Structural Analysis of Peptide Substrates for Mucin-Type O-Glycosylation†

Leonid Kirnarsky,‡ Mitsuharu Nomoto,‡ Yoshito Ikematsu,‡,§ Helle Hassan,|| Eric P. Bennett,|| Ronald L. Cerny,⊥ Henrik Clausen,|| Michael A. Hollingsworth,‡ and Simon Sherman*,‡

Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska 68198-6805, School of Dentistry, University of Copenhagen, DK-2200 Copenhagen N, Denmark, and Department of Chemistry, University of Nebraska, Lincoln, Nebraska 68588-0304

Received May 5, 1998; Revised Manuscript Received July 2, 1998

ABSTRACT: The structures of three nine-residue peptide substrates that show differential kinetics of O-linked glycosylation catalyzed by distinct recombinant uridine diphosphate−*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferases (GalNAc transferases) were investigated by NMR spectroscopy. A combined use of NMR data, molecular modeling techniques, and kinetic data may explain some structural features required for O-glycosylation of these substrates by two GalNAc transferases, GalNAc-T1 and GalNAc-T3. In the proposed model, the formation of an extended backbone structure at the threonine residue to be glycosylated is likely to enhance the O-glycosylation process. The segment of extended structure includes the reactive residue in a β-like or an inverse γ-turn conformation and flanking residues in a β-strand conformation. The hydroxyl group of the threonine to be glycosylated is exposed to solvent, and both the amide proton and carbonyl oxygen of the peptide backbone are exposed to solvent. The exchange rate of the amide proton for the reactive threonine correlated well with substrate efficiency, leading us to hypothesize that this proton may serve as a donor for hydrogen bonding with the active site of the enzyme. The oxygens of the residue to be glycosylated and several flanking residues may also be involved in a set of hydrogen bonds with the GalNAc-T1 and -T3 transferases.

The molecular parameters that govern the specificity and kinetics of mucin-type O-linked glycosylation of serine and threonine with *N*-acetylgalactosamine by uridine diphosphate−*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferases (GalNAc transferases)[1] remain poorly understood. O-Glycosylation of glycoproteins is influenced by protein trafficking, parameters of substrate protein folding, and levels of enzyme activity (*1*, *2*); in contrast, the relative importance of acceptor substrate primary amino acid sequence to this reaction has been widely debated (*3*−*11*). A consensus primary amino acid sequence for O-glycosylation has not been found; however, there is substantial evidence that sequences flanking serine and threonine residues significantly affect catalytic activity of GalNAc transferases (*4*, *5*, *8*−*10*, *12*).

O-Glycosylation is carried out by a large family of enzymes (for review, see ref *13*) that recently has been characterized by biochemical and cDNA cloning techniques. Recombinant forms of these enzymes have been investigated for relative catalytic efficiency and substrate specificity (*12*). The results of initial studies show that distinct GalNAc transferases have both common and distinct activities and specificities with different substrates. This fact partially explains previous difficulties in defining distinct parameters of acceptor substrates that influence O-glycosylation.

Mucins and mucin-like proteins are important substrates for GalNAc transferases because they are among the most heavily O-glycosylated glycoproteins. O-Linked carbohydrates on mucins are important to their function, which includes lubrication and protection of cellular surfaces of ductal epithelia and a possible role in innate immunity against certain pathogenic bacteria and viruses. Some human tumors overproduce mucins that are glycosylated differently than those produced by corresponding normal tissues. This may be of some advantage to the survival and metastatic properties of those tumors (*14*−*17*). Mucins and many mucin-like glycoproteins contain distinct domains that are glycosylated at different densities. The heavily glycosylated domains frequently include a tandem repeat, which consists of tandem arrays of identical or nearly identical peptide sequences (within one domain) containing a high percentage of potential sites of O-glycosylation (serine and threonine residues), often in combination with proline, alanine, glycine,

† This work was partially supported by NIH Grant RO1 CA69234 to M.A.H. The Molecular Modeling Core Facility and the Nuclear Magnetic Resonance Core Facility of the UNMC Eppley Cancer Center used in this work are supported by Cancer Center Support Grant P30 CA36727.

* To whom correspondence should be addressed.

‡ University of Nebraska Medical Center.

§ Present address: Second Department of Surgery, Nagasaki University School of Medicine, 1-7-1 Sakamoto, Nagasaki, 852-8501 Japan.

|| University of Copenhagen.

⊥ University of Nebraska.

[1] Abbreviations for amino acids conform to the recommendations of IUPAC−IUB Joint Commission on Nomenclature [(1972) *J. Biol. Chem. 247*, 977−983]. Other abbreviations: MUC1, human mucin; GalNAc transferase, uridine diphosphate−*N*-acetylgalactosamine: polypeptide N-acetylgalactosaminyltransferase; NMR, nuclear magnetic resonance; ROESY, rotating frame nuclear Overhauser enhancement spectroscopy; TOCSY, total correlation spectroscopy; COSY correlation spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; DMSO-d_6, dimethyl sulfoxide labeled with deuterium; 2D, two-dimensional; 3D, three-dimensional; PDB, Protein Data Bank; RMSD, root-mean-square deviation.

