Glycoconjugate Journal 15, 457–467 (1998)


# Structural features of the human salivary mucin, MUC7#

Tarikere L. Gururaja, Narayanan Ramasubbu, Paloth Venugopalan, Molakala S. Reddy, Kalaiyarasi Ramalingam and Michael J. Levine*

*Department of Oral Biology and Research Center in Oral Biology, 109 Foster Hall, School of Dental Medicine, State University of New York at Buffalo, Buffalo, NY 14214, USA*

**Human salivary mucin (MUC7) is characterized by a single polypeptide chain of 357 aa. Detailed analysis of the derived MUC7 peptide sequence reveals five distinct regions or domains: (1) an N-terminal basic, histatin-like domain which has a leucine-zipper segment, (2) a moderately glycosylated domain, (3) six heavily glycosylated tandem repeats each consisting of 23 aa, (4) another heavily glycosylated MUC1- and MUC2-like domain, and (5) a C-terminal leucine-zipper segment. Chemical analysis and semi-empirical prediction algorithms for O-glycosylation suggested that 86/105 (83%) Ser/Thr residues were O-glycosylated with the majority located in the tandem repeats. The high ( $\sim$ 25%) proline content of MUC7 including 19 diproline segments suggested the presence of polyproline type structures. CD studies of natural and synthetic diproline-rich peptides and glycopeptides indicated that polyproline type structures do play a significant role in the conformational dynamics of MUC7. In addition, crystal structure analysis of a synthetic diproline segment (Boc-Ala-Pro-OBzl) revealed a polyproline type II extended structure. Collectively, the data indicate that the polyproline type II structure, dispersed throughout the tandem repeats, may impart a stiffening of the backbone and could act in consort with the glycosylated segments to keep MUC7 in a semi-rigid, rod shaped conformation resembling a 'bottle-brush' model.**

***Keywords*: Salivary mucin (MUC7), O-glycosylation, APP segments, tandem repeat, (glyco)peptide synthesis, crystallization, X-ray diffraction, poly-L-proline type conformation**

***Abbreviations*: aa, amino acid; MUC7, human salivary mucin; APP, Ala-Pro-Pro, Boc-Ala-Pro-Pro-OBzl; Boc, $N^{\alpha}$-t-butyloxycarbonyl; Bzl, benzyl; CD, circular dichroism; e.s.ds, estimated standard deviations; Fmoc, $N^{\alpha}$-fluorenylmethoxycarbonyl; Fuc, fucose; Gal, galactose; GalNAc, N-acetylgalactosamine; MALDI, matrix assisted laser desorption ionization; PPI, poly-L-proline type I conformation; PPII, poly-L-proline type II conformation; NeuAc, sialic acid; SPPS, solid-phase peptide synthesis; SSCP, secondary structure content prediction**

## Introduction

Human saliva contains a number of glycoproteins including high and low molecular weight mucins [1–4]. The molecular weight mucin, now designated MUC7, can occur as two isoforms that differ in their content of terminal sialic acid and fucose residues [5]. MUC7 consists of a single polypeptide chain having 357 amino acids primarily made up of Ser, Thr, Pro and Ala [6, 7]. The major O-linked carbohydrate units of MUC7 have been structurally defined and range from two to seven residues [8]. Approximately, 80% of the oligosaccharides are comprised of Gal$\beta$l,3GalNAc, Fuc$\alpha$1,2Gal$\beta$1,3GalNAc,NeuAc$\alpha$2,3Gal$\beta$l,3GalNAc and are found to be conjugated to a serine and/or threonine residue via an $\alpha$-O glycosidic linkage involving GalNAc. While the structures of the major O-linked oligosaccharide chains are well characterized, the location and distribution of these oligosaccharides along the peptide chain of MUC7 are unknown. Although only MUC1, MUC2 and MUC7 human mucin cDNAs have been fully sequenced, data obtained from all the mucins indicate that they share certain common structural characteristics [9, 10]. For example, a predominant feature of most mucins is a central region consisting of repetitive peptide sequences or tandem repeats that are flanked on either side by non-repetitive domains. The tandem repeats of each mucin gene differ in their amino acid sequence and length; however they contain a high content of threonine/serine as potential O-glycosylation sites [9, 10].

There are a number of predictive methods now available for estimating the relative propensity for a given Ser or Thr

*To whom correspondence should be addressed. Tel.: (716)829-2114; Fax: (716)829-3942; E-mail: mj_levine@sdm.buffalo.edu

#Portions of this work were presented at the 4th Jenner International Glycoimmunology meeting and as an abstract published in *Glycoconjugate J* (1996) **13**: 883.

to be glycosylated [11–18]. Based on the O-glycosylation probability factor $h$, as described by Elhammer *et al.* [11], most of the mucins are heavily glycosylated in the tandem repeat region when $h$ is considered to be $> 0.75$. Investigations based on statistical analyses to define an O-glycosylation consensus signal have suggested that the flanking sequences around an O-glycosylation site influence the state of glycosylation of serine and threonine residues [15–18]. The presence of a Pro, Ala, Ser or Thr at positions $+3$, $-1$, $-3$ and $-6$ (with the glycosylated residue being 0) was often associated with glycosylation of Ser or Thr. Since the tandem repeats of MUC7 are comprised primarily of Pro, Ala, Ser, and Thr, chemical analyses and semi-empirical prediction algorithms were used to identify the potential O-glycosylation sites of MUC7.

Earlier CD studies suggested that the peptide backbone of MUC7 strongly influenced the mucin's conformation which most likely occurred as an extended structure [3]. In the present study, CD analyses of MUC7 and several MUC7-derived peptides and glycopeptides indicated the presence of a poly-L-proline conformation. Poly-L-proline conformation is an extended structure with a characteristic three-amino acid repeat and occurs as either type I (PPI) containing *cis* peptide bonds or type II (PPII) containing *trans* peptide bonds. The induction of PPII conformation in a protein requires at least two successive proline residues or proline residues regularly repeated along a sequence [19–22]. In MUC7, there are 19 diproline sequences mostly located in the tandem repeat domain suggesting that this mucin is prone to adopt a polyproline conformation. An earlier study on the conformation of Boc-Pro-Pro-OH [23] revealed that this synthetic dipeptide derivative adopted both PPI and PPII conformations and suggested that diprolines may be too short to be used as model peptides for determining the conformational preferences of MUC7. Since the majority of diproline segments in the MUC7 tandem repeats were comprised of Ala-Pro-Pro (APP), an APP-derivative, Boc-Ala-Pro-Pro-OBzl, was synthesized and its polyproline, type characteristics was determined by X-ray crystallographic analysis.

## Materials and methods

### Materials

Protected Fmoc-amino acid derivatives, Fmoc-Glu(OtBu)-Wang-Resin and peptide synthesis reagents were purchased from Sigma (St. Louis, MO) unless otherwise specified. Boc-Ala-Pro-OH and Pro-OBzl. HCl were purchased from Bachem Bioscience (Philadelphia, PA). Analytical thin layer chromatography (TLC) was performed on Merck Silica Gel 60 $F_{254}$ plates whereas column chromatography was performed on Silica Gel (60–200 mesh, JT Baker) using ethylacetate and hexane as solvent systems. All the reagents were of analytical grade and were used without further purification.

