Nuclear Magnetic Resonance and Molecular Modeling Studies on O-β-D-Galactopyranosyl-(1→4)-O-β-D-xylopyranosyl-(1→0)-L-serine, a Carbohydrate-Protein Linkage Region Fragment from Connective Tissue Proteoglycans*

(Received for publication, March 14, 1990)

N. Rama Krishnaf§, Bo-Young Choe|, Muthuothu Prabaharakaran‡‡, Göran C. Ekberg††, Lennart Rodén‡‡, and Stephen C. Harvey§§

From the §Comprehensive Cancer Center and the Departments of $Biochemistry, ¶Physics and ¶¶Medicine, University of Alabama, University Station, Birmingham, Alabama 35294

The solution conformation of O-β-D-galactopyranosyl-(1→4)-O-β-D-xylopyranosyl-(1→0)-L-serine (GXS), a carbohydrate-protein linkage region fragment from connective tissue proteoglycans, was investigated by two-dimensional NMR spectroscopy and molecular modeling calculations. Specifically, the 1H and 13C resonances were assigned by 2D-COSY and by 1H-13C heteronuclear correlation spectroscopy methods. 2D-NOESY was used to generate distance constraints between the galactose and xylose residues. The 1H vicinal coupling constants for the sugars and the serine were also determined. A general molecular modeling methodology suitable for complex carbohydrates was developed. This methodology employed molecular dynamics and energy minimization procedures together with the application of inter-residue spatial constraints across the linkages derived from 2D-NOESY. The first step in this methodology is the generation of a wide variety of starting conformations that span the (φ, ψ) space for each linkage. In the present study, nine such conformations were constructed for each linkage using the torsion angles φ and ψ corresponding to the gauche+, gauche−, and trans configurations across each of the two bonds constituting the linkage. These conformations were subjected to a combined molecular dynamics/energy minimization refinement using the NOESY-derived constraints as pseudoenergy functions. Families of conformations for the whole molecule were then constructed from the structures derived for each linkage. Characterization of GXS using this methodology identified a single family of conformations that are consistent with the solution phase NMR data on this molecule.

Proteoglycans and hyaluronan are ubiquitous constituents of the extracellular matrix in multicellular organisms, and their importance for the maintenance of normal tissue architecture has long been recognized. More recently, it has also become evident that the glycosaminoglycans and their parent proteoglycans serve important functions in intracellular events as well as in the interaction of the cells with their environment (Hök et al., 1984). Some of the known functions of the proteoglycans and hyaluronan may be viewed as an expression of the general physicochemical properties of these substances, among which their large size and polyanionic character are most prominent (Comper and Laurent, 1978). Other functions, however, are the result of specific interactions involving distinct segments of the complex carbohydrate molecules, as is strikingly illustrated by the anticoagulant activity of heparin which is due to the binding of antithrombin to a unique pentasaccharide segment in the polysaccharide (Lindahl et al., 1980). A special group of interactions is the recognition of substrate structures by the proteoglycan-synthesizing enzymes (Rodén, 1980), but knowledge in this area is still rudimentary. Regardless of the nature of the interactions, however, knowledge of the conformations of the complex carbohydrates in aqueous solution is a prerequisite for our understanding of the molecular basis of these processes. During the past decade, the structures of proteoglycans and hyaluronan have been investigated by NMR spectroscopy in several laboratories (e.g. Bociek et al., 1980; Casu, 1979; Casu et al., 1981; Cowman, 1985; Cowman et al., 1984; Ekborg et al., 1987; 1989; Fransson et al., 1978; Heatley and Scott, 1988; Holme and Perlin, 1989; Hounsell et al., 1986; Meyer et al., 1981; Mulloy and Johnson, 1987; Ragazzi et al., 1987; Sanderson et al., 1983; Scott et al., 1984; Scott, 1989; Sugahara et al., 1988; Sweeley and Nunez, 1982; Torchia et al., 1977; Van Halbeek et al., 1982, We1 et al., 1979). These investigations have been concerned with the repeating disaccharide segments of the glycosaminoglycan chains as well as with fragments from the carbohydrate-protein linkage region of the xylose/serine-linked proteoglycans.

