Journal Articles: Eppley Institute   Eppley Institute

Winter 2-20-2004

NMR-based Structural Studies of the Glycosylated MUC1 Tandem Repeat Peptide

G. Suryanarayanan
University of Nebraska Medical Center

P. A. Keifer
University of Nebraska Medical Center

G. Wang
University of Nebraska Medical Center

L. Kinarsky
University of Nebraska Medical Center

M. A. Hollingsworth
University of Nebraska Medical Center, mahollin@unmc.edu

See next page for additional authors

Tell us how you used this information in this *short survey*.

Follow this and additional works at: https://digitalcommons.unmc.edu/eppley_articles

Part of the **Neoplasms Commons**, and the **Oncology Commons**

Recommended Citation

Authors

G. Suryanarayanan, P. A. Keifer, G. Wang, L. Kinarsky, M. A. Hollingsworth, and S. Sherman

NMR-based Structural Studies of the Glycosylated MUC1 Tandem Repeat Peptide

G. Suryanarayanan, P.A. Keifer, G. Wang, L. Kinarsky, M.A. Hollingsworth and S. Sherman*

Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE, 68198-6805, USA.
* Corresponding author. E-mail: ssherm@unmc.edu

Abstract: MUC1 is a glycoprotein that plays an important role in cancer pathogenesis. In order to study the effect of glycosylation on the conformational propensities of the tandem repeat domain of MUC1, we have determined the structure of the MUC1 tandem repeat peptide AHGVTSAPDTRPAPGSTAPP, *O*-glycosylated with the trisaccharide (α-Glc-1,4-β-Glc-1,4-α-GalNAc-) at Thr5. This glycopeptide was synthesized to model a heavily *O*-glycosylated threonine residue in the tandem repeat domain. The NMR experiments used in this study included TOCSY, NOESY, ROESY, DQF-COSY, HSQC and 1D NMR. The peak volumes determined using the program SPARKY were converted into distance constraints using the program CALIBA. The programs FiSiNOE and HABAS were used to generate angle constraints. Using conformational restraints obtained from NMR, the program DYANA was used to determine the structures of the peptide. Finally, structural refinement was performed within the SYBYL software package using GLYCAM parameters and Kollman-all atom types. The presence of strong sequential αN connectivities suggested an extended conformation of the peptide backbone. Strong sequential αδ connectivities were indicative of a *trans* conformation of the Ala-Pro peptide bonds. In addition, presence of sequential NN connectivities in the peptide segments Gly3-Val4-Thr5-Ser6, Asp9-Thr10-Arg11 and Gly-Ser16 were indicative of twist-like conformations of the peptide backbone in these peptide segments.

Keywords: MUC1, glycopeptide, trisaccharide, NMR.

Introduction

MUC1 [1] is a glycoprotein expressed on the cell surface of normal and cancer cells. The large extracellular fragment of MUC1 consists of tandem repeats of a 20 amino acid motif. Each tandem repeat AHGVTSAPDTRPAPGSTAPP includes two serine and three threonine residues that are potential sites of *O*-glycosylation. In normal secretory epithelial cells, MUC1 is heavily glycosylated and expressed on the apical surface, whereas in cancer cells MUC1 is aberrantly glycosylated and contributes to reconfiguration of cell-cell interactions. The altered properties of MUC1 on tumor cells modulate adhesion to receptors on adjoining cells and tissues thereby facilitating metastasis [2]. Thus, glycosylation of MUC1 plays an important role in the pathophysiology of various cancers which include pancreatic adenocarcinoma, lung, breast and colon cancer.