### Peptide synthesis

*(i) Synthesis of Boc-Ala-Pro-Pro-OBzl (APP)*

Boc-Ala-Pro-Pro-OBzl was synthesized by standard solution-phase procedures using EDAC-HOBT mediated coupling [24]. Briefly, Boc-Ala-Pro-OH (286 mg, 1 mmol) and HOBT (135 mg, 1 mmol) were dissolved in dichloromethane (5 ml) and the reaction mixture was cooled to 0 °C. EDAC (191 mg, 1 mmol) was added to the reaction mixture and was solubilized by stirring. To the clear solution, Pro-OBzl. HCl (241 mg, 1 mmol) was added and the pH of the reaction mixture was adjusted to 8 using *N*,*N*-diisopropylethylamine. The reaction mixture was stirred for 3 h at room temperature, diluted with ethylacetate and the organic layer was washed extensively with 2 N HCl followed by 1 M $Na_2CO_3$ and brine. The ethylacetate layer was then dried and concentrated to yield the tripeptide as a white power (410 mg, yield: 90%).

*(ii) Solid-phase synthesis of APP-containing peptides and glycopeptides*

All the MUC7-derived APP-containing (glyco)peptides (**I–IV**, Table 1) were synthesized on an automated Beckman System 990 synthesizer following standard procedures of Fmoc chemistry using appropriate pre-loaded Fmoc-aa resins [25, 26]. In the case of glycopeptides (**III**) and (**IV**) having the sequence PAPPSSS*APPE where * denotes glycosylated amino acids having α-D-GalNAc as well as Galβ(1-3)-GalNAc, synthesis was carried out by coupling the preformed building blocks [27] $N^{\alpha}$-Fmoc-Ser-[$Ac_3$-α-D-$GalN_3$]-OPfp and $N^{\alpha}$-Fmoc-Ser-[$Ac_4$-β-D-Gal-(1-3)-$Ac_2$-α-D-$GalN_3$]-OPfp, respectively at the desired position in the peptide sequence as described before [28].

### Purification and characterization of synthetic (glyco)peptides

High performance liquid chromatography (HPLC) was carried out using a Beckman 344 gradient liquid chromatograph

**Table 1.** MUC7 (glyco)peptides[#]

| *(Glyco) peptides* | *Amino acid sequence* | *No. of APPs* | *% Pro* |
|---|---|---|---|
| **I** | TTAAPPTPPATTPAPPSSSAPPE | 3 | 39 |
| **II** | PAPPSSSAPPE | 2 | 46 |
| **III** | PAPPSSS*APPE | 2 | 46 |
| **IV** | PAPPSSS**APPE | 2 | 46 |
| **T1** | residues 142–345 (see Table 2) | 14 | 24 |
| **T2** | residues 70–141 (see Table 2) | 0 | 22 |

[#]Fragments **I–IV** are derived from the MUC7 tandem repeat sequence (see Table 2).
S* denotes Ser glycosylated with **α**-D-GalNAc whereas S** denotes Ser glycosylated with Galβ(1-3) GalNAc. **T1** and **T2** correspond to tryptic peptides isolated and purified by trypsinolysis of MUC7 [6].

having a variable wavelength detector. All peptides and glycopeptides were purified by repetitive HPLC runs on a Rainin Dynamax-60 Å reversed-phase semi-preparative C18 column (10 × 250 mm) coupled to a guard column (10 × 50 mm) employing 0.1% TFA in water (buffer A) – 0.1% TFA in acetonitrile (buffer B) linear gradient elution (1.5 ml $min^{-1}$) mode with detection at 230 nm). (Glyco)peptides eluted from the column using the linear gradient [Buffer A : B, 100 : 0 − > 60 : 40 (35 min)] were collected, pooled and lyophilized. The integrity of these purified (glyco)peptides was also confirmed by N-terminal sequencing [28], amino acid composition [5] and mass spectral analysis [26].

## Identification of MUC7 O- and N-glycosylation sites

### *(i) Extent of O-glycosylation*

The probability factor (*h*) for O-glycosylation was calculated by a computer program based on the parameters described by Elhammer *et al.* [11]. Prediction of O-glycosylation sites in MUC7 was performed by the *NetOglyc* WWW server (http://www.cbs.dtu.uk/netOglyc/cbsnet-Oglyc.html) specially maintained for mucin-type glycoproteins as described by Hansen *et al.* [12]. These data were then compared with the number of Ser + Thr residues in MUC7 that were O-glycosylated. For this purpose, MUC7 was purified from human submandibular-sublingual saliva as previously described [5]. Since the saliva donor had B blood group activity, all of the GalNAc present in MUC7 is involved in the glycopeptide linkage to Ser and/or Thr and is therefore an indicator of the number of O-linked units. For determining the GalNAc content, purified MUC7 was dried over phosphorous pentoxide for 48 h in a high vacuum desiccator, weighed and then dissolved in distilled water. Aliquots of the MUC7 solution were hydrolyzed for 4, 6, 8 and 10 h respectively at 100 °C in 2 N HCl in a test tube sealed under nitrogen atmosphere and then analyzed on a Beckman System 6300 amino acid analyzer [5]. The maximum yield of GalNAc was obtained after 8 h hydrolysis. This value was then divided by the destruction factor for GalNAc (0.94), determined following the hydrolysis of authentic N-acetylgalactosamine (Sigma Chemical Co., St. Louis, MO) for 8 h at 100 °C in 2 N HCl. To determine Ser and Thr content, aliquots of the MUC7 solution were dried and then hydrolyzed in 6 N HCl at 100 °C for 24, 48 and 72 h, respectively, as described above. Following amino acid analysis, the actual content of Ser and Thr was extrapolated to zero time. The percentage of Ser and Thr that were O-glycosylated in MUC7 was calculated by dividing the GalNAc content with the Ser + Thr content.

### *(ii) Prediction of N-glycosylation sites and secondary structure*

For determining N-linked glycosylation sites in MUC7, the consensus triplet sequence Asn-Xaa-Ser/Thr, where Xaa can be any residue except proline was utilized as a motif [29, 30]. Secondary structure content prediction (SSCP) utilizing the amino acid sequence was performed by WWW (http://www.embl-heidelberg.de/argos/sscp_info.html) server created by Dr. Frank Eisenhaber (EMBL, Heidelberg, FRG).

## Conformational analysis

Circular dichroism (CD) spectra of natural and synthetic APP-rich peptides and glycopeptides were recorded on a JASCO J-600 spectropolarimeter (Jasco, Eaton, MD). The device was routinely calibrated with the ammonium salt of (+)-10-camphorsulfonic acid as reported earlier [26]. Spectra were recorded between 250 and 180 nm at 0.2 nm intervals with a time constant of 2 s at 25 °C. Data were collected from five separate scans and averaged using an IBM PS/2 computer interfaced to the spectropolarimeter. A cylindrical quartz cell of path length 0.1 cm was used for the spectral range with the sample concentration of ~ 0.05 mM as determined by amino acid analysis. Peptide solutions were made in 20 mM sodium phosphate buffer, pH 7.2. The molar ellipticity values are expressed in deg $cm^2$ $dmol^{-1}$, using a mean residue molecular weight of 128.1. CD plots were made using Microsoft Excel software package as described earlier [26].

