Glycoproteins

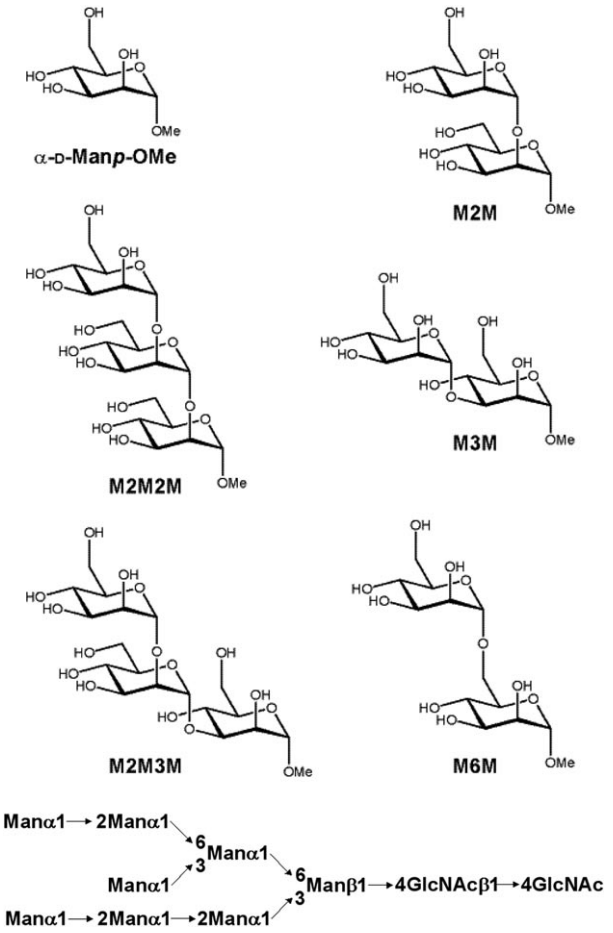
DOI: 10.1002/ange.201008258

Glycan Structure of a High-Mannose Glycoprotein from Raman Optical Activity**

*Christian Johannessen, Robert Pendrill, Göran Widmalm, Lutz Hecht, and Laurence D. Barron**

The structural analysis of glycoproteins, which play many fundamental roles in biochemistry[1a] and are of great interest to the pharmaceutical and biotechnology industries, is difficult. In the few cases for which X-ray crystal or solution NMR structures of intact or partially deglycosylated glycoprotein structures have been obtained, most of the oligosaccharide structure is poorly defined.[1b] Vibrational Raman optical activity (ROA)[2] is a promising chiroptical technique to address this problem. It is a powerful probe of biomolecular structure in aqueous solution,[3] with protein[4] and carbohydrate[5] ROA spectra exhibiting rich and informative band patterns. Preliminary ROA studies of intact glycoproteins such as α_1-acid glycoprotein (AGP) demonstrated that bands originating in the polypeptide and carbohydrate components can be identified.[6]

Here we report an ROA study of the glycan structure of intact yeast external invertase (*Saccharomyces cerevisiae*), a high-mannose glycoprotein used as a biocatalyst in the sugar industry. It consists of approximately 140-kDa subunits with over 50% of its mass being carbohydrates distributed as families of $Man_{8-14}GlcNAc_2$ and $Man_{>50}GlcNAc_2$ oligosaccharide chains N-linked to asparagine.[7] To facilitate the analysis, pertinent mono-, di- and trisaccharides were studied, namely α-D-Man*p*-OMe, α-D-Man*p*-(1→2)-α-D-Man*p*-OMe (M2M), α-D-Man*p*-(1→3)-α-D-Man*p*-OMe (M3M), α-D-Man*p*-(1→6)-α-D-Man*p*-OMe (M6M), α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-α-D-Man*p*-OMe (M2M2M), and α-D-Man*p*-(1→2)-α-D-Man*p*-(1→3)-α-D-Man*p*-OMe (M2M3M). Their structures, together with the linkage scheme of the $Man_8GlcNAc_2$ unit, are displayed in Scheme 1.



Scheme 1.