S0006-2960(98)01034-4 CCC: $15.00
Published on Web 08/28/1998

or valine. There is great variability in the length and primary sequence of different tandem repeats; however, these motifs have been found in many mucin-like proteins described to date (*2*). It is hypothesized that one function of the tandem repeat motif is to serve as a scaffold for extensive O-glycosylation of these proteins. In this case, each repeated unit is predicted to contain sequences that are effective substrate sites for GalNAc transferases, which, when placed in tandem array, may be capable of producing a structure that receives a high density of O-glycosylation.

The repeated nature of mucin-like tandem repeats and the presence of multiple acceptor substrate sites for GalNAc transferases provide a useful paradigm to evaluate the influence of acceptor substrate sequence on GalNAc transferase activity. O-Glycosylation of the tandem repeat domain of MUC1 by total GalNAc transferase activity has previously been investigated in human tumor cells (*8, 9*). Two-thirds of MUC1 is composed of multiple copies of the tandem repeat (from 18 to over 100, depending on allele length, which is highly polymorphic) of the sequence GVTSAP-DTRPAPGSTAPPAH. Previous in vitro studies with the pancreatic tumor cell extracts showed that two of three threonines (those at GVTSA and GSTAP, but not at PDTRP) and neither of two serines in the native tandem repeat sequence of MUC1 are glycosylated. There was also evidence that the two sites were glycosylated at different relative rates and that amino acid sequence, peptide length, and relative position of the residue to be glycosylated dramatically affect the ability of the peptides to serve as acceptor substrates for GalNAc transferases. Subsequent studies have shown that these tumor cell extracts contained multiple GalNAc transferases, which show both distinct and common features in substrate specificity and kinetic properties (*12, 18*). Studies with three different human recombinant GalNAc transferases showed that all three enzymes glycosylated the MUC1 tandem repeat peptide at three sites GVTSAPDTRPAPGSTAPPAH under extended reaction conditions (time), albeit at different reaction sites. Other studies suggest that additional enzymes with novel substrate specificities for the other two potential glycosylation sites in the MUC1 tandem repeat exist and are expressed in vivo (*19*).

Determination of sites and relative kinetics of glycosylation of different sites by distinct GalNAc transferases for MUC1 is of more than academic interest, since different glycoforms of MUC1 are produced by different normal tissues and cell types. Some of these glycoforms of MUC1 may result from glycosylation of distinct positions along the tandem repeat. Differences among organs, cells, and tumors probably have functional significance and may be useful in designing diagnostic and therapeutic reagents for cancers that express MUC1 (*8, 14, 17*). A recent report demonstrates that site-specific glycosylation of MUC1 influences the binding affinity of some monoclonal antibodies (*20*).

Our lack of knowledge of the structure of the active site of GalNAc transferases and of the precise topography of O-glycosylation sites on acceptor substrates makes any analysis of structural features responsible for O-glycosylation hypothetical. Moreover, short conformationally unrestricted peptides are generally considered to be an ensemble of interconverting conformers in solution that yield an averaged weighted conformation on the NMR time scale. Such conformation as well as any particular conformer may or may not represent specific structural features required for substrate−enzyme interactions. Nonetheless, the seeds of self-stabilized protein structure that contribute to interactions between substrate and enzyme during enzymatic catalysis are probably contained in short primary amino acid sequences of peptide substrates. Therefore, comparative analysis of structures for effective substrates and inactive analogues with similar amino acid sequences, combined with kinetic data for enzymatic catalysis, should provide useful insights into the molecular mechanisms of substrate recognition by the GalNAc transferases. Previous investigations of O-linked glycosylation sites have been mostly restricted to comparisons of substrate primary amino acid sequences (*5, 7*). There has not been a systematic investigation of the spatial structural features of active and inactive peptide substrate analogues. In this report, we present NMR-derived structural data for three nine-residue peptides and an analysis of the catalytic efficiency of three recombinant GalNAc transferases with these peptide substrates.

MATERIALS AND METHODS

Peptide Synthesis and Sample Preparation. The peptides were synthesized on a 0.25 mmol scale synthesis by standard Fmoc solid-phase methodologies on an ABI (Foster City, CA) Model 430A synthesizer. Side-chain deprotection and cleavage from the resin were achieved in a single-step acidolysis reaction. The peptides were purified by analytical and preparative reverse-phase HPLC on columns packed with C_{18}-bonded silica and were characterized by amino acid compositional analysis and fast atom bombardment mass spectrometry.

Expression and Purification of Recombinant GalNAc Transferases. Recombinant GalNAc transferases (*12*) were produced in a baculovirus expression system and purified as previously described. Enzyme assays were as previously described (*12*).