In the present study, one of these fragments, O-β-D-galactopyranosyl-(1→4)-O-β-D-xylopyranosyl-(1→0)-L-serine (GXS), was characterized by 2D-NMR spectroscopy and molecular modeling calculations. In our work, 2D-NMR spectroscopy was employed to deduce information about the conformations of the individual sugars and to determine inter-residue distance constraints. Molecular modeling using those...
constraints in a combined energy minimization (EM) and molecular dynamics (MD) approach, commonly called simulated annealing, was then used to find a set of models consistent with the experimental data. While the methodology can be adapted to any complex carbohydrate, unbranched oligosaccharides particularly lend themselves to a simplified analysis by this methodology since the conformation for the whole oligosaccharide can be constructed from the conformations of the individual monomers composed of 2 residues linked together (a disaccharide or a monosaccharide linked to an amino acid). Implicit in this modular analysis approach is the assumption that interactions between non-neighboring 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. Further, this situation can be easily ascertained by comparing the NMR data for di- and higher oligosaccharides. The absence of NOE contacts between non-neighboring residues in an unbranched complex carbohydrate such as in GXS is also compatible with this assumption. This methodology was used to characterize the conformational properties of GXS in solution.

MATERIALS AND METHODS

GXS was synthesized as described (Ekborg et al., 1987). The NMR experiments were performed at 25 °C on solutions of GXS in D2O (pD 6.8). The 1H chemical shifts were referenced to the residual HDO signal (4.8 ppm). The 13C chemical shifts were referenced to internal dioxane (67.4 ppm). The NMR measurements were performed at 400 MHz on a Bruker WH-400 spectrometer equipped with an Aspect-2000 computer and a Diablo-31 Drive. All the 2D-NMR experiments (2D-COSY (Aue et al., 1976), 2D-J resolved spectroscopy (Nagayama et al., 1977), 2D-NOESY (Kumar et al., 1980), and the 13C-C correlation spectroscopy) were performed with 256 τ1 and 1K τ2 data points. The spectra were plotted in the absolute value mode. The 2D-NOESY experiments were performed at two different mixing times (200 ms and 400 ms). The spectra for 2D-COSY and 2D-NOESY experiments are given in Fig. 1. The observation of positive NOE enhancement in 1D-NOE experiments (spectra not shown) demonstrates that the rotational correlation time for GXS is in the small molecule limit, and hence spin-diffusion does not play a role in determining cross-peak intensities. The 2D-NOESY spectra at 200 ms and 400 ms clearly demonstrate that the cross-peak intensities are in the initial rate regime. The various cross-peak intensities in a 400-ms NOESY spectrum were calculated as volume integrals using the Dennis Hare FTO software (Hare Research Inc.) implemented on a microVAX-II computer. In the initial rate regime, the cross-peak intensities (Irs) are proportional to (1/τrs)1.1 where τrs is the internuclear separation.

The combined use of MD and EM with constraints derived from NOEs to refine model structures for biological molecules is now well established (Clare et al., 1986; Folkers et al., 1990; Kaptein et al., 1985; Levitt, 1983; McCammon and Harvey, 1987; Scarsdale et al., 1988). This procedure is one way to build models that are consistent with the experimental results. It should be emphasized that the purpose of the potential energy function in the calculations is to guarantee reasonable stereochemistry (acceptable bond lengths and bond angles, no unacceptable atomic overlaps, and so on). In this sense, the approach is similar to refined restrained refinement of x-ray crystallographic models (Konnert and Hendrickson, 1980). Small differences in the relative energies of the various models are not significant; any model that satisfies the observed distance constraints is an acceptable one, while those models that do not do so are not acceptable.

