High-Resolution NMR and Molecular Modeling Studies on Complex Carbohydrates: Characterization of O-β-D-Gal-(1→3)-O-β-D-Gal-(1→4)-O-β-D-Xyl-(1→0)-L-Ser, a Carbohydrate–Protein Linkage Region Fragment from Connective Tissue Proteoglycans

Bo-Young Choe,[‡] Göran C. Ekborg,[⊥] Lennart Rodén,[⊥] Stephen C. Harvey,[§] and N. Rama Krishna*[,§,†]

Contribution from the Comprehensive Cancer Center, Metabolic Diseases Research Laboratory, and the Departments of Biochemistry and Physics, University of Alabama at Birmingham, University Station, Birmingham, Alabama 35294. *Received September 24, 1990*

Abstract: High-resolution NMR spectroscopy and molecular mechanics calculations were employed to characterize the solution conformation of O-β-D-Gal-(1→3)-O-β-D-Gal-(1→4)-O-β-D-Xyl-(1→0)-L-Ser (G'GXS), a fragment from the carbohydrate–protein linkage region of serine-linked connective tissue proteoglycans. The ^1H and ^{13}C resonances were assigned by 2D NMR spectroscopy. The conformations of the individual sugars were determined from an analysis of vicinal proton coupling constants. Interresidue spatial constraints between the sugars and between xylose and serine were determined by 2D NOESY. A general molecular modeling methodology suitable for complex carbohydrates was described. It was used to generate families of conformations compatible with the NMR data. This methodology employs a set of starting conformations which are refined by molecular dynamics and energy minimization procedures together with NOESY derived distance constraints introduced in the form of a pseudoenergy function into the calculations. In the final stage, the structures are energy-minimized without NOE constraints to relieve any steric conflicts and bond length and bond angle distortions. On the basis of these calculations we have identified a single family of conformations for G'GXS that are compatible with the solution-phase NMR data.

Introduction

The connective tissue proteoglycans are complex carbohydrates that are composed of a core protein and at least one but in some proteoglycan species more than 100 covalently bound polysaccharide chains.[1,2] In addition, shorter oligosaccharides are usually also present and may be linked to the core protein via an *N*-glycosylamine linkage between *N*-acetylglucosamine and an asparagine residue or by an O-glycosidic linkage between *N*-acetylgalactosamine and threonine or serine. In most of the mammalian proteoglycan species, the polysaccharide component is linked to the core protein by an O-glycosidic linkage between xylose and serine, and two galactose residues are also part of a unique carbohydrate–protein linkage region which precedes the repeating disaccharides characteristic of these compounds.

The functions of the proteoglycans are, in part, an expression of their general physicochemical properties, notably their large molecular size and their high density of negative charges, which results from the presence of ester sulfate and uronic acid carboxyl groups in the repeating disaccharide units.[3] It has now also been recognized from some time that certain functions are due to precise interactions between specific segments of the proteoglycans and other macromolecules.[3] This is strikingly illustrated by the anticoagulant activity of heparin, which results from the binding of a unique pentasaccharide segment in the polysaccharide to antithrombin.[4] A detailed understanding of this and similar phenomena[3] can be attained only if we possess the necessary basic knowledge of the conformations of the proteoglycans in aqueous solution. Substantial progress is now being made in this general area, and a number of investigations by NMR spectroscopy have already provided valuable information about the glycosaminoglycan components of the proteoglycans as well as the closely related polysaccharide, hyaluronan.[5-15]

In a recent investigation in our laboratory,[13,14] a fragment from the carbohydrate–protein linkage region of the xylose/serine-linked proteoglycans was characterized by a combination of 2D NMR spectroscopy and molecular modeling methodology, i.e., O-β-D-Gal-(1→4)-O-β-D-Xyl-(1→0)-L-Ser (GXS). The present study extends this approach to a larger fragment, i.e., O-β-D-Gal-(1→3)-O-β-D-Gal-(1→4)-O-β-D-Xyl-(1→0)-L-Ser (G'GXS) (Figure 1), and we have identified a single family of conformations that are compatible with the solution-phase NMR data on G'GXS.

Structural determination of biological macromolecules by the combined use of 2D NMR and molecular mechanics calculations has now become a well-established method.[16-21] We have recently

* Address correspondence to this author at the Comprehensive Cancer Center.
† Comprehensive Cancer Center.
‡ Department of Physics.
§ Department of Biochemistry.
⊥ Metabolic Diseases Research Laboratory.

(1) Heinegård, D.; Paulsson, M. In *Extracellular Matrix Biochemistry*; Piez, K. A., Reddi, A. H., Eds.; Elsevier: 1984; p 277.
(2) Rodén, L. In *The Biochemistry of Glycoproteins and Proteoglycans*, Lennarz, W. J., Ed.; Plenum: 1980; p 267.
(3) Comper, W. D.; Laurent, T. C. *Physiol.* **1978**, *58*, 255.
(4) Lindahl, U.; Backstrom, G.; Thunberg, L.; Leder, I. G. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6551.
(5) Cowman, M. K.; Cozart, D.; Nakanighi, K.; Balazs, E. A. *Arch. Biochem. Biophys.* **1984**, *230*, 203.
(6) Ekborg, G. C.; Klinger, M.; Rodén, L.; Jensen, J. W.; Schutzbach, J. S.; Huang, D. H.; Krishna, N. R.; Anantharamaiah, G. M. *Glycoconjugate J.* **1987**, *4*, 255.
(7) Ekborg, G. C.; Curenton, T.; Krishna, N. R.; Rodén, L. J. *Carbohydr. Chem.* **1989**, *9*, 15.
(8) Heatley, F.; Scott, J. E. *Biochem. J.* **1988**, *254*, 489.
(9) Holme, K. R.; Perlin, A. S. *Carbohydr. Res.* **1989**, *186*, 301.
(10) Sugahara, K.; Yamashina, I.; DeWaard, P.; Van Halbeek, H.; Vliegenthart, J. F. G. *J. Biol. Chem.* **1988**, *263*, 10168.
(11) Sweeley, C. C.; Nunez, H. A. *Ann. Rev. Biochem.* **1982**, *54*, 765.
(12) Van Halbeek, H.; Dorland, L.; Veldink, G. A.; Vliegenthart, J. F. G.; Garegg, P. J.; Norberg, T.; Lindberg, B. *Eur. J. Biochem.* **1982**, *127*, 1.
(13) Krishna, N. R.; Choe, B. Y.; Harvey, S. C. In *Computer Modeling of Carbohydrate Molecules*; French, A., Brady, J. W., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, **1990**; Chapter 14, p 227.
(14) Krishna, N. R.; Choe, B. Y.; Prabhakaran, M.; Ekborg, G. C.; Rodén, L.; Harvey, S. C. *J. Biol. Chem.* **1990**, *265*, 18256.
(15) For additional references on NMR studies of proteoglycans, see: ref 14.
(16) McCammon, J. C.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge: London, 1987.
(17) Clore, C. M.; Gronenborn, A. M.; Brunger, A. T.; Karplus, M. *J. Mol. Biol.* **1985**, *186*, 435.