Figure 1 a–d displays the Raman and ROA spectra of three disaccharides with glycosidic linkage types common in high-mannose glycoproteins, namely M2M, M3M and M6M, together with the spectra of the monosaccharide α-D-Man*p*-OMe. Above 800 cm^{-1}, the overall ROA (and Raman) band patterns are similar, albeit with some variations in relative band intensities, reflecting identical anomeric configurations of the mannose units. Below 800 cm^{-1}, the ROA band patterns of the three disaccharides are quite different from that of the monosaccharide suggesting that, like in the case of glucose-based disaccharides[5c] (but with different ROA band patterns), the low-wavenumber bands are characteristic of the corresponding glycosidic linkages, with small differences of detail reflecting the different types of α-linkage.

Figure 2 a,b displays the Raman and ROA spectra of the trisaccharides M2M2M and M2M3M. Their ROA band patterns below 800 cm^{-1} are more similar to each other than are the patterns for the M2M, M3M, and M6M disaccharides in Figure 1, implying less conformational freedom around the glycosidic linkages in the trisaccharides than in the disaccharides, possibly due to cooperatively hydrogen-bonded networks mediated by water.[8]

The Raman and ROA spectra of yeast external invertase are displayed in Figure 3 a. The negative–positive ROA band

[*] Dr. C. Johannessen, Dr. L. Hecht, Prof. Dr. L. D. Barron
Department of Chemistry
University of Glasgow, Glasgow G12 8QQ (UK)
Fax: (+44) 141-330-4888
E-mail: Laurence.Barron@glasgow.ac.uk
R. Pendrill, Prof. Dr. G. Widmalm
Department of Organic Chemistry, Arrhenius Laboratory
Stockholm University, S-106 91 Stockholm (Sweden)

[**] This work was supported by grants from the UK Engineering and Physical Sciences Research Council, the Swedish Research Counci (VR), and the Knut and Alice Wallenberg Foundation.

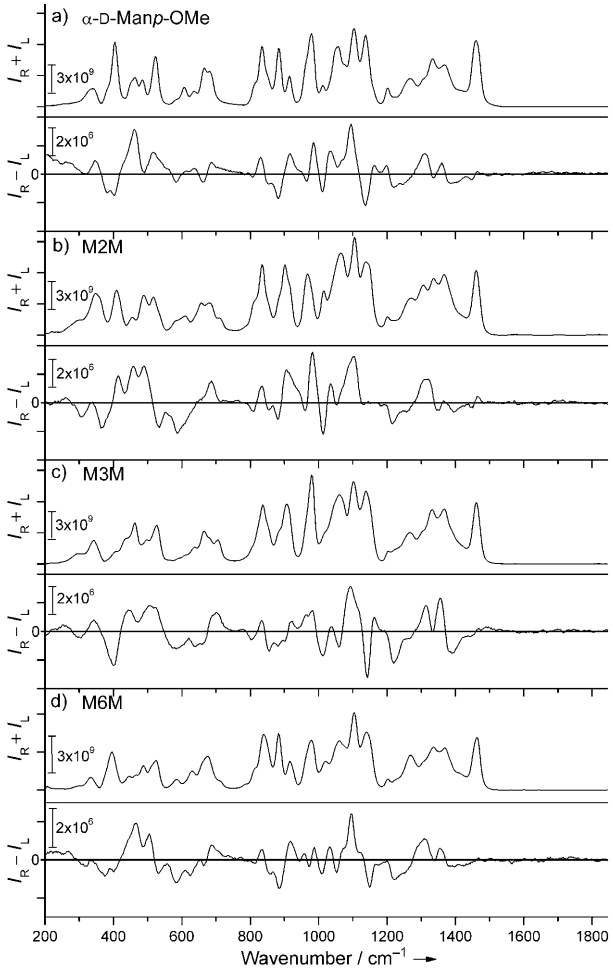


Figure 1. Backscattered SCP Raman ($I_R + I_L$) and ROA ($I_R - I_L$) spectra of a) α-D-Man*p*-OMe, b) α-D-Man*p*-(1→2)-α-D-Man*p*-OMe, c) α-D-Man*p*-(1→3)-α-D-Man*p*-OMe, and d) α-D-Man*p*-(1→6)-α-D-Man*p*-OMe in water.