^{1}H NMR Spectroscopy. All 2D NMR experiments were performed on a Varian Unity 500 NMR spectrometer. Peptide concentrations were approximately 5 mM and were dissolved in the DMSO-d_6, solvent used previously in studies of mucin-related peptides (*21*). Sodium 3-(trimethylsilyl)-[2,2,3,3-$^2H_4$]propionate was used as an internal reference at 0.00 ppm. Probe temperature was regulated at 30 ± 0.2 °C, and samples were not spinning. No line broadening or long-range ROE information was observed that would indicate peptide aggregation in these samples. All spectra were collected in phase-sensitive mode, except COSY, which was acquired in magnitude mode. Spectral widths were 6000 Hz in *F*1 and *F*2 with 2K complex points collected in *F*2 and 512 complex points collected in *F*1. Data were zero-filled once in *F*1 and apodized with Gaussian functions in *F*1 and *F*2, except for COSY data, which were apodized with sine bells. TOCSY spectra were acquired with an 80 ms spin locking time, and ROESY spectra were acquired with 100 and 200 ms mixing times. All initial processing of 2D data was done with VNMR (Varian, Inc.). Two-dimensional data sets were imported into NMRCompass (MSI, Inc.) for assignment and interpretation. Assignments for the ^{1}H NMR resonances for each of the peptides were made according to the standard procedures (*22*). $^3J_{N\alpha}$ values

were obtained from DQF-COSY spectra (*23*). Temperature coefficients for NH ^{1}H movement were obtained from 2D COSY spectra acquired from 20 °C to 40 °C in 5 °C intervals ($r > 0.99$).

Structure Determination Protocol. Three-dimensional structures of the peptides were generated from the NMR data set (NOE distance constraints and $^3J_{N\alpha}$ coupling constants) using the following protocol: (i) determination of allowed ranges for the ϕ, ψ, and χ_1 angles that are consistent with a given set of NMR data; (ii) refinement of the ϕ, ψ, and χ_1 angle ranges and stereospecific assignments of the β-methylene protons; (iii) generation of 3D structures consistent with the NMR data; (iv) energy minimization of NMR-matched structures and selection of low-energy structures; and (v) analysis of the selected structures.

On the first two steps of the protocol, the COMBINE procedure (*24*) was used to determine and refine the ranges for the ϕ, ψ, and χ_1 angles for each residue in the peptide and to make stereospecific assignments of the β-methylene protons. The ranges for the ϕ, ψ, and χ_1 angles were determined by the computer program, FiSiNOE-3 (*25*). The input data for FiSiNOE-3 included the set of upper limits for the intraresidue ($d_{N\beta}$, $d_{\alpha\beta}$) and sequential (d_{NN}, $d_{\alpha N}$, $d_{\beta N}$) distance constraints as well as $^3J_{N\alpha}$ coupling constants. The mathematical expectations, standard deviations, and ranges allowed for the ϕ, ψ, and χ_1 angles were obtained as output data from FiSiNOE-3. The range allowed for each torsion angle was considered as a confidence interval, determined as a product of the half-width (3.0) and the standard deviation of the angle from its mathematical expectation (i.e., statistical 3σ criterion was used for selection). To refine the ranges for the ϕ, ψ, and χ_1 angles and to make stereospecific assignments of the β-methylene protons, the HABAS program (*26*) was employed. HABAS used the allowed torsion angle ranges estimated by FiSiNOE-3, the intraresidue and sequential NOE distance constraints, and the coupling constants as input data.

The intervals allowed for torsion angles and the stereospecific assignments determined by the COMBINE procedure, in conjunction with the distance constraints, were then used as input data for the DIANA program (*27*) to generate 3D structures consistent with the NMR data. The standard selection of minimization levels and parameters for the DIANA program with REDAC strategy (*28*), within the SYBYL 6.3 software package (Tripos Associates, Inc., St. Louis, MO), were used to generate 50 structures consistent with the NMR data set.

All structures determined by the DIANA program were energy-minimized with SYBYL using the Powell method with a maximum of 5000 minimization cycles. Calculations were performed using Kollman "all-atom" force field and Kollman charges with distance-dependent dielectric constant equal to 4*r*.

Tightness of structures within each structural family was evaluated by RMSD between coordinates of the backbone heavy atoms of the structures within the set compared to the corresponding average structure. The average structure was calculated by superimposing and averaging the corresponding atomic coordinates of each structure within the structural family.

The energy-minimized, NMR-matched structures were divided into groups (clusters) according to their overall shape. Starting with the lowest energy structure as a reference for the first group, all other structures were compared with it using the RMSD of all non-hydrogen atoms. The value of the RMSD $\leq$ 1.0 Å was taken as a criterion of a structural similarity. From structures that were not included in the first group, the lowest energy conformation was taken as reference for a second family. The procedure was repeated until all structures were assigned to groups. Finally, the lowest energy structures from each of the groups were analyzed for possible hydrogen bonds. The presence of a hydrogen bond was assumed if a distance between a proton and an oxygen was less than 2.1 Å and if an angle formed by the NH group with oxygen was greater than 125°.

Table 1: Kinetic Parameters for Catalysis of Glycosylation of Peptide Acceptor Substrates by Purified Recombinant Polypeptide GalNAc Transferases

	GalNAc-T1		GalNAc-T2		GalNAc-T3	
acceptor peptide sequence	K_m (mM)	V_{max} (pmol/min)	K_m (mM)	V_{max} (pmol/min)	K_m (mM)	V_{max} (pmol/min)
I - GVTSAPDTR	0.05	0.33	ND[a]	ND	0.005	0.097
II - GVTSAGDTR	ND	ND	ND	ND	ND	ND
III - PDTRPAPGS	ND	ND	ND	ND	ND	ND

[a] ND, no activity detected. A positive control acceptor peptide (MUC2, see ref *12*) that is an efficient substrate for all three of these enzymes was used to verify enzyme activity.