0002-7863/91/1513-3743$02.50/0

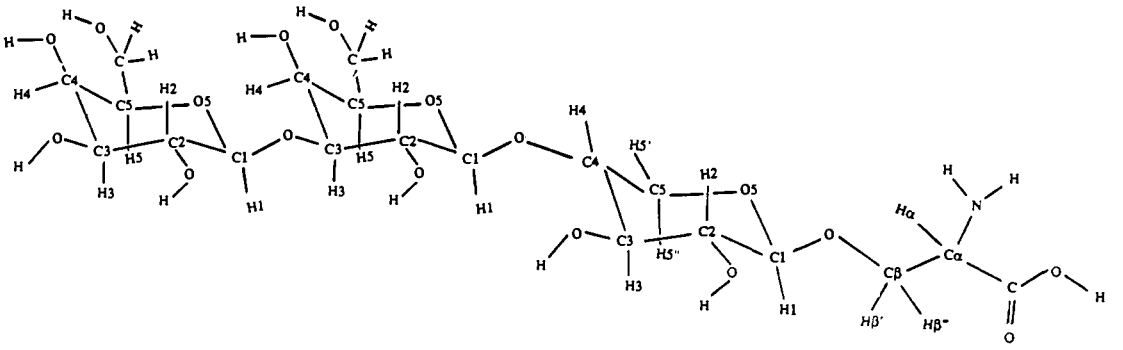

Figure 1. *O*-β-D-Gal-(1→3)-*O*-β-D-Gal-(1→4)-*O*-β-D-Xyl-(1→0)-L-Ser (G'GXS). The sugar residues are drawn in the $^4C_1$ conformations.

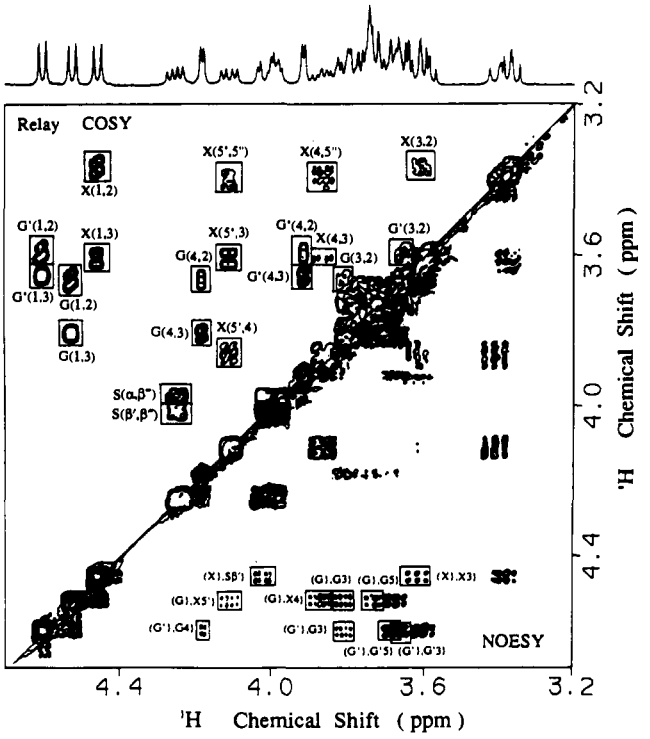

Figure 2. Combined ^1H 2D relay COSY/NOESY spectra of G'GXS. The relay COSY was obtained at 600 MHz, while the NOESY was obtained at 400 MHz with 400-ms mixing time. The corresponding 1D NMR spectrum is shown at the top. For brevity, a simplified notation is used to assign the crosspeaks. For example, G (1,3) stands for a relay COSY peak arising from the H1 and H3 protons of galactose, and (X1, Sβ') stands for a NOESY contact between the H1 of xylose and Hβ' of serine.

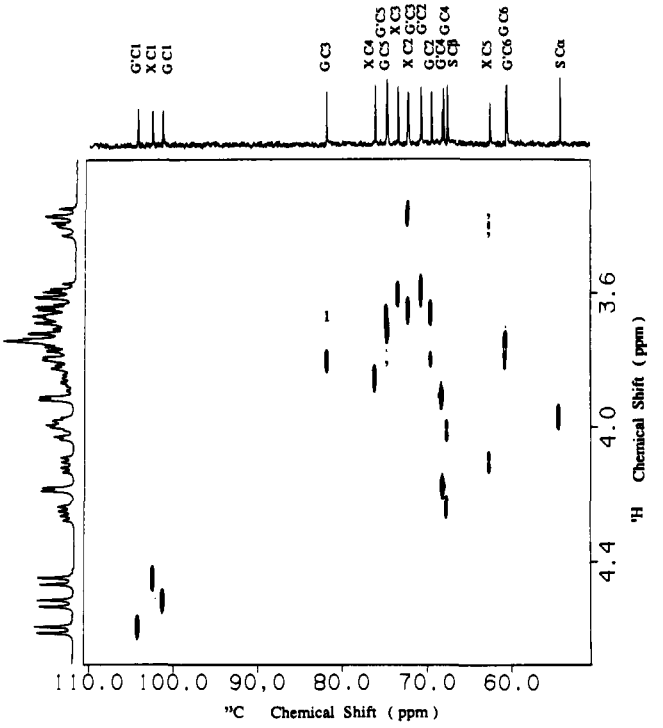

Figure 3. ^1H–^{13}C correlation NMR spectrum of G'GXS at 600 MHz. The assignments for the ^{13}C resonances are indicated at the top.

developed a molecular modeling methodology suitable for complex carbohydrates.[13,14] This methodology employs molecular dynamics (MD) and energy minimization (EM) algorithms together with NOESY derived distances constraints. In the following, we describe how this methodolgy can be applied for any complex carbohydrate, including highly branched carbohydrates. The first step in this methodology is the generation of several starting conformations, and three separate procedures are considered for this purpose. For unbranched oligosaccharides, such as G'GXS, the application of this methodology becomes particularly simple if the interactions between nonneighboring residues are negligible. Such a situation is commonly realized at the oligosaccharide level in connective tissue proteoglycans where 1→3 and 1→4 linkages are abundant. An absence of NOESY contacts between nonneighboring residues (e.g., in G'GXS) as well as a comparison of NMR data on di-, tri-, and higher oligosaccharides serve as additional criteria for testing the validity of this assumption.