## X-ray crystallography of APP

Crystals of APP, $M_r = 455.5$, were obtained by slow evaporation from an aqueous methanol solution. The crystals are monoclinic, space group $P2_1$, with unit cell parameters of $\mathbf{a} = 6.341(2)$, $\mathbf{b} = 40.395(6)$, $\mathbf{c} = 10.126(2)$ Å, $\beta = 90.9(1)°$, $V = 2593.1(10)$ Å$^3$ and $Z = 4$, $D_{cal} = 2.03$ g $cm^{-3}$, $\mu = 11.0$ $cm^{-1}$. Three dimensional intensity data (up to a Bragg angle of $\Theta = 65°$) were collected on an Enraf-Nonius CAD-4 diffractometer with a graphite monochromator using a crystal of dimensions 0.2 × 0.3 × 0.10 mm. Lattice parameters were obtained from a least-squares refinement utilizing the setting angles of 25 centered reflections in the range of $20° < \theta < 27°$. Data were collected at room temperature using $\omega - 2\theta$ scan technique to a maximum $2\theta$ of 130°. The scan rate varied from 1 to 3° per min and the scan width was calculated using the expression $(0.75 + 0.14 \tan\theta)°$. The horizontal aperture width varied from 2.4 to 3.5 mm and the vertical aperture width was set at 4.0 mm. The range of *hkl* was: *h*, 0 to 7; *k*, 0 to 42; and *l*, 0 to 12. A total of 4695 reflections were collected of which 3229 were unique and significant ($I > 4\sigma(I)$). A check on crystal and electronic stability was carried out by monitoring three reflections every hour during data collection. The plot of intensity versus time indicated a loss in intensity of 1.5%.

The structure was solved by direct methods using SHELXS-86 [31]. Sixty-eight non-hydrogen atoms were found from the best map. Refinement was done using full-matrix least-squares with non-hydrogen atoms refined anisotropically. The positions of most of the hydrogen atoms were found in subsequent difference Fourier maps; others were calculated from geometrical considerations. The thermal parameters of the hydrogen atoms were fixed and not refined and led to a weighted agreement factor of 6.0%. The least squares refinement was carried out using the program SHELXS-93, $(\Delta/\sigma)\max < 0.73$ for all atoms. The highest positive and negative peak in the difference electron density map were 0.42 and 0.30 $e\text{Å}^{-3}$ respectively at $R = 6.0\%$. Scattering factors were taken from *International Tables for X-Ray Crystallography* (1974) [32].

## Results and discussion

### MUC7 domains – analysis of the primary sequence

Analysis of the primary sequence of the MUC7 enable us to propose a new model for the structural and functional organization of MUC7 (Table 2). As shown, the secreted peptide core can be divided into 5 distinct domains.

*Domain 1*

This domain contains a histatin-like sequence ($aa^{23-37}$) similar in sequence (percent similarity is 53.33) to the small histidine-rich salivary peptides (e.g. histatins) that possess candidacidal activity [33]. Preliminary studies have shown that synthetic MUC7 peptides representing $aa^{23-37}$ also possess candidacidal activity albeit less than natural histatin 5 [34]. Domain 1 also contains a leucine zipper-like segment that consists of a Leu residue at i, i + 7 [35, 36]. A second leucine zipper sequence is present at the C-terminus of MUC7 (Table 2). Since leucine zippers are known to promote aggregation of monomers [35, 36], it is possible that MUC7 can undergo aggregation via its N- and C-terminal leucine zippers. Both of the half-cysteine residues are found in Domain 1.

*Domain 2*

This domain appears to be moderately glycosylated. It contains four of the five potential N-glycosylation sites and 18 potential Ser/Thr O-glycosylation sites. The tentative assignment of all the 18 potential Ser/Thr O-glycosylation sites and the presence of four potential N-glycosylation sites makes this domain a second richly glycosylated region of MUC7. This domain also retains an extended polyproline II structure due to the presence of prolines as suggested by CD experiments.

*Domain 3*

This domain contains the six 23 aa tandem repeats and is heavily glycosylated. There are 46 potential O-glycosylation sites and 14 APP segments. Three of the tandem repeats each contain 3 APP segments, two each contain 2 APP segments, and one contains 1 APP segment. This is the key domain which gives MUC7 a semi-rigid rod shaped conformation due to its polyproline type II conformation. It appears that the enriched APP-segments in this domain versus other domains make an overall contribution to the conformational rigidity of MUC7. It is noteworthy that Fontenot *et al.* [37] have recently found the importance of multiple proline-rich tandem repeats of the human MUC1 protein core in stabilizing the type II polyproline $\beta$-turn helix.

*Domain 4*

This domain contains 22 potential O-glycosylation sites and one N-glycosylation site. As compared to the sequence of MUC1 and MUC2, this domain has 27% sequence homology with mouse polymorphic epithelial mucin (MUC1) and 43% sequence homology with human intestinal mucin (MUC2) [38, 39]. Our rationale for comparing this domain to the sequences of MUC1 and MUC2 is as follows. Only MUC1, MUC2 and MUC7 human mucin cDNAs have been fully sequenced. Both MUC1 and MUC2 contains proline-rich tandem repeat sequences (MUC1, % proline = 25; MUC2, % proline = 22 and MUC7 % proline = 35). The high proline content in MUC1, MUC2 and MUC7 tandem repeats appears to be an important structural element in determining their conformational features. Interestingly, this domain showed almost 90% coil structure following SSCP analysis (Table 2).

*Domain 5*

The important characteristic feature of this domain is the presence of a leucine-zipper segment. Both of the methionine residues of secreted MUC7 are located in this domain. This domain appears to have a helical structure based on its secondary structure prediction (Table 2).

### O-glycosylation probability of Ser/Thr in MUC7

MUC7 was purified from the saliva of an individual with B blood group activity so that determination of the number of GalNAc residues per molecule would be an indicator of the number of O-linked units. Based on chemical analysis, approximately 82% or 86 of the 105 Ser + Thr residues were O-linked. Several prediction criteria are now available to determine potential O-glycosylation sites. *In vitro* studies on the influence of flanking residues on O-glycosylation suggest that the presence of Pro, Ala, Ser and Thr at positions close to the site of O-glycosylation was often associated with increased glycosylation [15–18]. Also, a semi-empirical analysis of several O-glycosylated proteins showed that neither strongly hydrophobic nor strongly hydrophilic residues may be present in the nonapeptide segment flanking the site of glycosylation (i.e., four residues preceding and four residues following the reactive Ser/Thr). It was also found that a Thr or Ser residue having a Pro three residues down the C-terminus was glycosylated [11].