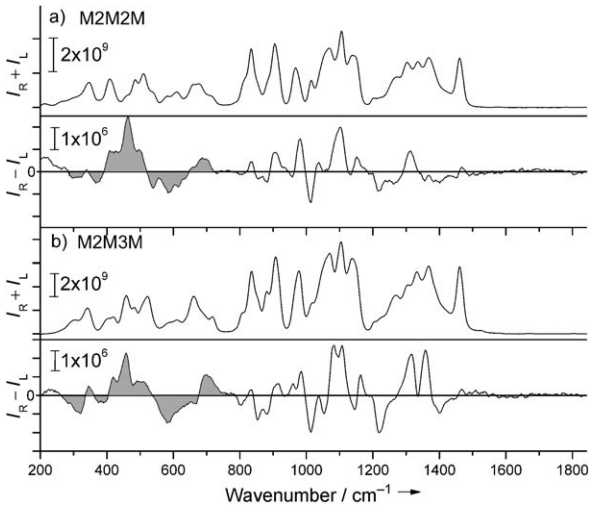


Figure 2. Backscattered SCP Raman ($I_R + I_L$) and ROA ($I_R - I_L$) spectra of a) α-D-Man*p*-(1→2)-α-D-Man*p*-(1→2)-α-D-Man*p*-OMe and b) α-D-Man*p*-(1→2)-α-D-Man*p*-(1→3)-α-D-Man*p*-OMe in water. Similar ROA bands in the low-wavenumber regions are highlighted by shading.

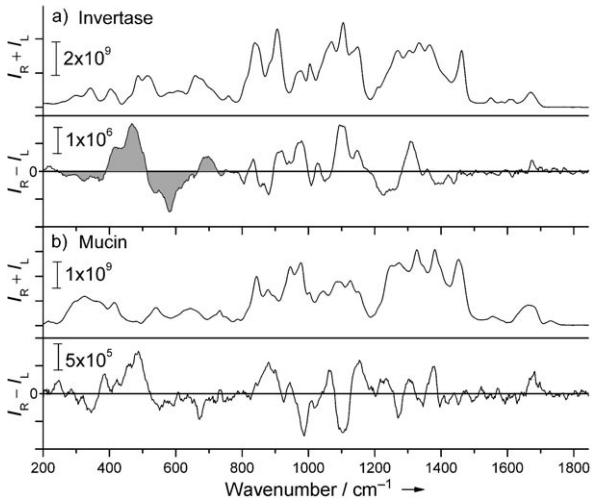


Figure 3. Backscattered SCP Raman ($I_R + I_L$) and ROA ($I_R - I_L$) spectra of a) yeast external invertase and b) bovine submaxillary mucin in water. The shaded ROA band pattern in the low-wavenumber region of the invertase spectrum is very similar to the shaded patterns in the ROA spectra of the trisaccharides M2M2M and M2M3M in Figure 2.

pattern centered at 1290 cm^{-1} in invertase falls within the extended amide III region (1230–1350 cm^{-1}) of proteins and may consequently suggest some secondary polypeptide structure.[4] However, all the mono-, di-, and trisaccharides discussed above show very similar ROA (and Raman) band patterns in this region. Furthermore, the absence of any significant ROA intensity in the amide I region (1630–1700 cm^{-1}) is powerful evidence that little, if any, α-helix or β-sheet is present,[4] with most of the ROA features of the glycoprotein originating in the glycan moieties. The ROA (and Raman) band pattern of invertase below 800 cm^{-1} is remarkably similar to those of the two trimannoses in Figure 2 and hence almost certainly originates in the α-(1→2)- and α-(1→3)-glycosidic linkages within the glycans with some minor contributions from the α-(1→6)-linkage. This close similarity also suggests that the conformations around these glycosidic links in the glycoprotein are very similar to those of the free trimannoses in aqueous solution.

The absence of signatures of secondary polypeptide structure in the ROA spectrum of invertase suggests that the high glycosylation forces the polypeptide into a completely disordered state. This interpretation is supported by the case of mucin, another highly glycosylated glycoprotein which is known to exist in an extended disordered state when fully glycosylated in aqueous solution but which adopts a globular shape when deglycosylated.[9a] Indeed, like the invertase, the ROA spectrum of bovine submaxillary mucin in aqueous solution (Figure 3b) shows little evidence of any secondary polypeptide structure, in agreement with a vibrational circular dichroism study;[10] most bands may be assigned to the glycans but with different band patterns since mucin is O-linked and is not of the high-mannose type. Furthermore, the ROA spectra of both invertase and mucin do not show the strong positive ROA band at 1320 cm^{-1} [11a] characteristic of the rather flexible poly(L-proline) II conformation, a major structural element in the peptide back-

bones of intrinsically disordered proteins.[11b,c] This indicates that the disordered polypeptide backbones in the two proteins imposed by the carbohydrate chains may be relatively stiff, something supported by biophysical data[9b] in the case of mucins.