The initiation of *O*-glycosylation of MUC1 in the tandem repeat domain is catalyzed by a family of enzymes called UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases (GalNAc transferases) [3-7]. Previous *in vitro* studies on the initiation of glycosylation of the tandem repeat domain of MUC1 have shown that attachment of GalNAc residues to specific sites within the tandem repeat peptide of MUC1 facilitates glycosylation at other sites and leads to high-density glycosylation [8-10]. It has been hypothesized that the GalNAc transferases following initial glycosylation, bind to the GalNAc residue through their lectin domain, which results in conformational changes in its catalytic domain or in the acceptor substrate peptide backbone that facilitate glycosylation at proximal and distant sites [9]. In order to characterize the conformational changes on acceptor substrates that take place following initial glycosylation, previous studies were focused on MUC1-based tandem repeat peptides monoglycosylated with a single GalNAc residue at Thr5 [11]. We extended those studies to glycosylation of tandem repeat peptides with bulky carbohydrate moieties. Using NMR, we have investigated the structure of a MUC1-based tandem repeat peptide *O*-glycosylated with the trisaccharide (α-Glc-1,4-β-Glc-1,4-α-GalNAc-) at Thr5. The trisaccharide α-Glc-1,4-β-Glc-1,4-α-GalNAc was designed to mimic natural glycoforms that glycosylate the tandem repeat domain of MUC1.

Materials and Methods

Total correlation spectroscopy (TOCSY), nuclear Overhauser enhancement spectroscopy (NOESY), double quantum filtered correlation spectroscopy (DQF-COSY), 1D NMR, and heteronuclear single quantum coherence (HSQC) NMR experiments were used to obtain structural data on the peptide. All data were acquired on a 600 MHz Varian INOVA spectrometer operating at 14.1 Tesla. The sample was dissolved in 90:10 H2O:D2O to a concentration of 3mM at pH 4.5 and the data were acquired at a sample temperature of 5°C. Water suppression was accomplished by using presaturation (50 Hz field) except for the HSQC experiment, which used pulsed field gradient solvent suppression. NMR data were apodized by a Gaussian function and processed using VNMR software (Varian, Palo Alto, CA).

Sequential resonance assignments of protons [12] in the glycopeptide were obtained using the TOCSY and NOESY spectra. $^3J_{N\alpha}$ vicinal coupling constants were obtained using the DQF-COSY spectra. In addition to data obtained from the NOESY and TOCSY spectra, the HSQC spectrum was used to assist resonance assignments of protons in the trisaccharide. The peak intensities in the NOESY spectrum (200 ms) were integrated using the Gaussian fit method implemented in the program SPARKY [13]. The peak integrals were input into the program CALIBA [14] to obtain upper distance constraints. The distance constraints and coupling constants were used by the programs FiSiNOE [15] and HABAS [16] to generate φ, ψ, and $χ^1$ torsion angle constraints and stereospecific assignments. Additional stereospecific assignments for β-methylene protons were obtained using the program GLOMSA [14]. A model of the peptide AHGVTSAPDTRPAPGSTAPPA, *O*-glycosylated with the trisaccharide (α-Glc-1,4-β-Glc-1,4-α-GalNAc-) at Thr5, was created within the software package SYBYL (version 6.6) [17] using Kollman forcefield [18] and Glycam parameters [19]. The torsion angle constraints, distance constraints, and stereospecific assignments were used by the program DYANA [20] within SYBYL to generate 100 structures of the glycopeptide that were further subjected to constrained energy minimization. In order to determine specific conformational features of the peptide segment GVTSA, which contains the glycosylated Thr and flanking residues, the 100 energy-minimized structures were clustered together using a root-mean-square-deviation (RMSD) criterion of 0.6 Å for the backbone N, Cα and C atoms. The conformational propensities of individual amino acid residues in the peptide segment GVTSA were determined by their relative occupancy in different regions of the Ramachandran plot [21, 22].

Results and Discussion

### *NOE Connectivities and Coupling Constants*

Figure 1 shows the summary of sequential connectivities observed in the peptide backbone. These data indicate strong to medium sequential αN NOE connectivities between majority of the amino acid residues. Strong sequential αδ connectivities within the dipeptide segments Xxx-Pro and absence of sequential αα connectivities were indicative of a *trans* conformation for these dipeptide segments [12]. Abundant αN sequential connectivities suggest that at least some or a majority of the conformers of the glycopeptide exist in solution having extended conformations of the peptide backbone. However, the presence of weak to medium sequential NN connectivities in the peptide segments Gly3-Val4-Thr5*-Ser6 (* = α-Glc-1,4-β-Glc-1,4-α-GalNAc-), Asp9-Thr10-Arg11 and Gly15-Ser16 (Figures 1 and 2), indicates that a significant population of conformers have twist-like conformations of the peptide backbone in these peptide segments. The absence of medium-range NOEs in the peptide excludes the presence of tight turns or α-helical structures in the peptide backbone. Thus, the majority of conformers of the glycopeptide exist in solution in a dynamic equilibrium between extended and twist-like conformations for the peptide segments GVSTA, DTR and GS.