**Table 2.** Classification of MUC7 into different domains based on its primary aa sequence.

| *MUC7 primary aa sequence* | *O-linked Ser/Thr*[c] | *Diproline segments* | *SSCP*[d] *(%)* |
|---|---|---|---|
| *Signal Peptide (1–20)*<br>MKTLPLFVCICALSACFSFS | | | |
| *Domain-1 (21–71)*<br>EGRERDHELRHRRHHHQ[a]SPKSH<br>FELPHYPGLL[b]AHQKPFIRKSYKC<br>LHKR#CR | 0 | 0 | h = 55%<br>e = 00%<br>c = 45% |
| *Domain-2 (72–164)*<br>PKLPP**S**PNKPPKFPNPHQPPKHP<br>DKN*SSVVNP**T**LVA**TT**QIP**S**V**T**FP<br>**S**A**ST**KI**TT**LPN*VTFLPQN*A**T**TI**S**<br>SR#ENVN*T**SSS**VA**T**LAPVN**S**PAP<br>QD | 3 | 18 | h = 00%<br>e = 13%<br>c = 87%<br>found = PPII |
| *Domain-3 (165–303)*<br>T**T**AAPP**T**P**S**A**T**TPAPP**SSS**APPE<br>T**T**AAPP**T**P**S**A**TT**QAPP**SSS**APPE<br>T**T**AAPP**T**PPA**TT**PAPP**SSS**APPE<br>T**T**AAPP**T**P**S**A**TT**PAPL**SSS**APPE<br>T**T**AVPP**T**P**S**A**TT**LDPS**S**ASAPPE<br>T**T**AAPP**T**P**S**A**TT**PAPP**SS**PAPQE | 46 | 16 | h = 00%<br>e = 00%<br>c = 100%<br>found = PPII |
| *Domain-4 (304–355)*[c]<br>T**T**AAPI**TT**PN*S**S**P**TT**LAPD**TS**ET<br>**S**AAP**T**HQ**TTTS**V**TT**Q**TTTT**K#QP<br>**TS**APGQN | 22 | 0 | h = 00%<br>e = 10%<br>c = 90% |
| *Domain-5 (355–377)*[c]<br>KISRFLLYMKNLL[b]NRIIDDMVEQ | 0 | 0 | h = 60%<br>e = 40%<br>c = 00% |

[a] Histatin-like sequence that resembles salivary histatins (Oppenheim *et al.* [33]). A sequence similarity of 53.33% was observed.
[b] Sequence depicts leucine-zippers consisting of Leu residue repeated at i, i + 7 positions [35]; [36].
[c] 86 of 105 Ser and Thr are O-glycosylated based on empirical predictions and chemical analysis. Bold letters, **T** and **S** represent the potential sites of O-glycosylation for threonine and serine, respectively.
[d] h represents helix, e is for extended $\beta$-strand and c is for coil structures.
[e] 27% sequence homology with mouse polymorphic epithelial mucin (MUC1; [38]) and 43% sequence homology with human intestinal mucin (MUC2; [39]).
N*designates N-glycosylation sites.
# Trypsin cleavage site. Tryptic glycopeptide, **T1** corresponds to $aa^{142-345}$ whereas tryptic glycopeptide, **T2** corresponds to $aa^{70-141}$.

In order to identify which of the 105 Ser/Thr residues in MUC7 might be O-glycosylated, we took advantage of the following strategies: (1) semi-empirical prediction algorithm for O-glycosylation as suggested by Elhammer *et al.* [11] (Figure 1) and Hansen *et al.* [12]; (2) usage of the three tetrapeptide sequence motifs [e.g., Thr(g)-Xaa-Xaa-Xaa, where Thr(g) and one Xaa only are glycosylated threonine residues] as described by Pisano *et al.* [15]; and (3) the presence of a proline, alanine, serine, or threonine at positions + 3, − 1, − 6, and − 3 was often associated with glycosylation of Ser/Thr, whereas a charged residue at these positions abolishes the glycosylation as reported by O'Connell *et al.* [18]. The combined strategies together with

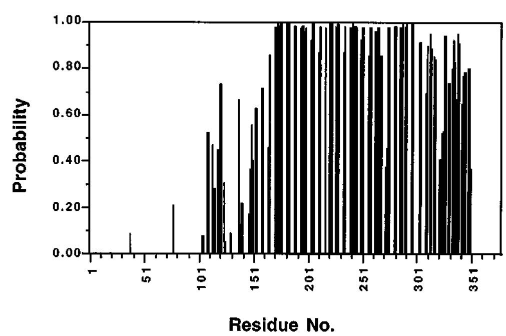


**Figure 1.** Plot showing the probability of Ser/Thr residues for O-glycosylation using semi-empirical specificity parameter data [11]. The plot is based on a window size of nine residues.

chemical analysis led to tentative assignment of all the potential O-glycosylation sites in MUC7. In the present study, the findings from semi-empirical prediction algorithms are in good agreement with the results from chemical analysis. In a recent detailed biochemical study, Gerken *et al.* [40] have determined the site-specific O-glycosylation pattern of a 81-residue tryptic tandem repeat glycopeptide from porcine submaxillary mucin. They found that the extent of O-glycosylation of this glycopeptide does not correlate well with the available predictive methods. In addition to sequence content, factors such as secondary structure and surface accessibility could influence O-glycosylation [40]. As a result, they have modeled an extended $\beta$-conformation to best account for the nearest neighbor and penultimate positional effects on O-glycosylation in the tandem repeat domain of porcine submaxillary mucin.

Table 3 gives the potential O-glycosylation sites in MUC7 using the available prediction algorithms of Elhammer *et al.* [11] and Hansen *et al.* [12]. The location of Ser/Thr glycosylation sites was pretty much straight forward in the case of domain 3, 4 and 5 (Table 2), and was mostly based on the Elhammer *et al.* [11] prediction method (Table 3) as it fits well for mucins [40]. The only exception is that residues Thr-165, Thr-188, Thr-211, Thr-234, Thr-257, Thr-280, Thr-303, Thr-322 and Thr-325 are assigned as non-glycosylated sites, due to the presence of acidic residues such as either Aps or Glu at $-1$ position which are known to disfavor the glycosylation [16]. In domain 2, both the prediction strategies gave different values for Thr-104 and Thr-125 leading to ambiguity in their assignment as O-glycosylation sites. However, these residues were assigned to be O-glycosylated due to the presence of $-1$ Pro. This assumption is based on the finding of Gerken *et al.* [40] where the presence of $-1$ Pro is thought to produce conformational effects favoring the O-glycosylation of Ser-43 in the tandem repeat glycopeptide of porcine submaxillary mucin. Likewise, the reason for assigning Thr-125 as O-glycosylated is based on the tetrapeptide sequence motif, Thr(g)-Xaa-Xaa-Xaa, where Thr(g) and one Xaa only are glycosylated threonine residues [15]. It is clear from Table 3 that the central region of MUC7, where the tandem repeats are located, contains the highest concentration of potential O-glycosylation sites.

In MUC7, there are 13 Asn aa residues (Table 2) of which five are determined to be N-glycosylated using the motif Asn-Xaa-Ser/Thr (where Xaa can be any residue except proline) [29, 30]. Four of the potential N-glycosylation sites are located in domain 2 while domain 4 has one such site.

## CD studies of diproline-containing peptides and glycopeptides

Trypsinolysis of the fucose-enriched isoform of MUC7 (formerly designated MG2a) resulted in two major tryptic glycopeptides designated **T1** and **T2** [6]. N-terminal sequencing and amino acid analysis revealed that glycopeptide **T1** corresponds to $aa^{142-345}$ while glycopeptide **T2** corresponds to $aa^{70-141}$ (Table 2). CD studies were performed on these glycopeptides and MUC7 to evaluate their content of polyproline type helical structures. Typically, a strong negative band at $\sim 205$ nm with a weak positive band at $\sim 229$ nm is characteristic of poly-L-proline type II helical structures [41]. The spectra measured for MUC7, **T1** and **T2** showed a $\lambda_{min}$ around 202–203 nm (Figure 2a). This $\lambda_{min}$ is characteristic of polyproline type II extended structures [41]; however, it is significantly shifted from the $\lambda_{min}$ of a typical random coil [42]. The relative content of polyproline type helical structures was **T1** > MUC7 > **T2** which is consistent with their number of diproline segments (Table 2).