Experimental Section

NMR Spectroscopy. G'GXS was synthesized as described.[7] Concentrations of 2–10 mM in D_2O for ^1H and 90 mM for ^{13}C were used. The sample pD was 6.7. Deuterium oxide was obtained from Merck Isotopes, Montreal, Canada. All NMR spectra (2D COSY,[22] 2D relayed COSY,[23] 2D *J* resolved,[24] 2D NOESY,[25] and ^1H–^{13}C correlation spectroscopy[26]) were obtained on spectrometers operating at either 600 MHz (Bruker AM-600) or 400 MHz (Bruker WH-400). These spectrometers were equipped with Aspect-3000 computers and Winchester disk drives. The experiments were all performed at ambient temperature (25 °C). The ^1H and ^{13}C chemical shifts were referenced to internal acetone (δ = 2.225 ppm) and internal dioxane (δ = 67.4 ppm), respectively. For 2D NMR spectra, 128 transients were collected for each of the 512 increments with a relaxation delay of 1.5 s between successive transients.

(18) Kaptein, R.; Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. R. *J. Mol. Biol.* **1985**, *182*, 179.
(19) Edge, C. J.; Singh, U. C.; Bazzo, R.; Taylor, G. L.; Dwek, R. A.; Rademacher, T. W. *Biochemistry* **1990**, *29*, 1971.
(20) Scarsdale, J. N.; Ram, P.; Prestegard, J. H.; Yu, R. K. *J. Comput. Chem.* **1988**, *9*, 133.
(21) For additional references, see: Scheek, R. M.; van Gunsteren, W. F.; Kaptein, R. In *Methods in Enzymology*; Oppenheimer, N. J., James, T. L., Eds.; Academic Press: 1989; Vol. 177, p 204.
(22) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229.
(23) Wagner, G. *J. Magn. Reson.* **1983**, *55*, 151.
(24) Nagayama, K.; Wuthrich, K.; Bachmann, P.; Ernst, R. R. *Biochem. Biophys. Res. Commun.* **1977**, *78*, 99.
(25) Kumar, A.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1.
(26) Bax, A. *J. Magn. Reson.* **1983**, *53*, 517.

Two-dimensional NMR spectra were obtained with quadrature detection in both dimensions. Typically, the time domain data consisted of 512 t_1 increments with 2K data points in the t_2 dimension. The data along the t_1 dimension were zero filled to 2K before 2D-Fourier transformation. The 2D-NOESY experiments were performed on the WH-400 at three different mixing times (200, 400, and 500 ms) to establish that the NOESY cross peaks were in the initial rate regime. In addition, steady-state 1D-NOE experiments were performed on some select resonances. The positive NOE enhancements observed in these experiments clearly establish that the rotational correlation times for this molecule are in the small molecule regime, and hence spin-diffusion related artifacts are not present in the NOESY experiments. All NMR data were transferred to a microVAX II computer and processed by using the FTNMR program (Hare Research, Inc., Woodinville, WA). The free-induction decays in both dimensions were multiplied by a sine bell weighting function prior to Fourier transformation. The final data matrix size of 2D spectra was 2K × 2K of real points ($\omega 1 \times \omega 2$), with digital resolution of 1.56 Hz per point for ^1H and 6.26 Hz per point for ^{13}C. Crosspeak volumes in the NOESY spectra were measured by using FTNMR. Figures 2 and 3 show typical results of 2D relay COSY, 2D NOESY, and ^1H–^{13}C correlation experiments.

MD/EM Calculations. All the calculations reported in this paper were performed with a slightly modified GROMOS package[27-29] which was used in earlier studies on cyclodextrin[30] and GXS.[13,14] The partial charges for G′GXS with all hydrogens were determined from the GAUSSIAN 80 (UCSF) program with minimal basis set and without geometrical optimization. The atomic charges for the serine residue were those of GROMOS. The amino group and the carboxyl group in serine were both assumed to be neutral so as to make the charge distribution compatible with that in the native core protein. A distance dependent dielectric constant was used in treating Coulombic interactions.[16]

We have previously described the molecular modeling protocol used in our studies on complex carbohydrates,[13,14] and here we summarize it briefly (Figure 4). Stage 1 deals with the generation of starting structures. Typically, several structures that span the (ϕ, ψ) space in a reasonable manner for each glycosidic linkage are used as starting structures.