Figure 1. Summary of sequential NOE connectivities and coupling constants observed between amino acid residues in the peptide backbone.

Figure 2. A section of the 200 ms NOESY spectra of the glycopeptide showing sequential amide-amide cross peaks (HN = amide proton of the amino acid residues, HAC = amide proton of the N-acetyl group in the α-GalNAc residue).

The degenerate chemical shifts of αH and NH protons of His2, Thr5, Ser6, Ala7, Val4, and Thr10 precluded the determination of $^3J_{N\alpha}$ coupling constants for these amino acid residues. The coupling constants for the remaining amino acid residues ranged from 6.2 – 8.5 Hz and did not provide any evidence for the presence of ordered secondary structures in the peptide backbone.

αH Chemical Shift Deviations for the Glycopeptide

The αH proton chemical shift deviations of amino acid residues from their corresponding random coil values [23], provides an indication of specific secondary structures in peptides and proteins. A summary of the chemical shift deviations of the amino acid residues in the glycopeptide is shown in Figure 3. Val4 and Thr5 showed significant downfield deviations (~0.3 ppm) in αH chemical shifts from random coil values. These deviations in chemical shifts from random coil values were probably caused by local conformational effects induced by the glycosylation at Thr5. In general, despite the presence of amino acids having significant deviation of chemical shifts from the random coil values, the lack of a dense grouping of amino acids with significant downfield or upfield shifts suggests the absence of ordered secondary structures, such as α-helices or β-strands in the peptide backbone.
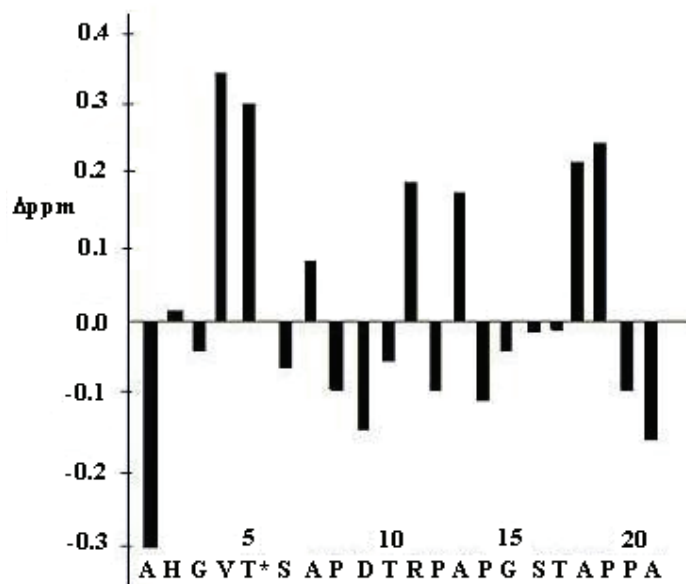

Figure 3. αH chemical shift deviations of amino acid residues in the glycopeptide from their corresponding random coil values (***** = α-*Glc-1,4-β-Glc-1,4-α-GalNAc-*).