To further examine the conformational effects imposed by diproline segments, several non-glycosylated and glycosylated portions of the polyproline-enriched MUC7 tandem repeat sequence were synthesized (Table 1). These included a 23 aa tandem repeat sequence with three diproline segments (**I**), a partial tandem repeat sequence with two diproline segments (**II**), a partial tandem repeat sequence containing GalNAc$\alpha$-Ser with two diproline segments (**III**) and a partial tandem repeat sequence containing Gal$\beta$(1-3)GalNAc$\alpha$-Ser with two diproline segments (**IV**). CD studies of peptides I and II showed a strong negative band around 201 nm which is indicative of the presence of poly-L-proline type II helical structures (Figure 2b). CD studies of glycopeptides III and IV containing GalNAc$\alpha$-Ser and Gal$\beta$(1-3)GalNAc$\alpha$-Ser, respectively, also showed a strong negative band at 201 nm with a weak broad positive band around 225–230 nm which is again a typical feature of a poly-L-proline type helical conformation (Figure 2b). Although, the CD spectra of MUC7 and its (glyco)peptides are characteristic of polyproline II structures, it is not possible to distinguish whether the CD contribution is from a population of fully extended polyproline structures or from conformations defined by segments of two or more residues. One can clearly visualize the structural effects conferred by the carbohydrates on the peptide backbone conformation, as the $\pi$–$\pi^*$ ellipticity values change significantly with concomitant red and/or blue shift in the CD bands. Though the overall band shape of these proline-rich (glyco)peptides remains unaffected, a change in the intensity of both the $n$–$\pi^*$ and $\pi$–$\pi^*$ bands was observed in the case of (glyco)peptides **II–IV**. This may be attributable to structural flexibility of these linear sequences as a consequence of decreased peptide chain length when compared to **I**, and/or contribution from carbohydrate chains separately to molar ellipticity values. Such variation in ellipticity values due to shorter peptide chain length has been observed and reported for helical peptides [44]. Since the CD spectra did not reveal any profound effect of carbohydrates on the APP-containing peptide backbone conformation, it appears

**Table 3.** Compilation of potential O-glycosylation sites in MUC7.

| *MUC7 aa Residue* | NetOglyc[a] *prediction* | *Probability*[b] *h value* | *Assignment (This study)* | *MUC7 aa Residue* | NetOglyc[a] *prediction* | *Probability*[b] *h value* | *Assignment (This study)* |
|---|---|---|---|---|---|---|---|
| Ser-38 | 0.00(−) | 0.09(−) | − | Thr-234 | 0.98(+) | 0.87(+) | − |
| Ser-41 | 0.00(−) | 0.00(−) | − | Thr-235 | 0.99(+) | 0.98(+) | + |
| Ser-62 | 0.00(−) | 0.00(−) | − | Thr-240 | 0.99(+) | 0.97(+) | + |
| Ser-77 | 0.12(−) | 0.21(+) | + | Ser-242 | 0.99(+) | 0.99(+) | + |
| Ser-98 | 0.06(−) | 0.00(−) | − | Thr-244 | 0.99(+) | 0.98(+) | + |
| Ser-99 | 0.10(−) | 0.00(−) | − | Thr-245 | 0.99(+) | 0.97(+) | + |
| Thr-104 | 0.77(+) | 0.08(−) | + | Ser-250 | 0.99(+) | 0.92(+) | + |
| Thr-108 | 0.67(+) | 0.52(+) | + | Ser-251 | 0.99(+) | 0.97(+) | + |
| Thr-109 | 0.98(+) | 0.52(+) | + | Ser-252 | 0.99(+) | 0.94(+) | + |
| Ser-113 | 0.83(+) | 0.47(+) | + | Thr-257 | 0.90(+) | 0.85(+) | − |
| Thr-115 | 0.82(+) | 0.28(+) | + | Thr-258 | 0.99(+) | 0.97(+) | + |
| Ser-118 | 0.98(+) | 0.45(+) | + | Thr-263 | 0.99(+) | 0.96(+) | + |
| Ser-120 | 0.99(+) | 0.30(+) | + | Ser-265 | 0.99(+) | 0.97(+) | + |
| Thr-121 | 0.78(+) | 0.73(+) | + | Thr-267 | 0.75(+) | 0.74(+) | + |
| Thr-124 | 0.16(−) | 0.31(+) | + | Thr-268 | 0.99(+) | 0.85(+) | + |
| Thr-125 | 0.75(+) | 0.06(−) | + | Ser-272 | 0.99(+) | 0.38(+) | + |
| Thr-130 | 0.65(−) | 0.09(−) | − | Ser-273 | 0.99(+) | 0.46(+) | + |
| Thr-137 | 0.46(−) | 0.67(+) | + | Ser-275 | 0.99(+) | 0.97(+) | + |
| Thr-138 | 0.65(−) | 0.13(−) | − | Thr-280 | 0.98(+) | 0.87(+) | − |
| Ser-140 | 0.23(−) | 0.22(+) | + | Thr-281 | 0.99(+) | 0.98(+) | + |
| Ser-141 | 0.01(−) | 0.08(−) | − | Thr-286 | 0.99(+) | 0.97(+) | + |
| Thr-147 | 0.30(−) | 0.17(−) | − | Ser-288 | 0.99(+) | 0.99(+) | + |
| Ser-148 | 0.35(−) | 0.37(+) | + | Thr-290 | 0.98(+) | 0.98(+) | + |
| Ser-149 | 0.45(−) | 0.56(+) | + | Thr-291 | 0.99(+) | 0.99(+) | + |
| Ser-150 | 0.25(−) | 0.40(+) | + | Ser-296 | 0.99(+) | 0.99(+) | + |
| Thr-153 | 0.98(+) | 0.63(+) | + | Ser-297 | 0.99(+) | 0.99(+) | + |
| Ser-159 | 0.95(+) | 0.71(+) | + | Thr-303 | 0.94(+) | 0.87(+) | − |
| Thr-165 | 0.94(+) | 0.46(+) | − | Thr-304 | 0.99(+) | 0.91(+) | + |
| Thr-166 | 0.98(+) | 0.86(+) | + | Thr-309 | 0.92(+) | 0.70(+) | + |
| Thr-171 | 0.99(+) | 0.97(+) | + | Thr-310 | 0.99(+) | 0.90(+) | + |
| Ser-173 | 0.99(+) | 0.99(+) | + | Ser-313 | 0.99(+) | 0.95(+) | + |
| Thr-175 | 0.99(+) | 0.98(+) | + | Ser-314 | 0.99(+) | 0.89(+) | + |
| Thr-176 | 0.99(+) | 0.99(+) | + | Thr-316 | 0.97(+) | 0.85(+) | + |
| Ser-181 | 0.99(+) | 0.98(+) | + | Thr-317 | 0.99(+) | 0.84(+) | + |
| Ser-182 | 0.99(+) | 0.99(+) | + | Thr-322 | 0.84(+) | 0.41(+) | − |
| Ser-183 | 0.99(+) | 0.99(+) | + | Ser-323 | 0.99(+) | 0.52(+) | + |
| Thr-188 | 0.98(+) | 0.87(+) | − | Thr-325 | 0.95(+) | 0.53(+) | − |
| Thr-189 | 0.99(+) | 0.98(+) | + | Ser-326 | 0.99(+) | 0.94(+) | + |
| Thr-194 | 0.99(+) | 0.97(+) | + | Thr-330 | 0.81(+) | 0.74(+) | + |
| Ser-196 | 0.99(+) | 0.98(+) | + | Thr-333 | 0.96(+) | 0.80(+) | + |
| Thr-198 | 0.98(+) | 0.96(+) | + | Thr-334 | 0.95(+) | 0.92(+) | + |
| Thr-199 | 0.99(+) | 0.97(+) | + | Thr-335 | 0.93(+) | 0.83(+) | + |
| Ser-204 | 0.99(+) | 0.92(+) | + | Ser-336 | 0.96(+) | 0.67(+) | + |
| Ser-205 | 0.99(+) | 0.99(+) | + | Thr-338 | 0.96(+) | 0.95(+) | + |
| Ser-206 | 0.99(+) | 0.99(+) | + | Thr-339 | 0.98(+) | 0.91(+) | + |
| Ser-227 | 0.99(+) | 0.98(+) | + | Thr-341 | 0.98(+) | 0.65(+) | + |
| Thr-211 | 0.98(+) | 0.87(+) | − | Thr-342 | 0.76(+) | 0.41(+) | + |
| Thr-212 | 0.99(+) | 0.98(+) | + | Thr-343 | 0.87(+) | 0.77(+) | + |
| Thr-217 | 0.99(+) | 0.97(+) | + | Thr-344 | 0.98(+) | 0.78(+) | + |
| Thr-221 | 0.99(+) | 0.99(+) | + | Thr-348 | 0.99(+) | 0.80(+) | + |
| Thr-222 | 0.99(+) | 1.00(+) | + | Ser-349 | 0.75(+) | 0.37(+) | + |
| Ser-228 | 0.99(+) | 0.98(+) | + | Ser-357 | 0.00(−) | 0.00(−) | − |
| Ser-229 | 0.99(+) | 0.99(+) | + | | | | |