A preliminary report of the protocol used in our methodology has been presented elsewhere (Krishna et al., 1980) and is briefly described in the following. Because of the lack of a general solution to treat multiple minima in high temperature MD simulations (Ha et al., 1983), it is essential to construct several starting structures so that they can span the available conformational space in a reasonable manner (Ha et al., 1988). Hence, Stage 1 of the protocol involves the construction of several starting structures. Our initial attempts to generate such structures by subjecting an arbitrary conformation of GXS to a high temperature (1000 K) MD simulation (a 50-ps MD simulation without constraints, where random atomic velocities corresponding to 1000 K were repeatedly assigned every 5 ps) showed that this MD simulation could not overcome the barriers across the inter-residue linkages, and, as a result, the conformations tended to localize around three points in the (ϕ, ψ) space rather than sample the available space in a reasonable fashion. Ha et al. (1988) have chosen, in their simulation of maltose, starting conformations defined on a 20° grid in the (ϕ, ψ) space, and these were subsequently energy-minimized. In our approach, we chose a total of nine starting conformations over the (ϕ, ψ) space for each inter-residue linkage (i.e., corresponding to 81 starting conformations for GXS, and in general 9° starting conformations for an oligosaccharide with "n" linkages) and subjected each of the starting conformations to high temperature MD evolution (1000 K, 5 ps). The starting conformations were defined by assigning the three torsion angle values corresponding to gauche+, gauche−, and trans configurations across the C-O and O-C' bonds defining the linkage. Our choice of 9 starting conformations for each linkage was dictated by our desire to limit the computations to a tractable number, while assuring that enough conformations were selected to span the (ϕ, ψ) space in a reasonable manner. Clearly, the protocol can be used for any arbitrary number of starting conformations spanning the (ϕ, ψ) space.

In order to relate the modeled structures to experimental observations during the high temperature MD simulation, spatial constraints between hydrogen atoms were included in the form of a potential energy function (see below). The resulting structures (Stage 2) were annealed by an additional 5 ps MD simulation at 300 K with NOE constraints to arrive at the set of structures in "Stage 3." These structures were subjected to 200 steps of energy minimization using the conjugate gradient method. The NOE constraints were retained during this step to produce models compatible with NOE data. The resulting set of nine structures (Stage 4) were then subjected to 200 steps of energy minimization without NOE constraints. This last step reduces the strains in the structures introduced by the NOE pseudoenergy function and produces models (Stage 5) with acceptable stereochimistry in terms of bond lengths and bond angles.

The parameter values of the potential energy function used in our calculations on GXS were the same as those used in an earlier study of cyclodextrin (Prabhatkaran and Harvey, 1987), which were derived from GROMOS (Aqvist et al., 1985; Van Gunsteren et al., 1983) with slight modification, as follows. The partial charges for a single sugar unit including the hydrogens were calculated (Prabhatkaran and Harvey, 1987) using the Gaussian 90 (UCSF) program with minimal basis set and without geometrical optimization. The atomic charges for the serine residue were those of GROMOS (Van Gunsteren et al., 1983). The amino group of serine was assumed to be deprotonated and the carboxyl was assumed to be protonated to remove the positive and negative charges so as to make the charge distribution compatible with that in the native GXS. Table I lists the pseudoenergy charges for the GXS molecule used in the computer modeling studies. A distance-dependent dielectric constant was used in the calculations (McCammon and Harvey, 1987).

In addition to the usual terms for the covalent and noncovalent interactions, a constraint energy penalty term was added to the potential function for some of the MD/EM runs in the protocol described above. A semiharmonic potential of the following form was used for this term (Kaptein et al., 1985):

$$E_{\text{NOE}} = K/2 (r - r_0)^2 \quad \text{for } r > r_0$$
$$E_{\text{NOE}} = 0 \quad \text{for } r \leq r_0$$

where $K$ is the force constant (3 kcal/mol Å² in the present calculations), $r$ is the interproton distance, and $r_0$ is the NOE cutoff distance. A value of 3.5 Å has been used as a conservative estimate for $r_0$ in our calculations. This constraint energy term was used for all proton pairs that showed NOE contacts.