Sugar-peptide NOE Connectivities for the Glycopeptide

The sugar-peptide NOE connectivities observed in the NOESY spectrum of the glycopeptide are represented in Figure 4. Several strong, medium and weak NOEs were observed between protons in the side chain of Thr5 and protons in the α-GalNAc residue. In addition, weak NOEs were observed between protons in α-GalNAc residue and protons in the peptide backbone. A weak NOE was

observed between the amide proton in the α-GalNAc residue and the amide proton of glycosylated Thr5 in the peptide backbone. The weak NOE between methyl group in α-GalNAc and αH proton of Ser6 and the medium NOE between H3 proton of α-GalNAc and NH proton of Ala7 indicate that the trisaccharide is oriented towards the C-terminus of the peptide backbone. These sugar-peptide interactions may mediate local conformational changes in the peptide segment GVTSA following glycosylation with the trisaccharide. These sugar-peptide interactions may also provide a local stabilizing effect for the peptide segment GVTSA. The trisaccharide α-Glc-1,4-β-Glc-1,4-α-GalNAc- due to its orientation towards the C-terminus of the peptide backbone may hinder addition of sugar residues by glycosyltransferases to the side chain of adjacent Ser6. However, our structural studies indicated that the side chains of Thr5 and Ser6 were facing away from each other. Hence, this orientation of the side chains may facilitate glycosylation of both residues with extended glycans by different glycosyltransferases.
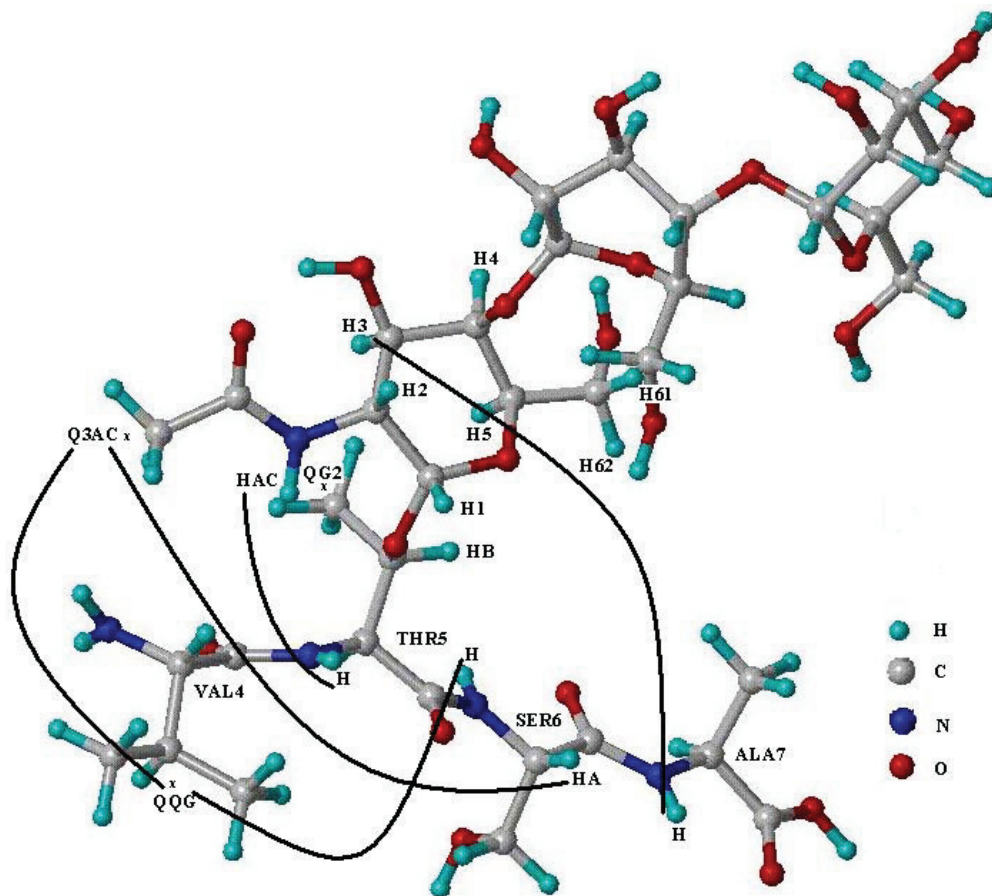
Figure 4. Schematic diagram showing connectivities between protons of α-GalNAc residue in the trisaccharide and protons of the amino acid residues in the peptide segment VTSA (H1, H2, H3, H4, H5, H61, H62 = protons of α-GalNAc residue, Q3AC = methyl group pseudo atom in the N-acetyl group, QQG = pseudo atom at the center of two methyl groups in Val4, QG2 = pseudo atom at the center of the methyl group in Thr5, HAC = amide proton in the N-acetyl group, HA = α proton in the peptide backbone, H = amide proton in the peptide backbone).