[a]Based on Hansen *et al.* [12] prediction method; [b]Following Elhammer *et al.* [11] prediction method; + symbol denotes glycosylation whereas − symbol denotes non-glycosylation.

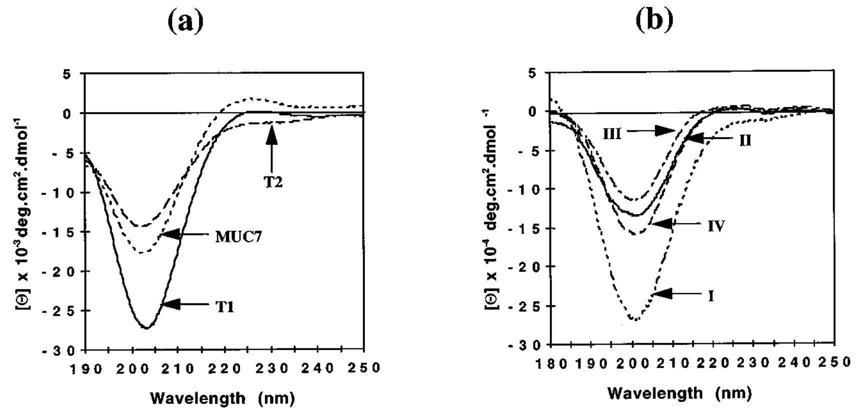


**Figure 2.** (a) CD spectra of MUC7, **T1** and **T2** recorded in 20 mM sodium phosphate buffer (pH 7.2) at 25 °C. (b) CD spectra of (glyco)peptide fragments (**I–IV**) recorded in 20 mM phosphate buffer (pH 7.2) at 25 °C.

that GalNAcα O-linked to Ser and Galβ(1-3)GalNAcα O-linked to Ser have either minimal or no influence on the polyproline type helical structures. Also, a five-fold increase or decrease in the peptide concentration in aqueous solutions does not significantly alter the CD band intensities, indicating that peptide association is not significant at the concentrations used for CD measurements.

Several studies have suggested that the influence of O-linked oligosaccharides on the conformation depends largely on the adjacent peptide sequence. For example, our results are consistent with the recent work of Arsequell *et al.* [45] and Liu *et al.* [46]. Arsequell *et al.* studied the structure of a synthetic O-glycosylated Sendai Virus Nucleoprotein analog, FAPS*NYPAL (*α-D-GalNAc) by NMR and found no differences in the amide resonances of glycosylated versus non-glycosylated peptide analogues indicating no influence of α-D-GalNAc on the peptide backbone structure [45]. It may be noted that this small peptide has two Pro in the sequence. Liu *et al.* synthesized MUC-1 derived 16 amino acid peptide analogs, GVT*S*APDTRPAPGSTA (*α-D-GalNAc) which contains three Pro and found that α-D-GalNAc attached to the threonine or serine did not impose any conformational changes to the peptide backbone nor has offered any severe steric resistance to the binding of antibodies [46]. In contrast, Butenhof and Gerken illustrated that specific hydrogen bond interactions between the carbohydrate and the peptide moiety can occur by stochastic dynamics simulation method utilizing an ovine submaxillary mucin-like glycopeptide, *viz*, Ncap-AAAT*T*AAA-Ccap(*α-D-GalNAc) devoid of Pro [47]. Similar observations were seen following NMR studies on synthetic proline-free model glycopeptides such as Z-T*AA-OMe (*α-D-GalNAc, Z = benzyloxycarbonyl) from antifreeze glycoproteins [48] where an intramolecular hydrogen bond formed between the amide proton of α-D-GalNAc and the carbonyl oxygen of threonine to which the α-D-GalNAc is covalently linked. Light scattering and circular dichroism studies of native and glycosidase-treated salivary mucin revealed that GalNAc residues attached to threonine on the peptide core are essential in maintaining a highly extended random coil configuration [49]. Removal of all disaccharides led to collapse or denaturation of the molecule. In a separate study, Liang *et al.* [50] have shown that different carbohydrate chain length can interact with the same peptide backbone to stabilize different conformations. Collectively, the aforementioned studies suggest that the conformational effects conferred by α-D-GalNAc and/or Galβ(1-3)GalNAcα depend on peptide sequence and/or the presence of constraining amino acid residues (e.g., Pro) around the O-glycosylation site.

## Molecular structure and crystal packing of APP

CD studies indicated that the diproline segments in the tandem repeat domain can contribute to the poly-L-proline type helical conformation in MUC7. However, these studies could not determine the type of poly-L-proline structure. An earlier study on the conformation of Boc-Pro-Pro-OH [23] revealed that this dipeptide derivative adopted both PPI and PPII conformations and suggested that diprolines may be too short to be used as model peptides for determining the conformational preferences of MUC7. Since the majority of diproline segments in the MUC7 tandem repeats were comprised of APP, an APP-derivative, Boc-Ala-Pro-Pro-OBzl, was synthesized and its polyproline type characteristics were determined by X-ray crystallographic analysis. The molecular structure of APP with the atomic numbering scheme is shown in Figure 3. There are two crystallographically independent molecules in the asymmetric unit, and these are referred to as A and B. The bond lengths and bond angles observed in the two independent molecules agree within $3\sigma$. Of the two benzyl groups, one associated with molecule B has a higher thermal motion. The relevant backbone torsion angles and those corresponding to the

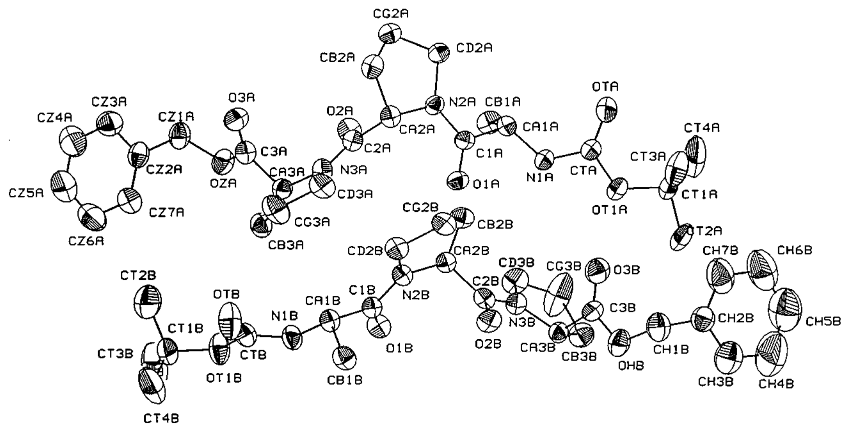


**Figure 3.** A perspective view of Boc-Ala-Pro-Pro-OBzl showing the atomic numbering scheme used. Thermal ellipsoids are drawn at 50% probability levels.