RESULTS

NMR Spectroscopy—The assignment of the 1H resonances from the Gal, Xyl, and Ser residues based on the sequential connectivities of cross-peaks in the 2D-COSY spectrum is straightforward and some of the assignments are indicated in Fig. 1. The 13C resonances were assigned from the 1H-13C correlation spectrum. The 1H and 13C chemical shifts for GXS based on these experiments are listed in Table II. Our assign-
ments for the two sugars in GXS based on 2D-NMR methods are in general agreement with those reported by Van Halbeek et al. (1982), based on 1D-NMR techniques and spectral simulation. We have, however, revised their assignment for the serine residue and have assigned the multiplet structure centered at 3.97 ppm assigned to the other \( \beta \)-hydrogen. The multiplet structure at 3.92 ppm was assigned to the \( \alpha \)-hydrogen. These proton assignments for serine are consistent with the vicinal coupling constant data in Table III, where the two \( \beta \)-hydrogens show a large geminal coupling constant, as expected. This assignment for the serine residue is further confirmed by the \( ^1H\)–\( ^13C \) correlation experiment which showed that the proton multiplets at 4.22 ppm and 3.97 ppm are coupled to a single carbon \((\text{C-}\beta) \) at 68.67 ppm, whereas the multiplet at 3.92 ppm is coupled to the \( \alpha \) carbon at 56.35 ppm.

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Residue} & \text{Atom} & \text{Atom form} & \text{Atom type} & \text{Sequence number} & \text{Atom partial charge} \\
\hline
\text{Gal} & \text{C}^1 & \text{C1} & \text{CS1} & 1 & 0.199 \\
\text{Gal} & \text{H}^1 & \text{H1} & \text{HC} & 2 & 0.050 \\
\text{Gal} & \text{C}^2 & \text{C2} & \text{CS1} & 3 & 0.036 \\
\text{Gal} & \text{O} & \text{O2} & \text{OA} & 4 & -0.386 \\
\text{Gal} & \text{H} & \text{H9} & \text{HO} & 5 & 0.268 \\
\text{Gal} & \text{H}^2 & \text{H2} & \text{HC} & 6 & 0.003 \\
\text{Gal} & \text{C}^3 & \text{C3} & \text{CS1} & 7 & 0.065 \\
\text{Gal} & \text{O} & \text{O3} & \text{OA} & 8 & -0.356 \\
\text{Gal} & \text{H} & \text{H10} & \text{HO} & 9 & 0.232 \\
\text{Gal} & \text{H}^3 & \text{H3} & \text{HC} & 10 & 0.062 \\
\text{Gal} & \text{C}^4 & \text{C4} & \text{CS1} & 11 & 0.056 \\
\text{Gal} & \text{O} & \text{O4} & \text{OA} & 12 & -0.300 \\
\text{Gal} & \text{H} & \text{H11} & \text{HO} & 13 & 0.341 \\
\text{Gal} & \text{H}^4 & \text{H4} & \text{HC} & 14 & 0.069 \\
\text{Gal} & \text{C}^5 & \text{C5} & \text{CS1} & 15 & 0.061 \\
\text{Gal} & \text{H}^5 & \text{H5} & \text{HC} & 16 & 0.059 \\
\text{Gal} & \text{C}^6 & \text{C6} & \text{CS2} & 17 & 0.020 \\
\text{Gal} & \text{O} & \text{O6} & \text{OA} & 18 & -0.529 \\
\text{Gal} & \text{H} & \text{H12} & \text{HO} & 19 & 0.385 \\
\text{Gal} & \text{H}^6 & \text{H6} & \text{HC} & 20 & 0.081 \\
\text{Gal} & \text{H}^7 & \text{H7} & \text{HC} & 21 & 0.051 \\
\text{Gal} & \text{O}^5 & \text{O5} & \text{OS} & 22 & -0.270 \\
\text{Xyl} & \text{O} & \text{O40} & \text{OS} & 23 & -0.266 \\
\text{Xyl} & \text{C}^1 & \text{C10} & \text{CS1} & 24 & 0.212 \\
\text{Xyl} & \text{H}^1 & \text{H8} & \text{HC} & 25 & 0.045 \\
\text{Xyl} & \text{C}^2 & \text{C20} & \text{CS1} & 26 & 0.057 \\
\text{Xyl} & \text{O} & \text{O20} & \text{OA} & 27 & -0.333 \\
\text{Xyl} & \text{H} & \text{H22} & \text{HO} & 28 & 0.223 \\
\text{Xyl} & \text{H}^2 & \text{H21} & \text{HC} & 29 & 0.100 \\
\text{Xyl} & \text{C}^3 & \text{C30} & \text{CS1} & 30 & 0.168 \\
\text{Xyl} & \text{O} & \text{O30} & \text{OA} & 31 & -0.391 \\
\text{Xyl} & \text{H} & \text{H32} & \text{HC} & 32 & 0.243 \\
\text{Xyl} & \text{H}^3 & \text{H31} & \text{HC} & 33 & -0.053 \\
\text{Xyl} & \text{C}^4 & \text{C40} & \text{CS1} & 34 & 0.101 \\
\text{Xyl} & \text{H}^4 & \text{H41} & \text{HC} & 35 & 0.008 \\
\text{Xyl} & \text{C}^5 & \text{C50} & \text{CS1} & 36 & 0.002 \\
\text{Xyl} & \text{H}^5 & \text{H51} & \text{HC} & 37 & 0.088 \\
\text{Xyl} & \text{H}^5 & \text{H52} & \text{HC} & 38 & 0.043 \\
\text{Xyl} & \text{O}^5 & \text{O10} & \text{OS} & 39 & -0.267 \\
\text{Ser} & \text{O} & \text{O60} & \text{OS} & 40 & -0.257 \\
\text{Ser} & \text{C}\beta & \text{C60} & \text{CH2} & 41 & -0.095 \\
\text{Ser} & \text{H}^6 & \text{H61} & \text{HC} & 42 & 0.110 \\
\text{Ser} & \text{H}^6 & \text{H62} & \text{HC} & 43 & 0.115 \\
\text{Ser} & \text{C} & \text{C70} & \text{CH1} & 44 & -0.004 \\
\text{Ser} & \text{H}^6 & \text{H71} & \text{HC} & 45 & 0.107 \\
\text{Ser} & \text{N} & \text{N1} & \text{NT} & 46 & -0.465 \\
\text{Ser} & \text{H} & \text{H72} & \text{H} & 47 & 0.131 \\
\text{Ser} & \text{H} & \text{H73} & \text{H} & 48 & 0.230 \\
\text{Ser} & \text{C} & \text{C80} & \text{C} & 49 & 0.381 \\
\text{Ser} & \text{O} & \text{O80} & \text{O} & 50 & -0.294 \\
\text{Ser} & \text{O} & \text{O90} & \text{OA} & 51 & -0.400 \\
\text{Ser} & \text{H} & \text{H90} & \text{HO} & 52 & 0.317 \\
\hline
\end{array}
\]