Generation of Starting Conformations. While, in principle, one could have a starting structure at each point on a finely divided grid surface (e.g., at 10° intervals) in the (ϕ, ψ) space, we chose a total of nine starting conformations for each linkage (i.e., a total of 729 starting conformations for G′GXS and in general 9^n starting conformations for a carbohydrate with "n" linkages) to keep the computations tractable. These nine conformations are generated by assigning the torsion angles ϕ and ψ that correspond to gauche$^+$, gauche$^-$, and trans configurations across the C–O and O–C′ bonds that constitute the linkage. For a complex carbohydrate composed of "n" linkages, the number of starting conformations ideally required in our protocol will be 9^n. However, since this number can become prohibitively large even for a small oligosaccharide, we chose to develop alternative, computationally less prohibitive procedures for generating suitable starting conformations. Here we summarize three of these procedures. In all these procedures, the starting conformation for each monosaccharide unit is first determined from a vicinal coupling constant analysis. **Procedure I.** The simplest procedure can be realized when interactions between nonneighboring (i.e., nonlinked) residues in a carbohydrate are negligible. In this case, each disaccharide defined by nine starting linkage conformations can be subjected separately to the protocol in Figure 4, and the final conformation(s) for the whole carbohydrate can be constructed from that of individual disaccharides. Thus, this procedure results in $9n$ starting conformations for a carbohydrate composed of "n" linkages. Many unbranched (e.g., G′GXS used in this study) and possibly some branched complex carbohydrates may be amenable to this procedure. A brief summary of this procedure has been reported elsewhere.[13,14] **Procedure II.** To incorporate interactions between nonlinked residues such as those that are likely to occur in highly branched and possibly some unbranched complex carbohydrates, a two-step procedure may be considered. First, a set of approximate conformations using procedure I and the protocol in Figure 4 may be deduced for the whole molecule by considering only interactions between neighboring (i.e., linked) residues. Next, these approximate conformations can be used as the stage 1 starting conformations in the protocol for a second cycle of refinement in which NOESY contacts between nonlinked residues are also now incorporated (i.e., in addition to those between linked residues) as pseudoenergy terms during restrained molecular dynamics and energy minimization steps. **Procedure III.** An entirely different procedure for generating a wide variety of starting conformations will involve the adaptation of distance geometry methods[31] to complex carbohydrates. Typically, a number of experimentally determined constraints such as inter- and intra residue distances constraints (through NOESY contacts) and torsion angle constraints for individual residues and possibly for glycosidic linkages from homo- and heteronuclear vicinal coupling constants, respectively, can be used in generating a large number of distance geometry conformations for the whole carbohydrate molecule. Each of these can be then refined by using the protocol outlined in Figure 4.

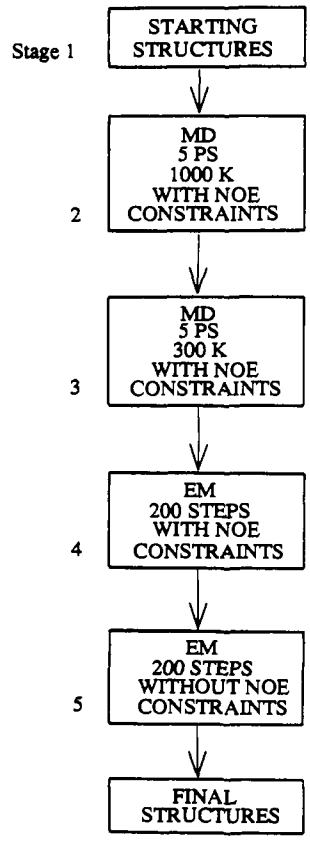

Figure 4. The protocol used in the molecular modeling studies on complex carbohydrates.

These starting structures are subjected to a high-temperature MD simulation for 5 ps (stage 2). To relate the modeled structures to experimental observables, distance constraints between hydrogen atoms were introduced in the form of pseudoenergy function, E(NOE) during the simulations. During stage 3 these structures are annealed by performing MD at 300 K for 5 ps with NOE constraints. These structures are subjected to 200 steps of conjugate gradient energy minimization with NOE constraints (stage 4). In the final stage, the NOE constraints are lifted, and the structures are subjected to 200 steps of EM by using the conjugate gradient method. These steps relieve any bond length and bond angle distortions introduced by the E(NOE) term.

On the basis of the ^1H vicinal coupling constant data, the three sugars were assumed to exist in $^4C_1$ chair conformations while defining the starting conformations. The atomic coordinates for β-D-galactose in $^4C_1$ conformation were obtained from crystallographic data.[32] The atomic coordinates for β-D-xylose in $^4C_1$ conformation were obtained from that of β-D-galactose by appropriate substitutions at the various carbons. While working with Gal′-Gal, the linkage oxygen connecting the second galactose to the xylose residue was replaced by a hydroxyl group. On the basis of ^1H vicinal coupling constant data and rotamer population analysis, the side-chain orientation of the serine residue was assigned to correspond to the "c" rotamer configuration. The glycosidic torsion angle ϕ is defined by the four atoms O5′–C1′–O–C_x, and ψ is defined by the four atoms C1′–O–C_x–C_{x-1} with right-handed rotations taken as positive torsion angles following the IUPAC convention.[33] The torsion angles

(27) van Gunsteren, W. F.; Karplus, M. *Macromolecules* **1982**, *15*, 1528.
(28) van Gunsteren, W. F.; Berendsen, H. J. C.; Hermans, J.; Hol, W. G. J.; Postma, J. P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4315.
(29) Aqvist, J.; van Gunsteren, W. F.; Leijonmarck, M.; Tapia, O. *J. Mol. Biol.* **1985**, *183*, 471.
(30) Prabhakaran, M.; Harvey, S. C. *Biopolymers* **1987**, *26*, 1087.
(31) Kuntz, I. D.; Thomason, J. F.; Oshiro, C. M. In *Methods in Enzymology*; Oppenheimer, N. J., James, T. L., Eds.; Academic Press: 1989; Vol. 177, p 159.
(32) Sheldric, B. *Acta. Crystallogr.* **1976**, *B32*, 1016.

Table I. ^1H and ^{13}C Chemical Shifts (ppm) of G'GXS

residue	proton	$^1H^a$	carbon	$^{13}C^b$
Gal'	H-1	4.61	C-1	104.21
	H-2	3.60	C-2	70.95
	H-3	3.66	C-3	72.47
	H-4	3.92	C-4	68.48
	H-5	3.69	C-5	74.98
	H-6'	3.80	C-6	61.05
	H-6''	3.77		
Gal	H-1	4.53	C-1	101.26
	H-2	3.67	C-2	69.73
	H-3	3.82	C-3	81.90
	H-4	4.19	C-4	68.38
	H-5	3.73	C-5	74.83
	H-6'	3.75	C-6	60.91
	H-6''	3.72		
Xyl	H-1	4.46	C-1	102.48
	H-2	3.38	C-2	72.55
	H-3	3.62	C-3	73.61
	H-4	3.87	C-4	76.29
	H-5' (eq)	4.12	C-5	62.92
	H-5'' (ax)	3.41		
Ser	H-α	3.98	C-α	54.54
	H-β'	4.03	C-β	67.89
	H-β''	4.25		

[a] Measured from 2D COSY spectra using Dennis Hare software. [b] Measured from 1D NMR spectrum.