RESULTS

Enzyme Kinetics. Recombinant GalNAc transferase enzymes were purified (*12*) and employed to derive kinetic values for selected mucin peptide substrates. Three nine-residue peptides were tested for GalNAc transferase activity: the acceptor substrate with the wild-type sequence, GVTSAPDTR (I); the analogous peptide GVTSAGDTR (II), with substitution of G for P at position 6; and the wild-type peptide PDTRPAPGS (III), which includes a threonine that is not glycosylated in vitro by recombinant GalNAc-T1, -T2, or -T3. Table 1 presents kinetic data derived from plots of 1/V vs 1/[S] for selected peptide substrates with the three recombinant enzymes: GalNAc-T1, GalNAc-T2, and GalNAc-T3. Values of K_m and V_{max} obtained with the peptide GVTSAPDTR indicate that this is an efficient substrate for GalNAc-T1 and -T3; however, this nine-residue peptide is not an effective substrate for GalNAc-T2. It is notable that the K_m values for these short (nine-residue) peptide acceptors are approximately 10-fold lower than those for slightly longer peptides (with two additional amino acids at the amino terminus, reported in ref *12*) for both GalNAc-T1 and GalNAc-T3. The lack of catalysis by GalNAc-T2 suggests that this enzyme requires additional sequence at the amino terminus to catalyze glycosylation of the substrate. Analysis of the reactive products by mass spectrometry confirmed that the GVTSAPDTR peptide was glycosylated on only one position (data not shown): a single glycosylated threonine was observed at GVTSA for both GalNAc-T1 and GalNAc-T3, as determined by MS/MS analysis of the fragmentation pattern produced following β-elimination of the glycosylated residue (data not shown). It is interesting that substitution of G for P at position 6 of peptide I (GVTSA**G**DTR) significantly reduced the ability of this peptide to serve as a substrate of these enzymes, consistent with previous reports

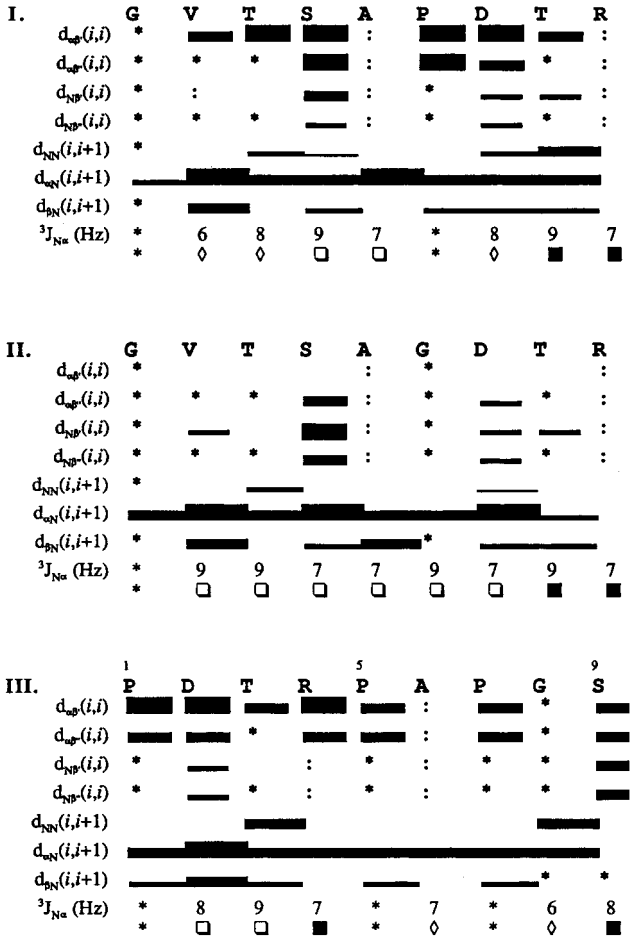

FIGURE 1: Summary of NMR data observed. Schematic diagram shows the magnitude of various NOE connectivities (strong, medium, weak, and very weak). An empty space indicates that the intensity of the corresponding cross-peak is undistinguishable from background. The single and double primes are used to differentiate two β protons without stereospecific assignments. Asterisks indicate lack of one proton. The colon indicates overlapping cross-peaks that were not resolved. In case of proline, NH refers to δH. The values of the $^3J_{N\alpha}$ coupling constants are expressed in round numbers. The slow (<0.003 ppm/K) and medium rates (in the interval between 0.003 and 0.005 ppm/K) of the amide proton movements are represented by closed (■) and open (□) squares. The rhombus (◇) denotes the fast rate (>0.005 ppm/K) of the amide proton movements.

(*8*). This observation prompted us to investigate the structure of the GVTSAPDTR peptide and the analogous GVTSAGDTR peptide. NMR spectroscopy and molecular modeling techniques were used to deduce structural features that may contribute to the relative effectiveness of these peptides as substrates. Since a threonine residue within sequence PDTR was not glycosylated, the structure of the nine-residue peptide PDTRPAPGS containing this sequence at a different relative position was also investigated.