Table I
List of partial atomic charges in GXS

A rotamer population analysis (Bystrov, 1976) of the data for the serine residue indicated that it exists predominantly in the \( ^{\text{c}}\) rotamer configuration with \( x = 60^\circ \). The coupling constant data for the two sugars are in excellent agreement with the predicted values for the \( ^{13C} \) conformations of \( \beta\)-D-pyranosides (Altona and Hansnot, 1980).

Table II
\( ^1H \) and \( ^{13C} \) chemical shift (ppm) of GXS

Table III
Vicinal \( ^1H \) coupling constants (in Hz) of GXS

* Measured by 2D-J resolved spectroscopy.
The 2D-NOESY spectrum in Fig. 1 shows several interesting cross-peaks, and some of these have been identified in the figure. Some cross-peaks are intraresidue in origin while others are inter-residue (Gal-Hl', Xyl-H4'; Gal-Hl', Xyl-H5', Xyl-H1', Ser-Hβ etc.). Many of these cross-peaks are too weak to observe at 200 ms mixing time, but they begin to build up in intensity at 400 ms.

Generation of Starting Conformations (Stage 1)—On the basis of the 1H vicinal coupling constant analysis (Rikborg et al., 1987; Van Halbeek et al., 1982), both the sugars were assumed to adopt the C5 chair conformations as starting structures in Stage 1. No sugar repuckering was observed in any of the simulations (including the 50-ps simulation mentioned earlier). The exocyclic torsion angle θ (O5'-C5'-C6'-O6') for the galactose residue has been assigned a starting value of 178.2° as observed in the crystal structure for β-D-galactose (Sheldric, 1976). Since the side chain of the serine residue exists predominantly in the c rotamer configuration, a value of χ = 60° was used for all the Stage 1 conformations. In defining the stage 1 conformations, the torsion angles for each linkage (i.e. φ1, φ2, for X-S and φ1, φ2 for G-X) were given values that corresponded to the gauche+, gauche−, and trans configurations. This procedure generated 9 conformations for each linkage (See Tables IV and V).