Our study has reinforced the earlier indications[6] that ROA is a powerful probe of both glycan and polypeptide structure of intact glycoproteins, but provides more information about the glycan structure due to the availability of spectral data from key constituent di- and trisaccharides forming the high-mannose chains of yeast external invertase. It has also revealed a clear ROA signature of high-mannose glycoproteins. The way in which the glycan and polypeptide constituents modulate each other's structure is illustrated by the observation that the high glycosylation suppresses secondary polypeptide structure. The additional stability of the free trisaccharides in water relative to the disaccharides, ascribed to water-mediated hydrogen bonded networks, is maintained in the glycan chains, which accords with molecular mechanics studies showing that the flexibility of a large oligosaccharide is less than would be predicted from the inherent flexibilities of the individual linkages and may depend on the solvent properties.[1b]

Experimental Section

The samples of α-D-Man*p*-OMe, yeast external invertase (grade VII) and bovine submaxillary mucin (type S) were supplied by Sigma–Aldrich and used without further purification.

The disaccharides M2M, M3M, and M6M were available from previous studies.[12] The trisaccharides[13] M2M2M and M2M3M were synthesized using *N*-iodosuccinimide/silver triflate-promoted glycosylations[14] using suitably protected donor and acceptor molecules, namely ethyl 2,3,4,6-tetra-*O*-benzoyl-1-thio-α-D-mannopyranoside, ethyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-1-thio-α-D-mannopyranoside, methyl 3,4,6-tri-*O*-benzyl-α-D-mannopyranoside, and methyl 2-*O*-benzyl-4,6-*O*-benzylidene-α-D-mannopyranoside. Subsequent workup followed by deprotection and purification by gel-permeation chromatography on an ÄKTApurifier equipped with a SuperdexTM 30 column, prep. grade gel (GE Healthcare, Uppsala, Sweden), using water containing 1 % *n*-butanol as eluent gave after lyophilization the target trisaccharides. The identity of the trisaccharides was confirmed by phase-sensitive $^{1}H,^{1}H$-DQF-COSY, $^{1}H,^{13}C$-HSQC, $^{1}H,^{13}C$-H2BC, and $^{1}H,^{13}C$-HMBC NMR experiments in D_2O at 303 K for M2M2M and at 288 K for M2M3M. The α-anomeric configurations of the mannopyranosyl residues in the trisaccharides were deduced from the magnitudes of $J_{H1,H2}$ and $J_{H1,C1}$ being ca. 1.8 Hz and ca. 172 Hz, respectively. ESI-MS of trisaccharides: *m/z* $[M+Na]^+$ calcd for $C_{19}H_{34}O_{16}Na$ 541.1739, found 541.1754 for M2M2M, found 541.1752 for M2M3M.

The ROA spectra were measured at ambient temperature in water using the previously described ChiralRAMAN instrument (BioTools, Inc.),[2d] which employs the scattered circular polarization (SCP) measurement strategy. The ROA spectra are presented as circular intensity differences ($I_R - I_L$) and the parent Raman as circular intensity sums ($I_R + I_L$), with I_R and I_L denoting the Raman intensities with right- and left-circular polarization, respectively. The sample concentrations were 300 mM for α-D-Man*p*-OMe, 150 mM for the disaccharides, 100 mM for the trisaccharides, and ca. 50 mg mL^{-1} for the proteins. Experimental conditions: laser wavelength 532 nm; laser power at the sample 350 mW; spectral resolution 10 cm^{-1}; acquisition times 33 h for mucin and 17 h for the other samples. Solvent water spectra were subtracted from the parent Raman spectra and all spectra were subsequently smoothed using a second level Savitzky-Golay filter.

Received: December 30, 2010
Published online: April 26, 2011

Keywords: glycoproteins · glycosylation · invertase · Raman optical activity

[1] a) R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683 – 720; b) M. R. Wormald, A. J. Petrescu, Y. L. Pao, A. Glithero, T. Elliot, R. A. Dwek, *Chem. Rev.* **2002**, *102*, 371 – 386.