NMR Conformational Analysis and Molecular Modeling. The effective substrate I, the poorly glycosylated substrate II, and the nonglycosylated peptide III were analyzed by NMR. Semiquantitative estimates of intensities for sequential and intraresidue cross-peaks, values of vicinal coupling constants $^3J_{\alpha N}$, and thermal protection measurements for amide protons of these peptides are summarized in Figure 1. These data were processed by the COMBINE procedure (*24*) and the DIANA program (*27*) to determine sets of 50 structures consistent with NMR-derived constraints.

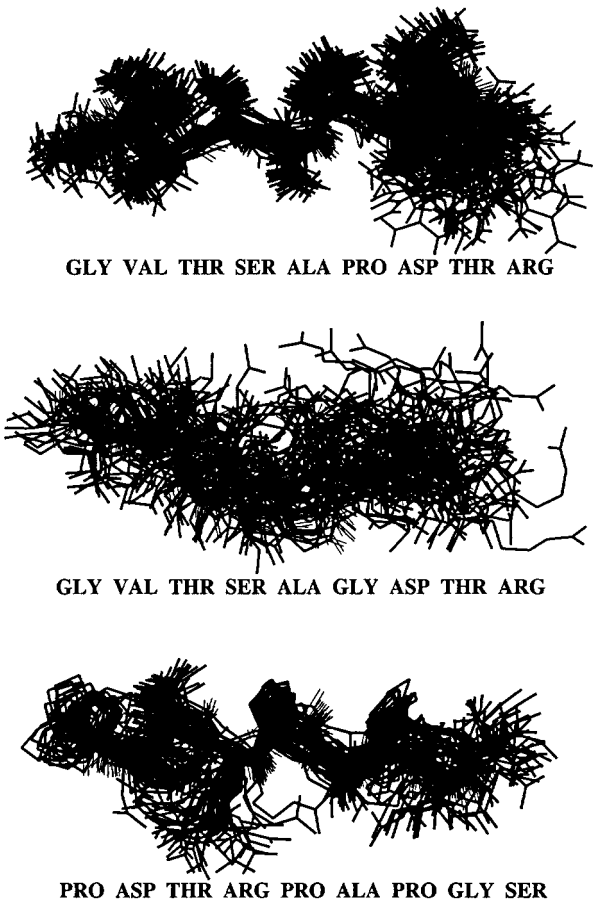

FIGURE 2: Structural families of the peptides I, II, and III consistent with the NMR data. Superimpositions were made by the backbone heavy atoms of the peptides.

After energy minimization, structural families of the effective substrate I, poorly glycosylated substrate II, and nonglycosylated peptide III were evaluated by tightness of the structures and by overall shape of the average structures. The low-energy (less than 15 kcal/mol), NMR-matched structures within each structural family were divided into clusters to define distinct structural features.

A structural family for substrate I (GVTSAPDTR) (Figure 2) was well-defined and showed highly consistent structures among members. The two other structural families (GVTSAGDTR and PDTRPAPGS) were less consistent (Figure 2). The mean pairwise RMSD for the backbone heavy atoms of the structures and the corresponding average structure for substrate I was 0.79 ± 0.30 Å. The mean pairwise RMSD values for substrates II and III were 2.38 ± 0.67 and 1.24 ± 0.41 Å, respectively. The GVTSAPDTR peptide has two distinct structural segments. The five residues at positions g − 1, g, g + 1, g + 2, and g + 3 have a high propensity for extended backbone conformations. (Position g refers to the residue to be glycosylated; − and + refers to residues that are N-terminal or C-terminal to a reactive hydroxyamino acid, respectively.) Chemical shifts of amide protons in this fragment, especially Thr3, are sensitive to temperature in a manner that is consistent with exposure of the protons to solution. The rate of amide proton movements for Ser4 and Ala5 are slightly lower, which may indicate partial involvement in intramolecular hydrogen bonding. The most populated structural cluster of GVTSAPDTR exhibited extended, β-like conformations for the residues Val2, Thr3, and Ser4,

with the backbone NH groups fully exposed to solvent. A second, slightly less populated structural cluster contained an inverse γ-turn conformation. This γ-turn forms a bump-like bulge centered on Thr3 with a backbone conformation of $\phi = -83°$ and $\psi = 66°$ and with flanking residues in an extended conformation. The distance between the CO group of Val2 and the NH group of Ser4 is close enough for H-bond formation.

In contrast, the substrate fragment PDTR, which includes residues at positions g + 3, g + 4, g + 5, and g + 6, adopts a much more compact structure. The relative insensitivity to temperature of chemical shifts for the amide proton of Arg9 (within the interval 0.0004−0.0007 ppm/K) is consistent with hydrogen-bond formation or inaccessibility to solvent. Likewise, the guanidinium group of Arg9 has rates of proton exchange lower than 0.0016 ppm/K. The low-field positions of the chemical shift for these protons combined with low rates of exchange are consistent with the hypothesis that they are involved in intramolecular hydrogen bonding. The low rate of amide proton movement for Thr8 (0.0023 ppm/K) may indicate partial involvement in hydrogen bonding or inaccessibility to solvent molecules. The set of short-range cross-peak intensities, including a cross-peak between the proton of Asp7 and the amide proton of Arg9, and values of the $^3J_{N\alpha}$ coupling constants of the fragment Pro6−Arg9 suggest the existence of a compact, turn-like conformation. Representative structures of the two most populated clusters both demonstrated close proximity of the CO groups of Pro6 and Asp7 with NH groups of Thr8 and Arg9, respectively, with a high probability of hydrogen bonding. The guanidinium group of Arg9 could be involved in hydrogen bonding with either the CO group of Pro6 or the carboxyl group of the side chain of Asp7.