Cluster Analysis of the Peptide Segment GVTSA

In order to determine structural features of the peptide segment GVTSA, we clustered all conformers of the glycopeptide at the segment GVTSA using an RMSD criterion of 0.6 Å for backbone atoms N, Cα, and C. The clustering of conformers resulted in three distinct clusters containing 25%, 45%, and 16% (Clusters 1, 2, and 3) of the conformers and several small clusters with 1-4 conformers. In order to determine distinct conformational propensities of individual amino acid residues in the peptide segment GVTSA, the average structures of the conformers in each cluster were calculated. The φ and ψ torsion angles values (in degrees) for each amino acid residue in the average structure of the peptide segment GVTSA in each cluster were determined (Table 1).

In the peptide segment GVTSA, Val4 showed conformational propensities for the extended β-strand-like and inverse γ turn-like regions of the Ramachandran plot. Thr5 and Ala7 showed conformational propensities for regions corresponding to extended β-strand-like conformations, whereas Ser6 showed conformational propensities for regions corresponding to extended β-strand-like or polyproline II-like conformations.

In order to determine the effects of extensive glycosylation on the conformational propensities of the peptide backbone, further studies will be focused on peptides that will be glycosylated with trisaccharides at other potential sites of *O*-glycosylation in the tandem repeat domain, which include Thr10 in the PDTR and Thr17 in the GSTA region. These conformational studies together with in vitro kinetic and affinity studies for these glycopeptides will provide important insights into the regulation of glycosylation of the tandem repeat domain of MUC1.

Table 1. Values of the torsion angles φ and ψ (in degrees) of amino acid residues in the average structure of each cluster for the peptide segment GVTSA.

	Gly3		Val4		Thr5		Ser6		Ala7	
	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ
Cluster1	171	−157	−56	113	−109	121	−142	131	−94	164
Cluster2	−152	−136	−83	75	−117	117	−128	134	−103	139
Cluster3	−87	−112	−149	164	−110	127	−81	175	−113	142