**Table 4.** Selected torsion angles[#] in Boc-Ala-Pro-Pro-OBzl.

| *Molecule A* | | *Molecule B* | | *Angle* |
|---|---|---|---|---|
| *Backbone torsions* | | | | |
| CTA-N1A-CA1A-C1A | – 142 | CTB-N1B-CA1B-C1B | – 153 | $\phi 1$ |
| N1A-CA1A-C1A-N2A | 156 | N1B-CA1B-C1B-N2B | 154 | $\phi 1$ |
| CA1A-C1A-N2A-CA2A | 174 | CA1B-C1B-N2B-CA2B | 177 | $\omega 1$ |
| C1A-N2A-CA2A-C2A | – 72 | C1B-N2B-CA2B-C2B | – 76 | $\phi 2$ |
| N2A-CA2A-C2A-N3A | 150 | N2B-CA2B-C2B-N3B | 151 | $\phi 2$ |
| CA2A-C2A-N3A-CA3A | 173 | CA2B-C2B-N3B-CA3B | 170 | $\omega 2$ |
| C2A-N3A-CA3A-C3A | – 70 | C2B-N3B-CA3B-C3B | – 68 | $\phi 3$ |
| N3A-CA3A-C3A-OZA | 150 | N3B-CA3B-C3B-OHB | 153 | $\phi 3$ |
| *Proline torsions* | | | | |
| CA2A-CB2A-CG2A-CD2A | – 38 | CA2B-CB2B-CG2B-CD2B | – 36 | $\tau 0$ |
| CB2A-CG2A-CD2A-N2A | 30 | CB2B-CG2B-CD2B-N2B | 28 | $\tau 1$ |
| CD2A-N2A-CA2A-CB2A | – 12 | CD2B-N2B-CA2B-CB2B | – 13 | $\tau 2$ |
| CA2A-N2A-CD2A-CG2A | – 12 | CA2B-N2B-CD2B-CG2B | – 9 | $\tau 3$ |
| N2A-CA2A-CB2A-CG2A | 31 | N2B-CA2B-CB2B-CG2B | 30 | $\tau 4$ |
| CA3A-CB3A-CG3A-CD3A | – 32 | CA3B-CB3B-CG3B-CD3B | – 19 | $\tau 0$ |
| CB3A-CG3A-CD3A-N3A | 26 | CB3B-CG3B-CD3B-N3B | 12 | $\tau 1$ |
| CA3A-N3A-CD3A-CG3A | – 10 | CA3B-N3B-CD3B-CG3B | 0 | $\tau 2$ |
| CD3A-N3A-CA3A-CB3A | – 9 | CD3B-N3B-CA3B-CB3B | – 11 | $\tau 3$ |
| N3A-CA3A-CB3A-CG3A | 25 | N3B-CA3B-CB3B-CG3B | 17 | $\tau 4$ |

[#]Supplementary material including experimental details, isotropic/anisotropic temperature factors, additional interatomic distances, bond angles, torsion angles, observed and calculated structure factors for APP are available upon request. The e.s.ds of torsion angle are in the range 0.6–1.3°.

proline pucker are given in Table 4. There are no *cis* peptide bonds in the structure. Only one of the prolines exhibits envelope conformation; all the other three exhibit twisted conformations in which the $\beta$ and $\gamma$ atoms are above and below the plane defined by the rest of the ring atoms [51]. Both the molecules exhibit a PPII conformation with characteristic backbone torsion angles $\phi$ (in the range $-68$ to $-70°$) and $\varphi$ (in the range 148 to 151°).

The molecular packing observed for APP in the crystal is shown in Figure 4. Only one hydrogen bond is possible between NI and any one of the six oxygen atoms per molecule. However, no hydrogen bonding is observed in the structure. The aligning of the molecules along the **b** axis (40 Å) reveals an interesting feature in that the polyproline conformation is extended to infinity along this axis. This suggests that the polyproline conformation dominates the

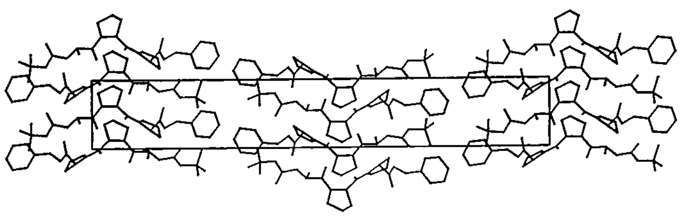

**Figure 4.** Crystal packing observed in the solid-state structure of Boc-Ala-Pro-Pro-OBzl.

packing. In contrast, along the other two shorter axes (6 and 10 Å), the neighboring polyproline chains generated by the two-fold screw pack in such a way that the proline rings are in close proximity (Figure 4) reminiscent of collagen packing [52].

## Overall conformational features of MUC7

The high content of O-glycosylated Ser and Thr residues is a characteristic of most mucins. Previous studies on ovine and porcine submaxillary mucins have indicated that these molecules adopt a semi-rigid rod like structure [47, 53, 54]. In these mucins, the sequential removal of carbohydrate units led to a random coil-like structure of the peptide backbone as evidenced by the reduction in the radius of gyration. On the other hand, addition of neutral and acidic oligosaccharides units increases the radius of gyration of mucins by a factor of 2.5–3 when compared with other proteins of similar chain length [55]. Thus, it was concluded that the O-linked carbohydrate chains play a significant role in the ability of these mucins to adopt filamentous, non-random coil structures and in the stiffening of the mucin structure [53, 54]. In contrast to these mucins, the peptide backbone of MUC7 contains numerous diproline segments. These diproline segments, the majority of which are located in the tandem repeat domain, enable MUC7 to adopt an extended structure characteristic of a PPII conformation. CD studies on synthetic diproline-rich peptides and glycopeptides suggested that O-linked GalNAc$\alpha$ or Gal$\beta$(1-3)GalNAc$\alpha$ did not significantly influence the PPII conformation adopted by the adjacent peptide domains. In conclusion, the data presented indicate that the polyproline type II structure adopted by the diproline segments, dispersed throughout the tandem repeats, may impart a semi-rigid, rod shaped conformation to the MUC7 peptide backbone and could act in consort with the numerous O-glycosylated units to give the mucin a 'bottle-brush' appearance.

## Acknowledgments

This work was supported by USPHS grants DE07585, DE08240 and DE11064.