The NMR patterns and conformational features of the P(G)DTR fragment in substrates I, II, and III are very similar, suggesting that the P(G)DTR fragment forms internally stabilized structures. The distances between C^α atoms of Pro (Gly in II) and Arg residues for the average structure of families I, II, and III were equal to 8.1, 8.4, and 8.6 Å. The mean pairwise RMSD values for the backbone heavy atoms calculated for the fragment PDTR within structural families I and III were equal to 0.31 ± 0.08 and 0.36 ± 0.11, respectively. Thus, the conformational preferences of the PDTR fragment are independent of its relative position within the substrate peptide, strongly suggesting the existence of specific interactions that provide internal stability for this region of the peptide.

Substitution Gly for Pro6 in the poorly glycosylated peptide GVTSAGDTR affected the conformation and flexibility of the N-terminal fragment. The Thr3 residue of the G-substituted peptide does not exhibit a distinct propensity to adopt either an elongated or twisted backbone conformation, with the amide proton showing intermediate sensitivity to temperature. The conformational propensities within structural families I and II can be characterized by distances between the N- and C-termini of the average structures. For the effectively glycosylated substrate I, the distance between N- and C-termini was equal to 23.1 Å, whereas the distance was only 18.0 Å for the poorly glycosylated substrate II. This reflects the more extended structure of the well-glycosylated peptide GVTSAPDTR. The mean pairwise RMSD values for the backbone heavy atoms calculated for the fragment GVTSAP(G) within structural families I and II were equal to 0.59 ± 0.24 and 0.87 ± 0.17 Å, respectively. This indicates that the fragment GVTSAP within structural family I is better stabilized internally than GVTSAG in peptide II. In contrast to peptide I, none of the structural clusters for peptide II demonstrated fully extended β-like conformations for Thr3 and the flanking residues. Instead, the three largest clusters of peptide II revealed a mix of β, inverse γ-turn, and polyproline II-like conformations for Thr3 and the flanking residues. The structural motif consisting of an inverse γ-turn for the reactive threonine with flanking residues in β-conformations, as observed in the second cluster of peptide I, was assigned to only 10% of the structures within structural family II.

DISCUSSION

Comparative analysis of two differentially glycosylated peptide substrates leads us to propose that structural features of the acceptor substrate GVTSAPDTR enhance its ability to be glycosylated through the catalytic activity of GalNAc-T1 and GalNAc-T3. There was a lack of reactivity of GalNAc-T2 with the short peptide substrates used in these experiments, even though this enzyme is capable of glycosylating this site on longer peptide substrates (*12*). This confirms a previous hypothesis that the molecular recognition of MUC1 tandem repeats by GalNAc-T2 is distinct from that by GalNAc-T1 and GalNAc-T3.

NMR data showed the presence of two distinct structural motifs within GVTSAPDTR. One motif contains five residues at position g − 1, g, g + 1, g + 2, and g + 3 and is characterized by an extended backbone conformation. In contrast, the four residues at positions g + 3, g + 4, g + 5, and g + 6 form a relatively compact substructure stabilized by a set of intramolecular hydrogen bonds. No long-range interactions between these two structural regions were observed. NMR data showed that the reactive threonine residues had a marked propensity to assume extended backbone conformations, whereas the nonreactive threonines were more likely to exist in twisted backbone conformations.

The importance of an extended β-like conformation at an O-glycosylated residue has been recognized (*3*, *4*, *7*, *11*, *29*). Previously proposed models of peptide-transferase interaction posited that extended β-like substrate conformations are favorable for enzyme−substrate binding (*7*, *11*). Our results are consistent with this hypothesis. The most populated structural cluster of the peptide GVTSAPDTR exhibited extended β-like conformations for the residues Val2, Thr3, and Ser4. However, for the second cluster, we observed a structural motif comprised of an inverse γ-turn conformation for Thr3 with the flanking residues in an extended, β-strand-like conformation. Thus, the peptide backbone of GVTSAPDTR exhibited a β-like or an inverse γ-turn conformation of the residue to be glycosylated. The peptide backbone has a very similar shape in both conformations, although the threonine protrudes more in the inverse γ-turn than in the β-like conformation.

Our observation of predominantly extended backbone conformations at positions g − 1, g, g + 1, g + 2, g + 3, and g + 4 is consistent with the general finding that substitution of a proline residue at these positions enhances catalytic efficiency (*5*, *7*). We propose that an extended

backbone conformation at and around the glycosylated residue facilitates interaction between some GalNAc transferases and the acceptor substrate by enhancing the probability (and the rate) of formation of intermolecular hydrogen bonds. All GalNAc transferases characterized to date display relatively broad activity for substrates in that they will glycosylate peptides with diverse amino acid sequences flanking the reactive residue (*7*). The lack of specific recognition of amino acid side chains by GalNAc transferases, together with the data presented here, leads us to propose that GalNAc-T1 and GalNAc-T3 exhibit molecular specificity for conformations assumed by the peptide backbone of different acceptor substrates. The acceptor substrate−enzyme interaction probably involves hydrogen bonding between the GalNAc transferase and some components of the extended backbone frame.