[2] a) L. D. Barron, M. P. Bogaard, A. D. Buckingham, *J. Am. Chem. Soc.* **1973**, *95*, 603 – 605; b) L. D. Barron, *Molecular Light Scattering and Optical Activity*, 2nd ed., Cambridge University Press, Cambridge, **2004**; c) J. Haesler, I Schindelholz, E. Riguet, C. G. Bochet, W. Hug, *Nature* **2007**, *446*, 526 – 529; d) L. D. Barron, A. D. Buckingham, *Chem. Phys. Lett.* **2010**, *492*, 199 – 213.

[3] a) L. D. Barron, L. Hecht, E. W. Blanch, A. F. Bell, *Prog. Biophys. Mol. Biol.* **2000**, *73*, 1 – 49; b) L. D. Barron, *Curr. Opin. Struct. Biol.* **2006**, *16*, 638 – 643.

[4] a) L. D. Barron, F. Zhu, L. Hecht, N. W. Isaacs in *Methods in Protein Structure and Stability Analysis. Part B. Vibrational Spectroscopy* (Eds.: V. N. Uversky, E. A. Permyakov), Nova Science, New York, **2007**, pp. 27 – 68; b) F. Zhu, G. E. Tranter, N. W. Isaacs, L. Hecht, L. D. Barron, *J. Mol. Biol.* **2006**, *363*, 19 – 26.

[5] a) F. Zhu, N. W. Isaacs, L. Hecht, G. E. Tranter, L. D. Barron, *Chirality* **2006**, *18*, 103 – 115; b) Z. Q. Wen, L. D. Barron, L. Hecht, *J. Am. Chem. Soc.* **1993**, *115*, 285 – 292; c) A. F. Bell, L. Hecht, L. D. Barron, *J. Am. Chem. Soc.* **1994**, *116*, 5155 – 5161; d) A. F. Bell, L. Hecht, L. D. Barron, *J. Raman Spectrosc.* **1995**, *26*, 1071 – 1074; e) N. R. Yaffe, A. Almond, E. W. Blanch, *J. Am. Chem. Soc.* **2010**, *132*, 10654 – 10655.

[6] a) A. F. Bell, S. J. Ford, L. Hecht, G. Wilson, L. D. Barron, *Int. J. Biol. Macromol.* **1994**, *16*, 277 – 278; b) F. Zhu, N. W. Isaacs, L. Hecht, L. D. Barron, *J. Am. Chem. Soc.* **2005**, *127*, 6142 – 6143.

[7] R. B. Trimble, P. H. Atkinson, *J. Biol. Chem.* **1986**, *261*, 9815 – 9824.

[8] E. C. Stanca-Kaposta, D. P. Gamblin, E. J. Cocinero, J. Frey, R. T. Kroemer, A. J. Fairbanks, B. G. Davis, J. P. Simons, *J. Am. Chem. Soc.* **2008**, *130*, 10691 – 10696.

[9] a) R. Bansil, B. S. Turner, *Curr. Opin. Colloid Interface Sci.* **2006**, *11*, 164 – 170; b) R. Bansil, E. Stanley, J. T. Lamont, *Annu. Rev. Physiol.* **1995**, *57*, 635 – 657.

[10] G. Shanmugam, P. L. Polavarapu, *Proteins Struct. Funct. Genet.* **2006**, *63*, 768 – 776.

[11] a) I. H. McColl, E. W. Blanch, L. Hecht, N. R. Kallenbach, L. D. Barron, *J. Am. Chem. Soc.* **2004**, *126*, 5076 – 5077; b) L. D. Barron, E. W. Blanch, L. Hecht, *Adv. Protein Chem.* **2002**, *62*, 51 – 90; c) F. Zhu, J. Kapitan, G. E. Tranter, P. D. A. Pudney, N. W. Isaacs, L. Hecht, L. D. Barron, *Proteins Struct. Funct. Genet.* **2008**, *70*, 823 – 833.

[12] a) P. E. Jansson, G. Widmalm, *J. Chem. Soc. Perkin Trans. 2* **1992**, 1085 – 1090; b) U. Olsson, E. Säwén, R. Stenutz, G. Widmalm, *Chem. Eur. J.* **2009**, *15*, 8886 – 8894.

[13] B. Hakkarainen, L. Kenne, M. Lahmann, S. Oscarson, C. Sandström, *Magn. Reson. Chem.* **2007**, *45*, 1076 – 1080.

[14] P. Konradsson, U. E. Udodong, B. Fraser-Reid, *Tetrahedron Lett.* **1990**, *31*, 4313 – 4431.