Table II. Vicinal ^1H Coupling Constants (Hz) of G'GXS

residue	coupling constants[a]
Gal'	J_{12} = 7.8, J_{23} = 9.8, J_{34} = 3.5, J_{45} = 1.2, $J_{56'}$ = 7.4, $J_{56''}$ = 4.7, $J_{6'6''}$ = −12.0
Gal	J_{12} = 7.8, J_{23} = 10.1, J_{34} = 3.4, J_{45} = <1, $J_{56'}$ = 8.7, $J_{56''}$ = 5.1, $J_{6'6''}$ = −11.4
Xyl	J_{12} = 7.8, J_{23} = 9.3, J_{34} = 9.2, $J_{45'}$ = 5.3, $J_{45''}$ = 10.2, $J_{5'5''}$ = −11.8
Ser	$J\alpha\beta'$ = 3.1, $J\alpha\beta''$ = 5.6, $J\beta'\beta''$ = −11.1

[a] Measured from 2D J resolved spectroscopy.

(ϕ_1, ψ_1), (ϕ_2, ψ_2), and (ϕ_3, ψ_3), respectively, define the conformations for (Xyl-Ser), (Gal-Xyl), and (Gal'-Gal) linkages.

In addition to the normal terms for the covalent and noncovalent interactions, a NOE constraint energy term of the following form was added to the typical potential energy function:

$$E(\text{NOE}) = (K/2)(r - r_0)^2 \quad \text{for } r > r_0$$
$$= 0 \quad \text{for } r \le r_0 \qquad (1)$$

where K is the force constant (3 kcal/mol Å^2), r is the distance between two protons, and r_0 is the NOE cutoff distance. This form of potential function is identical with the one used by Kaptein et al.[18] A value of 3.5 Å has been used as a conservative estimate for r_0 in our calculations. This potential function gives equal weight to all distances less than or equal to cutoff distances and contributes a harmonic potential for distances greater than 3.5 Å.

Results

NMR Spectroscopy. Figure 2 shows the 2D relay COSY/NOESY (400 ms mixing time) spectra for G'GXS. The assignments for the crosspeaks were made in a straightforward manner from the J coupled networks within each residue. The sequence specific assignments for the sugars followed from the interresidue NOESY contacts. The ^{13}C assignments were obtained from the ^1H–^{13}C correlation spectrum (Figure 3). The vicinal ^1H coupling constants were determined from the 2D J resolved spectrum (not shown) with digital resolution of 0.1 Hz per point for the $\omega 1$ region. The ^1H and ^{13}C chemical shifts and coupling constants for G'GXS are listed in Tables I and II, respectively.

(33) IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC) Rules for the nomenclature of organic chemistry: *Pure Appl. Chem.* **1983**, *55*, 1269.

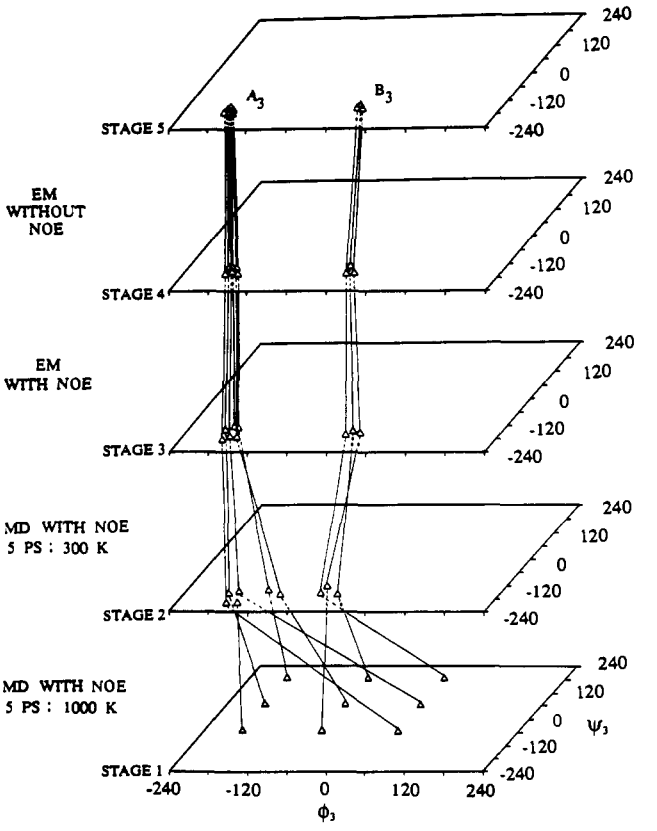

Figure 5. Refinement (from bottom to top) of the linkage torsional angles (ϕ_3, ψ_3) for the G'G linkage of G'GXS at different stages of the MD/EM protocol in Figure 4.

Our assignments based on 2D NMR are in general agreement with those of Van Halbeek et al.[12] based on 1D NMR techniques and spectral simulation. We have, however, revised their assignments for serine and for Gal-H2.

The 2D NOESY shows several intraresidue NOE contacts (Gal'-H1, Gal'-H3; Gal'-H1, Gal'-H5; Gal-H1, Gal-H3; Gal-H1, Gal-H5; Xyl-H1, Xyl-H3) expected for $^4C_1$ conformations and several interesting interresidue NOE contacts (Gal'-H1, Gal-H3; Gal'-H1, Gal-H4; Gal-H1, Xyl-H4; Gal-H1, Xyl-H5'$_{eq}$; Xyl-H1, Ser-Hβ'). All of the observed NOE contacts were used as distance constraints for the psedopotential energy function, E(NOE) (eq 1).

Under the initial rate approximation, the crosspeak intensities are proportional to r_{AB}^{-6} where r_{AB} is the distance between two protons A and B showing a NOESY contact. We have ascertained, from experiments at several mixing times, that the 400 ms mixing time NOESY experiment falls into this category. The crosspeak intensities between proton pairs were measured as volume integrals by using the FTNMR software (Hare Research Inc., Woodinville, WA). The known crystallographic distance (2.5 Å) between the H1 and H3 of *O*-β-D-galactose was used to calibrate all the other distances. Table III summarizes distances derived from the NOESY data under the assumption of a single preferred conformation for G'GXS. The validity of this assumption may be ascertained from a comparison of the derived distances with distances from theoretical models.