A comparison of the NMR data for the peptide substrates and enzyme kinetic data shows that the rates of exchange of the amide protons of the reactive threonines are correlated with the relative rates of glycosylation. The amide proton of the reactive Thr residue in GV<u>T</u>SAPDTR demonstrates a higher temperature coefficient and, thus, greater accessibility to solvent than the corresponding amide proton in the inefficient substrate GV<u>T</u>SAGDTR. Thus, in the effective substrate, this proton is exposed and may serve as a donor for hydrogen bonding with the enzyme. In contrast, the amide protons of threonines that are not efficiently glycosylated have relatively low temperature coefficients, suggesting these protons are involved in intramolecular hydrogen bonding that adversely affects their potential for binding enzyme to facilitate catalysis of the glycosylation reaction.

The finding that prolines near the reactive residue are favorable for substrate efficacy supports a previously stated hypothesis that backbones of effective substrates accept hydrogen bonds from GalNAc transferases at the carbonyl oxygen (*7*) rather than donating bonds at the amide proton as we propose here. The hypothesis that there is hydrogen bonding between enzyme and substrate carbonyl oxygen is plausible and not mutually exclusive with the hypothesis that there is interaction between amide proton and enzyme. Regardless of these hypothesized intramolecular interactions, one major effect of the occurrence of proline at positions g − 1, g + 1, g + 2, and, particularly, at g + 3 is a significant increase in the propensity of the peptide backbone to adopt an extended conformation. A second effect that could enhance substrate efficacy is that prolines may restrict local flexibility of the peptide backbone to a conformation that is suitable for glycosylation.

The NMR data (including short-range cross-peak intensities and coupling constants) for different peptides tested here show that the PDTR fragment, which is a tumor-associated epitope recognized by a number of monoclonal antibodies (*17*, *20*), has similar structural properties whether it is placed at the C- or N-terminus of short peptides. Presumably, this substructure is stabilized by a set of hydrogen bonds formed by the backbone and side-chain groups of the PDTR residues. The relative stability of this short peptide form suggests that it may represent a structural component that is a self-stabilized building block of the mature structure assumed by the MUC1 protein. The substructure of the PDTR region of the peptide exists in a predominantly twisted conformation and is not glycosylated by GalNAc-T1, -T2, or -T3 (Figure 2). There is recent evidence that the PDTR site is glycosylated by other polypeptide GalNAc transferases. In a Herculean effort, Müller et al. (*19*) showed that this site is glycosylated in a fraction of MUC1 protein purified from normal human milk. It has recently shown that the newly described GalNAc-T4 (*13*) catalyzes glycosylation of the threonine in PDTR (Bennet, Hassan, and Clausen, unpublished data). These data lead us to propose that some polypeptide GalNAc transferases catalyze glycosylation of peptide substrates with twisted backbone conformations, whereas GalNAc-T1 and -T3 are more effective at glycosylating substrates with extended backbone conformations. These differences in propensity to catalyze glycosylation of peptides with different backbone structures may explain in part the surprisingly large number of polypeptide GalNAc transferases with distinct and overlapping substrate specificities (*13*).

Previous investigations (*4*, *5*, *8*−*10*) showed a correlation between substrate efficacy and the length of the fragments flanking the reactive threonine residue. Increased enzyme activity with peptide substrates was seen when at least five to six residues were adjacent to both sides of the reactive threonine, as compared to shorter peptides. It was suggested that the binding site of the GalNAc transferase recognizes extended substructures in these adjoining residues.

The peptide substrates based on the MUC1 tandem repeat used in these studies were designed to investigate some aspects of substrate conformation that relate to the minimal requirements for O-glycosylation of a particular site on this tandem repeat. These findings may be applicable to other substrates with similar structural character. Undoubtedly, glycosylation is also influenced by other structural parameters conferred by substrates of different primary sequence and higher order structures that can be assumed by native protein substrates.

On the basis of the comparative analysis of NMR data, enzyme kinetic data, and molecular modeling for well-glycosylated and nonglycosylated analogues, we propose here a structural model to explain the O-glycosylation of the GVTSAPDTR substrate by GalNAc-T1 and -T3. We hypothesize that most GalNAc transferases interacting with peptide substrates require at least three points of contact to stabilize the spatial orientation of the hydroxyl group that is the target of catalysis in the O-glycosylation reaction. Presumably, one of these sites is the hydroxyl group that is to be modified. The active site of GalNAc-T1 and -T3 probably recognizes an extended substructure on the GVTSAPDTR peptide backbone that may include a distinct structural bulge formed by the reactive threonine. We propose that for GalNAc-T1 and GalNAc-T3, a second site is located on the backbone somewhere in vicinity of g + 3, g + 4, g + 5, and g + 6 positions. Another site could be located on the backbone from position g − 2 to g − 6. In the process of interacting with the GalNAc transferase, the main chain of the central part of the substrate is predicted to form a set of hydrogen bonds with the enzyme. The amide proton of the reactive residue may serve as an anchor position for substrate−enzyme interaction by forming a hydrogen bond with the active site of the enzyme. The oxygens of the reactive residue and several flanking residues may also be involved in hydrogen bonding with the GalNAc transferase. This binding site would be predicted to extend from

the catalytic site to approximately 9 Å on both sides. This hypothetical model will be tested in future experiments.