## References

1 Schenkels LCPM, Gururaja TL, Levine MJ (1996) In *Oral Mucosal Drug Delivery* (Rathbone MJ, ed) pp 191–220. New York: Marcel Dekker.
2 Cohen RE, Levine MJ (1989) In *Human Saliva: Clinical Chemistry and Microbiology*, Vol I (Tenovuo JO, ed) pp 101–30. Boca Raton: CRC Press.
3 Loomis RE, Prakobphol A, Levine MJ, Reddy MS, Jones PC (1987) *Arch Biochem Biophys* **258**: 452–64.
4 Prakobphol A, Levine MJ, Tabak LA, Reddy MS (1982) *Carbohydr Res* **108**: 111–22.
5 Ramasubbu N, Reddy MS, Bergey EJ, Haraszthy G, Soni S-D, Levine MJ (1991) *Biochem J* **280**: 341–52.
6. Reddy MS, Bobek LA, Haraszthy GG, Biesbrock AR, Levine MJ (1992) *Biochem J* **287**: 639–43.
7 Bobek LA, Tsai H, Biesbrock AR, Levine MJ (1993) *J Biol Chem* **268**: 20563–9.
8 Reddy MS, Levine MJ, Prakobphol A (1985) *J Dent Res* **64**: 33–6.
9 Strous GJ, Dekker J (1992) *CRC Rev Biochem Mol Biol* **27**: 57–92.
10 Kim YS, Gum JR, Brockhausen I (1996) *Glycoconjugate J* **13**: 693–707.
11 Elhammer AP, Poorman RA, Brown E, Maggiora LL, Hoogerheide JG, Kezdy FJ (1993) *J Biol Chem* **268**: 10029–38.
12 Hansen JE, Lund O, Engelbrecht J, Bohr H, Nielsen JO, Hansen JES, Brunak S (1995) *Biochem J* **308**: 801–13.
13 Chou KC (1995) *Protein Sci* **4**: 1365–83.
14 Yoshida A, Suzuki M, Ikenaga H, Takeuchi M (1997) *J Biol Chem* **272**: 16884–8.
15 Pisano A, Redmond J, Williams K, Gooley A (1993) *Glycobiology* **3**: 429–35.
16 Nehrke K, Hagen FK, Tabak LA (1996) *J Biol Chem* **271**: 7061–5.
17 Wilson IBH, Gavel Y, Heijne G (1991) *Biochem J* **275**: 529–34.
18 O'Connell B, Tabak LA, Ramasubbu N (1991) *Biochem Biophys Res Commun* **180**: 1024–30.
19 Williamson MP (1994) *Biochem J* **297**: 249–60.
20 Hay DI, Moreno EC (1989) In *Human Saliva: Clinical Chemistry and Microbiology*, Vol I (Tenovuo JO, ed) pp 131–50. Boca Raton: CRC Press.
21 Blundell TL, Pitts JE, Tickle IJ, Wood SP, Wu CW (1981) *Proc Natl Acad Sci USA* **78**: 4175–9.
22 Darbon H, Bernassau J-M, Deleuze C, Chenu J, Roussel A, Cambillau C (1992) *Eur J Biochem* **209**: 765–71.
23 Thomas LM, Ramasubbu N, Bhandary KK (1994) *Int J Peptide Protein Res* **44**: 207–14.
24 Bodanszky M, Bodanszky A (1984) In *The Practice of Peptide Synthesis*, New York: Springer-Verlag.
25 Fields GB, Noble RL (1990) *Int J Peptide Protein Res* **35**: 161–214.
26 Gururaja TL, Levine MJ (1996) *Peptide Res* **9**: 283–9.
27 Paulsen H, Peters S, Bielfeldt T, Meldal M, Bock K (1995) *Carbohydr Res* **268**: 17–34.
28 Gururaja TL, Ramasubbu N, Levine MJ (1996) *Letts Peptide Sci* **3**: 79–88.
29 Bause A (1983) *Biochem J* **209**: 331–6.

30 Shakin-Eshleman SH, Spitalnik SL, Kasturi L (1996) *J Biol Chem* **271**: 6363–6.
31 Sheldrick GM (1985) In *Crystallographic Computing 3* (Sheldrick GM, Kruger C, Goddard R, eds) pp 175–89. New York: Oxford University Press.
32 *International Tables for Crystallography* (1974) Vol IV (present distributor – D. Riedel, Dordrecht) Birmingham: Kynoch Press.
33 Oppenheim FG, Xu T, McMillian FM, Levitz SM, Diamond RD, Offner GD, Troxler RF (1988) *J Biol Chem* **263**: 7472–7.
34 Levine JH, Tran DT, Gururaja TL, Ramalingam K, Ramasubbu N, Levine MJ (1997) *J Dent Res* **76**: Abstr. No. 1650.
35 Landschulz WH, Johnson PF, McKnight SL (1988) *Science* **240**: 1759–64.
36 O'Shea EK, Rutkowski R, Stafford WF 3rd, Kim PS (1989) *Science* **245**: 646–8.
37 Fontenot JD, Tjandra N, Bu D, Ho C, Montelaro RC, Finn OJ (1993) *Cancer Res* **53**: 5386–94.
38 Vos HL, De Vries Y, Hilkens J (1991) *Biochem Biophys Res Commun* **181**: 121–30.
39. Gum JR, Hicks JW, Toribara NW, Siddiki B, Kim YS (1994) *J Biol Chem* **269**: 2440–6.
40 Gerken TA, Owens CL, Pasumarthy M (1997) *J Biol Chem* **272**: 9709–19.
41 Ronish EW, Krimm S (1974) *Biopolymers* **13**: 1635–51.
42 MacArthur MW, Thornton JM (1991) *J Mol Biol* **218**: 397–412.
43 Woody RW (1977) *J Polymer Sci* **12**: 181–321.
44 Vijayakumar EKS, Sudha TS, Balaram P (1984) *Biopolymers* **23**: 877–86.
45 Arsequell G, Haurum JS, Elliot T, Dwek RA, Lellouch AC (1995) *J Chem Soc Perkin Trans* 1, 1739–45.
46 Liu X, Sejbal J, Kotovych G, Koganty RR, Reddish MA, Jackson L, Gandhi SS, Mendonca AJ, Longenecker BM (1995) *Glycoconjugate J* **12**: 607–17.
47 Butenhof KJ, Gerken TA (1993) *Biochemistry* **32**: 2650–63.
48 Mimura Y, Inoue Y, Maeji NJ, Chujo R (1992) *Int J Biol Macromol* **14**: 242–8.
49 Shogren R, Gerken TA, Jentoft N (1989) *Biochemistry* **28**: 5525–35.
50 Liang R, Andreotti AH, Kahne D (1995) *J Am Chem Soc* **117**: 10395–6.
51 Hasnoot CAG, De Leeuw FAAM, De Leeuw HPM, Altona C (1981) *Biopolymers* **20**: 1211–45.
52 Carver JP, Blout ER (1967) In *Treatise on Collagen*, Vol 1 (Ramachandran GN, ed) pp 441–526. New York: Academic Press.
53 Gerken TA, Butenhof KJ, Shogren R (1989) *Biochemistry* **28**: 5536–43.
54 Gerken TA (1993) *Crit Rev Oral Biol Med* **4**: 261–70.
55 Jentoft N (1990) *Trends Biochem Sci* **15**: 291–4.

Received 21 April 1997, revised 2 September 1997, accepted 6 October 1997