MD/EM Calculations. Under the assumption that the interactions between nonneighboring residues are negligible, it is permissible to subject each linkage in an oligosaccharide separately to the protocol in Figure 4 (procedure I mentioned earlier). In vacuo simulations were performed for the disaccharide Gal'-Gal. Earlier we performed similar calculations for Gal-Xyl and Xyl-Ser disaccharides which characterize the GXS fragment. The variations in the torsion angles due to refinement at different stages for Gal'-Gal are shown in Table IV. The exocyclic torsion angles, χ' (Gal') and χ (Gal), converge into two values centered around trans and gauche$^+$ in the final stage. On the other hand, the

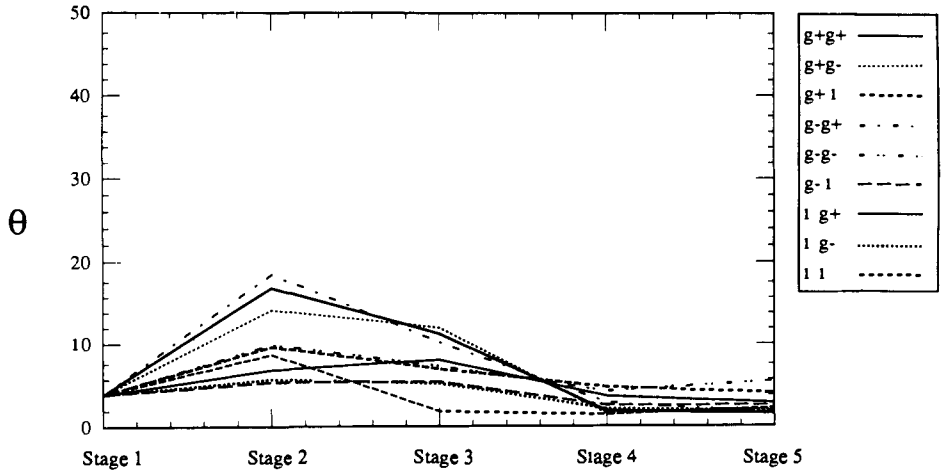

Figure 6. Changes for the puckering distortion factor θ of the terminal galactose.

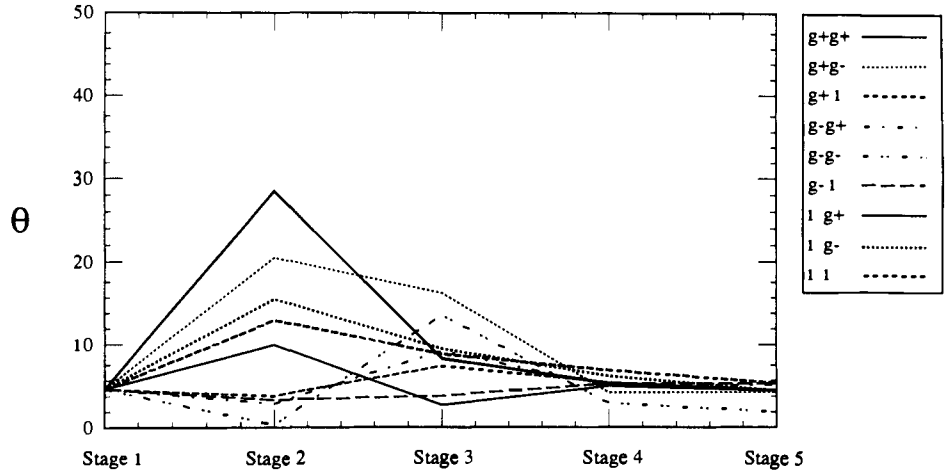

Figure 7. Changes for the puckering distortion factor θ of the intermediate galactose.

linkage torsion angles ϕ_3 and ψ_3 experience considerable variations at different stages of the MD/EM refinement. These variations are also plotted in Figure 5 to emphasize convergence of the nine starting conformations into two distinct families, A_3 and B_3. For the final structures in stage 5, the mean values for the total puckering amplitude[34] Q was 0.590 ± 0.005 (Å) and 0.582 ± 0.005 (Å) for Gal′ and Gal, and the mean values for the puckering distortion factor[34] θ for Gal′ and Gal were found to be 2.67° (Figure 6) and 4.40° (Figure 7), respectively. We have shown that the Gal-Xyl linkage conformations (ϕ_2, ψ_2) converge into two families (A_2 and B_2), while the Xyl-Ser linkage conformations (ϕ_1, ψ_1) converge into three distinct families (A_1, B_1, and C_1).

Discussion

The inter- and intraresidue NOESY contacts observed for the GXS segment in G′GXS are identical with those reported for free GXS. In addition, the vicinal coupling constants defining the conformations of Gal and Xyl and the side-chain orientation of Ser also remained virtually unchanged. We have previously studied[13,14] the GXS fragment in detail, by using the protocol outlined above, and have identified a single family of conformations that were compatible with the NMR data. Under the assumption of negligible interactions between nonneighboring residues, it is permissible to compute the conformations for the G′–G disaccharide and combine the data with our earlier calculations on GXS. The shorter fragment, GXS, was found to exist predominantly in a single family of conformations ($A_2'A_1'$) that was compatible with NMR data, with glycosidic torsion angles of ϕ_2 = −62.8°, ψ_2 = 128.3° for Gal-Xyl (A_2) and ϕ_1 = −176.2°, ψ_1 = 72.3° for Xyl-Ser (A_1). Our present calculations on the G′-G linkage show that the nine starting conformations converge into two families A_3 and B_3.

Table III. Comparison of Experimental Distances with Those Predicted for the $A_3'A_2'A_1'$ and $B_3'A_2'A_1'$ Families of G′GXS

NOE proton pairs	NOESY crosspeak volume integrals[a]	NOE derived distances[b,c]	calcd distances[b]	
			$A_3'A_2'A_1$	$B_3'A_2'A_1'$
G′H1–G′H3	35.58	2.5[d]	2.6	2.6
G′H1–G′H5	43.15	2.4	2.5	2.5
G′H1–GH3	19.47	2.8	3.0	3.4
G′H1–GH4	20.12	2.7	2.4	4.0
G′H2–GH4	*e*	*e*	4.3	2.3
GH1–GH3	30.26	2.6	2.6	2.6
GH1–GH5	32.97	2.5	2.5	2.5
GH1–XH4	27.34	2.6	2.4	2.4
GH1–XH5′	19.82	2.8	2.8	2.8
XH1–XH3	29.47	2.6	2.5	2.5
XH1–SHβ'	20.10	2.7	2.9	2.9

[a] Cross peak volumes (in arbitrary units, error ≃ ±23%) in the NOESY spectrum recorded with a mixing time of 400 ms. [b] The distances are expressed in Å. [c] Distances calculated assuming a single conformation for G′GXS (error ≃ ±0.1 Å). [d] The distance 2.5 Å for the GH1–GH3 pair was used to calibrate the volume integrals. [e] No crosspeak could be observed.