ACKNOWLEDGMENT

We gratefully acknowledge Dr. W. Gmeiner for collecting and processing the NMR data and Drs. N. Kaufman and S. Sanderson for intellectual contributions to this work.

REFERENCES

1. Rademacher, T. W., Parekh, R. B., and Dwek, R. A. (1988) *Annu. Rev. Biochem. 57*, 785−838.
2. Strouss, G. J., and Decker, J. (1992) *Crit. Rev. Biochem. Mol. Biol. 27*, 57−92.
3. Wilson, I. B. H., Gavel, Y., and von Heijne, G. (1991) *Biochem. J. 275*, 529−534.
4. O'Connell, B., Tabak, L. A., and Ramasubbu, N. (1991) *Biochem. Biophys. Res. Commun. 180*, 1024−1030.
5. O'Connell, B., Hagen, F. K., and Tabak, L. A. (1992) *J. Biol. Chem. 267*, 25010−25018.
6. Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L., and Kezdy, F. J. (1991) *J. Biol. Chem. 266*, 14554−14561.
7. Elhammer, A. P., Poorman, R. A., Brown, E., Maggiora, L. L., Hoogerheide, J. G., and Kezdy, F. J. (1993) *J. Biol. Chem. 268*, 10029−10038.
8. Nishimori, I., Johnson, N. R., Sanderson, S. D., Perini, F., Mountjoy, K., Cerny, R. L., Gross, M. L., and Hollingsworth, M. A. (1994) *J. Biol. Chem. 269*, 16123−16130.
9. Nishimori, N., Perini, F., Mountjoy, K. P., Sanderson, S. D., Johnson, N. R., Cerny, R., Gross, M. L., Fontenot, D. R., and Hollingsworth, M. A. (1994) *Cancer Res. 54*, 3738−3744.
10. Nehrke, K., Hagen, F. K., and Tabak, L. A. (1996) *J. Biol. Chem. 271*, 7061−7065.
11. Gerken, T. A., Owens, C. L., and Pasumarthy, M. (1997) *J. Biol. Chem. 272*, 9709−9719.
12. Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennet, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) *J. Biol. Chem. 272*, 23503−24202.
13. Clausen, H., and Bennet, E. (1996) *Glycobiology 6*, 635−646.
14. Devine, P. L., and McKenzie, I. F. C. (1992) *BioEssays 14*, 619−625.
15. Girling, A., Bartkova, J., Burchell, J., Gendler, S., Gillet, C., and Taylor-Papadimitriou, J. (1989) *Int. J. Cancer 43*, 1072−1076.
16. Barnd, D. L., Lan, M., Metzger, R., and Finn, O. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7159−7163.
17. Gendler, S. J., Spicer, A. P., Lalani, E. N., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Boshell, M., and Taylor-Papadimitriou, J. (1991) *Am. Rev. Respir. Dis. 144*, S42−S47.
18. Sutherlin, M. E., Nishimori, I., Caffrey, T., Bennett, E. P., Hassan, H. H., Mandel, U., Mack, D., Iwamura, T., Clausen, H., and Hollingsworth, M. A. (1997) *Cancer Res. 57*, 4744−4748.
19. Müller, S., Goletz, S., Packer, N., Gooley, A., Lawson, A. M., and Hanisch, F.-G. (1997) *J. Biol. Chem. 272*, 24780−24793.
20. Karsten, U., Diotel C., Klich, G., Paulsen, H., Goletz, S., Müller, S., and Hanisch, F.-G. (1997) *Cancer Res. 58*, 2541−2549.
21. Scanlon, M. J., Morley, S. D., Jackson, D. E., Price, M. R., and Tendler, S. J. B. (1992) *Biochem. J. 284*, 137−144.
22. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*. John Wiley and Sons, New York.
23. Kim, Y., and Prestegard, J. H. (1989) *J. Magn. Reson. 84*, 9−13.
24. Sherman, S., Sclove, S., Kirnarsky, L., Tomchin, I., and Shats, O. (1996) *J. Mol. Struct.: THEOCHEM 368*, 53−162.
25. Shats, O., and Sherman, S. (1996) *Third Electronic Computational Chemistry Conference (ECCC-3)* (URL http://www.unmc.edu/Eppley/ECCC3/fisinoe3.htm).
26. Güntert, P., and Wüthrich, K. (1989) *J. Am. Chem. Soc. 111*, 3997−4004.
27. Güntert, P., Braun, W., and Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.
28. Güntert, P., and Wüthrich, K. (1991) *J. Biomol. NMR 1*, 447−456.
29. Aubert, J.-P., Biserte, G., and Loucheux-Lefebvre, M.-H. (1976) *Arch. Biochem. Biophys. 175*, 410−418.

BI981034A