The observed inter-residue NOE contacts, viz. Gal-H1' to Xyl-H4', Gal-H1' to Xyl-H5'eq, and Xyl-H1' to Ser-Hβ, together with the intraresidue contacts were used as distance constraints for the constraint energy function, E(NOE), as described earlier (Equation 1).

MD/EM Calculations—Under the assumption that the interactions between non-neighboring residues are negligible (i.e. between galactose and serine in the case of GXS), it is permissible to subject each linkage in an oligosaccharide separately to the protocol described above. Thus inter-residue linkages for the disaccharides X-S and G-X were subjected separately to the protocol and the resulting variations in the torsion angles due to refinement at different stages are shown in Tables IV and V, respectively. While working with G-X, the linkage oxygen connecting to the serine residue was replaced by an OH group. In a similar fashion, while performing calculations on X-S, the linkage oxygen connecting to the galactose was also replaced by an OH group. The side chain orientation of serine defined by the torsion angle χ remained relatively invariant at the various stages of the refinement (see Table IV). The exocyclic torsion angle, θ (O5'-C5'-C6'-O6') for the galactose residue grouped into two values centered around 177.2° and 60.7° in the final stage (not shown). On the other hand, the linkage torsion angles φ1, ψ1, and ψ2 experienced considerable variations at different stages of the MD/EM refinement. These variations are also plotted in Figs. 2 and 3 to emphasize the convergence of nine starting conformations into distinct families. It is noted that for X-S, the conformations converged into three distinct families, A1, B1, and C1. For the other linkage, G-X, the conformations converged into two distinct families, A2 and B2.

**DISCUSSION**

Table VI shows a comparison of the average torsion angles and the inter-residue proton distances obtained for the different families. In each case, there was only one set of conformations in the final stage that correctly reproduced the observed NOE contacts. These are the A1 conformations for X-S and the A2 conformations for G-X. The other conformations (B1, C1, and B2) represent models that are trapped in local minima of the potential energy function. Interestingly, the calculated potential energies of these other structures are 0.5-5.0 kcal/mol higher than those of conformations A1 and A2. We emphasize that these differences are not an indication that A1 and A2 conformations are “better” than the other conformations in a structural sense. The purpose of our

### Table IV

| Torsion angles in X-S° | Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 | Final family |
|------------------------|--------|--------|--------|--------|--------|--------------|
| φ1 +60.0                | −62.0  | −63.4  | −61.6  | +178.7 | −59.9  | B1           |
| ψ1 +60.0 +179.8        | +155.5 | +178.6 | +178.7 | +178.7 | +178.7 | B1           |
| χ +60.0                | +65.1  | +65.5  | +61.1  | +61.3  |        |              |
| φ1 +60.0 +165.1(−194.9)| +165.6(−194.4) | −177.7 | −178.5 |        |        |              |
| ψ1 −60.0 +92.4         | +94.7  | +68.7  | +67.9  | +58.0  |        |              |
| χ +60.0                | +69.8  | +58.9  | +57.9  | +58.0  |        |              |
| φ1 +60.0 +6.5          | +54.5  | −60.5  | −58.7  |        |        |              |
| ψ1 ±180.0 +190.3       | +134.8 | +170.0 |         |        |        |              |
| χ +60.0 +73.5          | +64.1  | +73.1  | +73.2  | +73.1  |        |              |
| φ1 +60.0 −144.5(−215.5)| +167.6(−192.3) | −175.2 | −175.8 | −175.8 |        |              |
| ψ1 +60.0 +95.2         | +96.8  | +96.8  | +61.6  | +61.6  | +61.6  | A1           |
| χ +60.0 +68.7          | −63.3  | −70.8  | −87.9  | −87.9  |        |              |
| φ1 −60.0 −140.8        | −92.2  | −76.8  | −75.7  | −75.7  |        |              |
| ψ1 +60.0 +70.0         | +64.1  | +67.3  | +66.7  |        |        |              |
| χ +60.0 −49.3          | −64.5  | −61.4  | −61.4  | −61.4  |        |              |
| φ1 ±180.0 −175.8       | +176.9 | ±180.0 | ±180.0 | ±180.0 | ±180.0 |              |
| χ +60.0 +62.7          | +58.1  | +62.1  | +62.1  | +62.1  |        |              |
| φ1 ±180.0 −120.5       | +167.4(−192.6) | −177.5 | −177.9 | −177.9 |        |              |
| ψ1 +60.0 +63.7         | +85.5  | +70.9  | +69.3  | +69.3  |        |              |
| χ +60.0 +50.9          | +61.7  | +58.7  | +57.8  |        |        |              |
| φ1 ±180.0 −110.6      | −118.8 | −172.2 | −171.7 |        |        |              |
| χ +60.0 −50.4          | +59.4  | +91.8  | +91.8  | +91.8  |        |              |
| φ1 +180.0 +64.7        | +52.0  | +59.8  | +59.8  |        |        |              |
| ψ1 ±180.0 +175.9(−184.1)| +163.9(−196.1) | +179.8(−180.2) | −177.5 |        |        |              |
| χ +60.0 +58.0          | +60.2  | +58.9  | +58.5  |        |        |              |