As can be seen from Figures 6 and 7, sugar repuckering was not observed in any of the simulations. The puckering distortion factor[34] θ describing the conformational state of the pyranosyl ring shows that all of the sugars always maintain the stable $^4C_1$ chair conformation. While the pyranosyl ring geometry was not significantly perturbed during even high-temperature simulations,

(34) Cremer, D.; Pople, J. R. *J. Am. Chem. Soc.* **1975**, *97*, 1354.

Table IV. Results of MD/EM Calculations on the G'G Linkage

torsion angles[a] in G'–G	stage 1	stage 2	stage 3	stage 4	stage 5	final family
ϕ_3	+60.0	−155.5	−166.6	−166.5	−172.0	
ψ_3	+60.0	−167.4	−166.7	−171.0	−173.2	A_3
χ	+178.2	+77.2	+67.4	+59.1	+58.6	
χ'	+178.2	−170.7	+171.6	+171.2	+179.8	
ϕ_3	+60.0	−158.6	−170.9	−168.8	−168.8	
ψ_3	−60.0	+170.2 (−189.8)	+179.6 (−180.4)	−173.6	−172.1	A_3
χ	+178.2	+56.0	+58.3	+58.8	+58.6	
χ'	+178.2	+160.1	+175.9	+177.7	+178.1	
ϕ_3	+60.0	−32.5	+9.6	+11.7	+34.4	
ψ	±180.0	−162.1	−169.4	−169.0	−168.9	B_3
χ	+178.2	+27.8	+71.6	+59.9	+61.6	
χ'	+178.2	+172.7	+165.6	+174.3	+174.2	
ϕ_3	−60.0	−90.7	−160.7	−166.6	−166.8	
ψ_3	+60.0	−169.9	−177.4	−173.0	−172.7	A_3
χ	+178.2	−173.6	+56.0	+56.5	+58.1	
χ'	+178.2	+57.6	−178.2	+177.8	+177.8	
ϕ_3	−60.0	−32.1	+42.9	+32.5	+34.9	
ψ_3	−60.0	−130.4	−173.6	−169.3	−169.3	B_3
χ	+178.2	+60.9	+61.6	+60.3	+61.6	
χ'	+178.2	−169.7	−177.9	+173.6	+174.2	
ϕ_3	−60.0	−3.8	+28.8	+20.7	+34.8	
ψ_3	±180.0	−175.3	−176.4	−166.6	−169.5	B_3
χ	+178.2	+73.1	+64.2	+62.8	+61.9	
χ'	+178.2	+154.7	−179.5	−177.8	+174.6	
ϕ_3	±180.0	−165.3	−161.7	−167.1	−168.6	
ψ_3	+60.0	−177.6	−173.3	−171.9	−170.4	A_3
χ	+178.2	+68.0	+63.1	+58.6	+58.7	
χ'	+178.2	+168.5	+178.1	+177.6	+177.4	
ϕ_3	±180.0	−147.1	−169.4	−168.3	−168.4	
ψ_3	−60.0	+167.0 (−193.0)	−170.1	−171.1	−170.9	A_3
χ	+178.2	+70.9	+68.7	+59.4	+59.6	
χ'	+178.2	−176.4	−157.1	+177.9	+178.1	
ϕ_3	±180.0	−113.8	−158.8	−160.3	−164.1	
ψ_3	±180.0	−155.7	−174.7	−173.1	−170.3	A_3
χ	+178.2	+157.8	+64.1	+58.7	+58.9	
χ'	+178.2	+69.8	−173.9	+176.6	+176.9	

[a] ϕ_3 (O5′–C1′–O–C3′) and ψ_3 (C1′–O–C3–C2′) define the linkage torsion angles between Gal′ and Gal. χ (O5′–C5′–C6′–O6′) and χ' (O5′–C5′–C6′–O6′) define the exocyclic torsion angles in Gal and Gal′. The angles are expressed in deg.

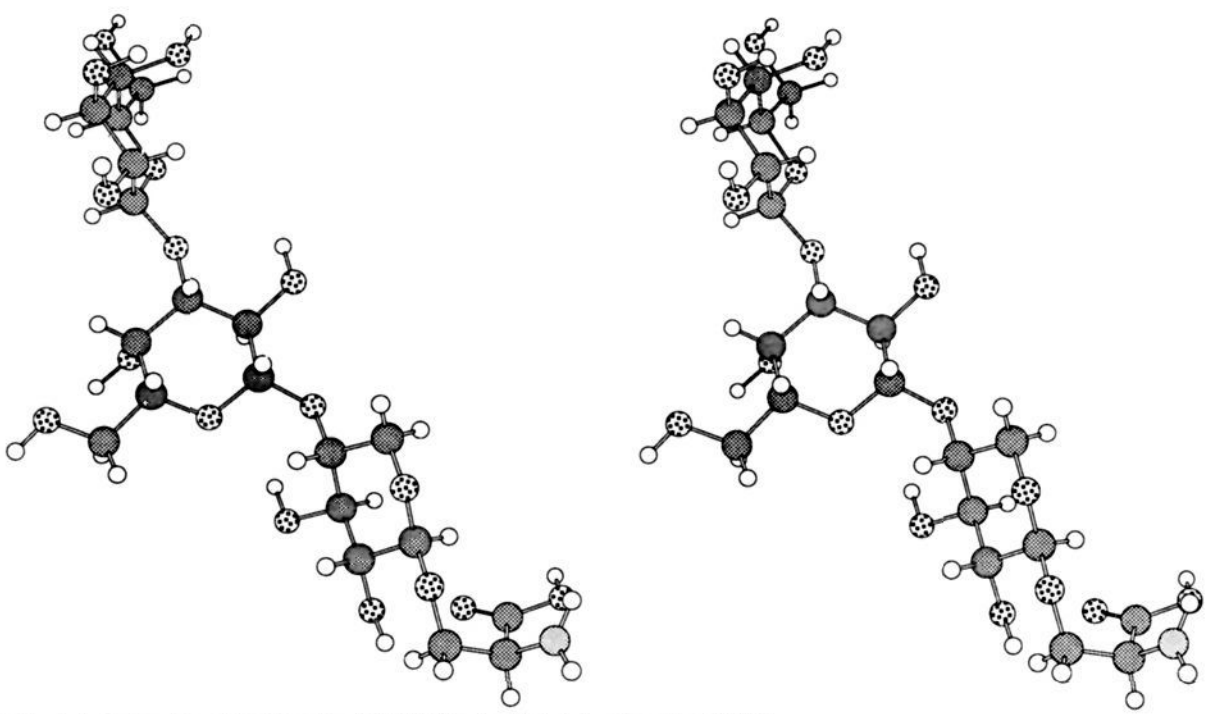

Figure 8. A typical conformation from the $A_3'A_2'A_1'$ family of final structures for G'GXS.