Acknowledgements

References

1. Lan, M. S.; Batra, S. K.; Qi, W. N.; Metzgar, R. S.; Hollingsworth, M. A. Cloning and sequencing of a human pancreatic tumor-mucin cDNA. *J. Biol. Chem*. **1990**, *265*, 15294-15299.
2. McDermott, K. M.; Crocker, P. R.; Harris, A.; Burdick, M. D.; Hinoda, Y.; Hayashi, T.; Imai, K.; Hollingsworth, M. A. Overexpression of MUC1 reconfigures the binding properties of tumor cells. *Int. J. Cancer* **2001**, *94*, 783-791.
3. Wandall, H. H.; Hassan, M.; Mirgorodskaya, F.; Kristensen, A. K.; Roepstorff, P.; Bennett, E. P.; Nielsen, P. A.; Hollingsworth, M. A.; Burchell, J.; Taylor Papadimitriou, J.; Clausen, H. Specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine: Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *J. Biol. Chem*. **1997**, *272*, 23503-23514.
4. Stadie, T.; Chai, W.; Lawson, A. M.; Byfield, P.; Hanisch, F. G. Studies on the order and site-specificity of GalNAc-transfer to MUC1 tandem repeats by UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase from milk or mammary carcinoma cells. *J. Biol. Chem*. **1995**, *274*, 27867-27874.
5. Nishimori, I.; Johnson, N. R.; Sanderson, S. D.; Perini, R.; Mountjoy, K.; Cernag, R. L.; Gross, M. L.; Hollingsworth, M. A. Influence of acceptor substrate primary amino acid sequence on the activity of human UDP-N-acetylgalactosamine: polypeptide N-acetylgalactosaminyltransferase. Studies with the MUC1 tandem repeat. *J. Biol. Chem*. **1994**, *269*, 16123-16130.
6. Bennett, E. P.; Hassan, H.; Mandel, U.; Hollingsworth, M. A.; Akisawa, N.; Ikematsu, Y.; Merkx, G.; van Kessel, A. G.; Olofsson, S.; Clausen, H. Cloning and characterization of a close homologue of human UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *J. Biol. Chem*. **1998**, *274*, 25362-25370.
7. Hanisch, F. G.; Muller, S.; Hassan, H.; Clausen, H.; Zachara, N.; Gooley, A. A.; Paulsen, H.; Alving, K.; Peter-Katalinic, J. Dynamic epigenetic regulation of initial O-glycosylation by UDP-N-Acetylgalactosamine: Peptide N-acetylgalactosaminyl- transferases. Site-specific glycosylation of MUC1 repeat peptide influences the substrate qualities at adjacent or distant Ser/Thr positions. *J. Biol. Chem*. **1999**, *274*, 9946-9954.
8. Bennett, E. P.; Hassan, H.; Hollingsworth, M. A.; Clausen, H. A novel human UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc- glycosylated acceptor substrates. *FEBS Lett.* **1999**, *460*, 226–230.
9. Hassan, H.; Reis, C.; Bennett, E. P.; Mirgorodskaya, E.; Roepstorff, P.; Hollingsworth , M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J. Biol. Chem.* **2000**, *275*, 38197–38205.
10. Hanisch, F. G.; Reis C. A.; Clausen, H.; Paulsen, H. Evidence for glycosylation- dependent activities of polypeptide N-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides. *Glycobiology* **2001**, *11*, 731-740.
11. Kirnarsky, L.; Prakash, O.; Vogen, S. M.; Nomoto, M.; Hollingsworth, M. A.; Sherman, S. Structural effects of *O*-glycosylation on a 15-residue peptide from the mucin (MUC1) core protein. *Biochemistry* **2000**, *39*, 12076-12082.

12. Wüthrich, K. NMR of proteins and nucleic acids, John Wiley and Sons, New York, **1986**.

13. Goddard, T. D.; Kneller, D. G. SPARKY 3, University of California, San Francisco.

14. Güntert, P.; Braun, W.; Wüthrich, K. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **1991**, *217*, 517-530.

15. Shats, O.; Sherman, S. The FiSiNOE-3 Program for Determination of Protein and Peptide Conformations from NMR Data. *Third Electronic Computational Chemistry Conference (ECCC-3)* **1996**, http://bims.unmc.edu/ECCC3/fisinoe3.htm

16. Güntert, P.; Wüthrich, K. Automated Stereo Specific ^{1}H NMR Assignments and their Impact on the Precision of Protein Structure Determinations. *J. Am. Chem. Soc.* **1989**, *111*, 3997-4004.

17. Tripos Associates, St. Louis, version 6.6.

18. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S. Jr.; Weiner, P. A New Forcefield for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc*. **1984**, *106*, 765-784.

19. Woods, R. J.; Edge, C. J.; Wormald, M. R.; Dwek, R. A. A generalized parameter set for molecular dynamics simulations of glycoproteins and oligosaccharides. Application to the structure and dynamics of a disaccharide related to oligomannose. *In* Complex Carbohydrates in Drug Research; Bock, K.; Clausen, H.; Krogsagaard-Larsen, P.; Kofod, H.; Eds.; *Alfred Benzon Symposium* **1993**, *36*, 15-26.

20. Güentert, P.; Mumenthaler, C.; Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **1997**, *273*, 283-298.

21. Rubinstein, A.; Shats, O.; Sherman, S. Role of the Local Interactions in Intrinsic Conformational and Structural Propensitites of Amino Acid Residues in Proteins. *Seventh Electronic Computational Chemistry Conference (ECCC-7)*, **2001**, http://bims.unmc.edu/ECCC7.

22. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain conformations. *J. Mol. Biol.* **1963**, *7*, 95-99.

23. Wishart, D. S.; Sykes, B. D.; Richards, F. M. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, *31*, 1647-1651.