packer, N., Gooley, A. A., Lawson, A. M., and Hanisch, F.-G. (1997) *J. Biol. Chem. 272*, 24780−24793.
7. Hanisch, F. G., Muller, S., Hassan, H., Clausen, H., Zachara, N., Gooley, A. A., Paulsen, H., Alving, K., and Peter-Katalinic, J. (1999) *J. Biol. Chem. 274*, 9946−9954.
8. Andreotti, A. H., and Kahne, D. (1993) *J. Am. Chem. Soc. 115*, 3352−3353.
9. Liang, R., Andreotti, A. H., and Kahne, D. (1995) *J. Am. Chem. Soc. 117*, 10395−10396.
10. Huang, X., Smith, M. C., Berzofsky J. A., and Barchi, J. J., Jr. (1996) *FEBS Lett. 393*, 280−286.
11. Huang, X., Barchi, J. J., Jr., Lung, F.-D. T., Roller, P. P., Nara, P. L., Muschik, J., and Garrity, R. R. (1997) *Biochemistry 36*, 10846−10856.
12. Liu, X., Sejbal, J., Kotovych, G., Koganty, R. R., Reddish, M. A., Jackson, L., Gandhi, S. S., Mendonca, A. J., and Longenecker, B. M. (1995*) Glycoconj. J. 12*, 607−617.
13. Satyanarayana, J., Gururaja, T. L., Naganagowda, G. A., Ramasubbu, N., and Levine, M. J. (1998) *J. Pept. Res. 52*, 165−179.
14. Spencer, D. I. R., Missailidis, S., Denton, G., Murray, A., Brady, K., Matteis, C. I. D., Searle, M. S., Tendler, S. J. B., and Price, M. R. (1999) *Biospectroscopy 5*, 79−91.
15. Live, D. H., Williams, L. J., Kuduk, S. D., Schwarz, J. B., Glunz, P. W., Chen, X.-T., Sames, D., Kumar, R. A., and Danishefsky, S. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 3489−3493.
16. Schuster, O., Klich, G., Sinnwell, V., Kränz, H., Paulsen, H., and Meyer, B. (1999*) J. Biomol. NMR 14*, 33−45.
17. Burchell, J., Gendler, S., Taylor-Papadimitrou, J., Girling, A., Lewis, A., Millis, R., and Lamport, D. (1987) *Cancer Res. 47*, 5476−5482.
18. Burchell, J., Taylor-Papadimitrou, J., Boshell, M., Gendler, S., and Duhig, T. (1989) *Int. J. Cancer 44*, 691−696.
19. Girling, A., Bartkova, J., Bruchell, J., Gendler, S., Gillett, C., and Taylor-Papadimitriou, J. (1989) *Int. J. Cancer 43*, 1072−1076.
20. Scanlon, M. J., Morley, S. D., Jackson, D. E., Price, M. R., and Tendler, S. J. B. (1992) *Biochem. J. 284*, 137−144.
21. Fontenot, J. D., Mariappan, S. V. S., Catasti, P., Domenech, N., Finn, O. J., and Gupta, G. (1995) *J. Biomol. Struct. Dyn. 13*, 245−260.
22. Dokurno, P., Bates, P. A., Band, H. A., Stewart, L. M. D., Lally, J. M., Burchell, J. M., Taylor-Papadimitriou, J., Snary, D., Sternberg, M. J. E., and Freemont, P. S. (1998) *J. Mol. Biol. 284*, 713−728.
23. Sanderson, S. D., Kirnarsky, L., Sherman, S. A., Ember, J. A., Finch, A. M., and Taylor, S. M. (1994) *J. Med. Chem. 37*, 3171−3180.
24. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, John Wiley and Sons, New York.
25. Kirnarsky, L., Nomoto, M., Ikematsu, Y., Hassan, H., Bennett, E. P., Cerny, R. L., Clausen, H., Hollingsworth, M. A., and Sherman, S. (1998) *Biochemistry 37*, 12811−12817.
26. Sherman, S., Sclove, S., Kirnarsky, L., Tomchin, I., and Shats, O. (1996) *J. Mol. Struct.: THEOCHEM 368*, 153−162.
27. Shats, O., and Sherman, S. (1996) *Third Electronic Computational Chemistry Conference (ECCC-3)*, (URL − http:/www.unmc.edu/Eppley/ECCC3/fisinoe3.htm).
28. Güntert, P., and Wüthrich, K. (1989) *J. Am. Chem. Soc. 111*, 3997−4004.
29. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
30. Wishart, D. S., Sykes, B. D., and Richards, M. F. (1992) *Biochemistry 31*, 1647−1651.
31. Dyson, H. J., and Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−538.
32. Smith, J. A., and Pease, L. G. (1980) *CRC Crit. Rev. Biochem. 8*, 315−399.
33. Ghiara, J. B., Stura, E. A., Stanfield, R. L., Profy, A. T., and Wilson, I. A. (1994) *Science 264*, 82−85.
34. Ghiara, J. B., Ferguson, D. C., Satterhwait, A. C., Dyson, H. J., and Wilson, I. A. (1997) *J. Mol. Biol. 266*, 31−39.
35. IUPAC−IUB Joint Commission on Nomenclature (1972) *J. Biol. Chem. 247*, 977−983.

BI0010120