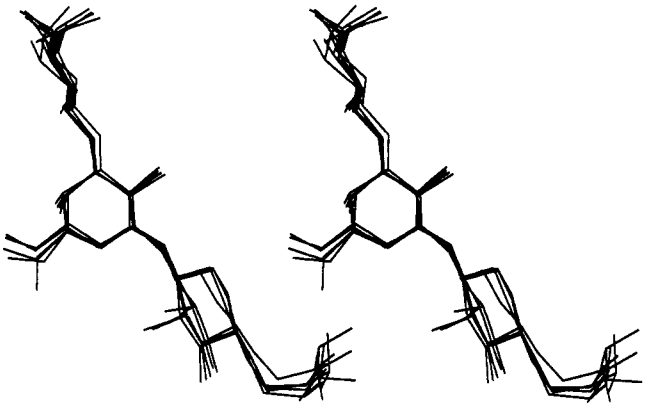

Figure 9. Structures showing the conformational flexibility of the $A_3'A_2'A_1'$ family, in Figure 8. To generate this figure, the conformation in Figure 8 was first energy minimized, then heated to 300 K for 3 ps, and equilibrated for 5 ps at this temperature, followed by a 25-ps MD simulation. A harmonic constraint with a force constant corresponding to a standard deviation of 5° for all the linkage torsion angles was employed during the MD. The figure shows a superposition of five structures sampled at 5-ps intervals during the last stage. These calculations were performed on a Silicon Graphics Iris work station using CHARMM.[36] The orientation of G'GXS is slightly different from that in Figure 8. The hydrogens were omitted for clarity.

there were significant conformational changes in the exocyclic groups. As can be seen in Table IV, the exocyclic torsion angles in Gal' and Gal converged into distinct sets in the final stage although these angles are dramatically changed in MD simulation at 1000 K like the linkage torsion angles. Since the exocyclic torsion angles of χ' and χ have different sets of values (Table IV) in the final stage, the A_3 and B_3 family for Gal'-Gal linkage can be viewed as consisting of subfamilies corresponding to these values. According to the definition of the exocyclic conformations for a sugar ring,[35] the exocyclic hydroxymethyl groups in Gal' and Gal were found to contain three different orientations (trans–gauche, TG; gauche–trans, GT; gauche–gauche, GG). While the exocyclic conformations for the terminal galactose (Gal') remained unchanged in the TG orientation at a value of around 180°, the GT orientation at a value of around +60° (gauche$^+$) was predominantly found in the intermediate galactose (Gal). The GG orientation, however, was not found in either galactose. Since the hydrogen bonding term was excluded in the present potential energy function of the slightly modified GROMOS package for in vacuo simulations, it was not surprising that no intramolecular hydrogen bond in galactoses was detected in the final stages.

(35) Ha, S. N.; Madsen, L. J.; Brady, J. W. *Biopolymers* **1988**, *27*, 1927.
(36) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

On the basis of a comparison of experimental distances with those predicted for two different families (A_3 and B_3) of Gal'-Gal linkage, only a single family (A_3) correctly reproduced the observed NOESY contacts in the final stage. Since the strong interresidue NOE contact (2.3 Å; between Gal'-H2 and Gal-H4) expected for the B_3 family was not observed in 2D NOESY, this family was ignored in constructing the model for G'GXS (see Table III). Therefore, the average linkage torsion angles of the final families (A_3 for Gal'-Gal, A_2 for Gal-Xyl, and A_1 for Xyl-Ser) for each linkage were used to construct the final structure of G'GXS. In order to relieve any steric conflicts that might arise owing to the modular construction of the whole molecule, the combined family ($A_3A_2A_1$) for G'GXS was subjected to additional energy minimization without NOE constraints. The resulting final model, $A_3'A_2'A_1'$, is representative of the conformations of G'GXS compatible with the NMR data in aqueous solution. Figure 8 shows a typical example from the final $A_3'A_2'A_1'$ sets of structures for G'GXS. Since the solution conformations are not static as in Figure 8, but are characterized by a certain degree of conformational flexibility, we have shown this flexibility in Figure 9 by a superposition of five separate structures obtained by sampling at 5-ps intervals during a 25-ps MD simulation at 300 K. A harmonic constraint with a standard deviation of 5° was used for all the linkage torsion angles. No distance constraints were used. The figure legend gives further details. The set of conformations in Figure 9 is more reflective of the solution conformation of G'GXS.

Concluding Remarks

In the present study, all the ^1H and ^{13}C resonances of G'GXS were completely assigned by 2D NMR spectroscopy at 600 MHz. The conformation of this molecule has been investigated by using a general molecular modeling methodology developed earlier. This methodology is based upon MD/EM calculations with NOE constraints introduced into the calculations to produce conformations compatible with the NMR experimental data. These calculations suggest that G'GXS exists as a single family of conformations, $A_3'A_2'A_1'$. The contribution of $B_3'A_2'A_1'$, if any, to the conformational equilibrium appears to be negligible, since no NOESY contacts predicted by this family were observable. These studies are relevant in understanding the mechanism of biosynthesis of proteoglycans.

Acknowledgment. We thank Dr. Ted Sakai for helpful discussions and Dr. Arun Malhotra for his assistance in the calculation of structures shown in Figure 9. This work was supported by the National Institutes of Health under Grants AR-39301, CA-13148, GM-34015, NS-27353, RR-03373, and DE-08252 and by the National Science Foundation under Grants DMB-8502666, BBS-8611303, and DMB-8706551. The BIOGRAF molecular modeling software supplied by BioDesign, Inc. (Pasadena, CA) through an academic grant program was used in some stages of the calculations.