* φ1(O5'-C5'-O-C-P) and ψ1(C1'-O-C5'-Ca) define the linkage torsion angles between Xyl and Ser. χ(O-C5'-Ca-N) defines the side chain orientation of Ser. The angles are expressed in degrees.
TABLE V

Results of MD/EM calculations on the GX linkage

| Linkage torsion angles in G-X | Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5 | Final family |
|------------------------------|---------|---------|---------|---------|---------|--------------|
| \( \phi_1 \)                | +60.0   | -46.2   | -56.2   | -64.9   | -67.1   | A2           |
| \( \psi_1 \)                | +60.0   | +115.0  | +127.2  | +129.0  |          | B1           |
| \( \phi_1 \)                | +60.0   | -51.2   | -80.9   | -69.0   | -75.2   | B3           |
| \( \psi_2 \)                | -60.0   | -40.2   | -63.3   | -51.9   | -54.8   | A2           |
| \( \phi_2 \)                | +60.0   | -1.2    | -60.9   | -57.8   | -62.1   | A2           |
| \( \psi_3 \)                | \( \pm 180.0 \) | +122.6  | +123.2  | +123.7  | +127.6  | A2           |
| \( \phi_3 \)                | -60.0   | -48.9   | -60.9   | -57.4   | -58.4   | A2           |
| \( \psi_4 \)                | +60.0   | +118.2  | +129.9  | +124.9  | +126.8  | A2           |
| \( \psi_5 \)                | -60.0   | -77.0   | -71.8   | -77.8   | -70.0   | A2           |
| \( \phi_4 \)                | -60.0   | -67.6   | -61.4   | -55.9   | -55.1   | A2           |
| \( \psi_6 \)                | -60.0   | -55.8   | -58.9   | -57.1   | -60.7   | A2           |
| \( \psi_7 \)                | \( \pm 180.0 \) | +140.6  | +120.9  | +124.4  | +126.4  | A2           |
| \( \phi_7 \)                | \( \pm 180.0 \) | -56.9   | -60.3   | -64.2   | -64.7   | A2           |
| \( \psi_8 \)                | +60.0   | +134.5  | +119.5  | +129.3  | +128.9  | A2           |
| \( \psi_9 \)                | \( \pm 180.0 \) | -65.7   | -59.5   | -63.2   | -65.9   | A2           |
| \( \phi_9 \)                | \( \pm 180.0 \) | -158.2  | +123.8  | +131.9  | +132.2  | A2           |
| \( \psi_{10} \)             | \( \pm 180.0 \) | -78.9   | -57.3   | -56.9   | -59.6   | A2           |

\( \phi_8 (O5'-C1'-O-C4') \) and \( \psi_8 (C1'-O-C4'-C3') \) define the linkage torsion angles between Gal and Xyl. The angles are expressed in degrees.

even though the conformations of \( A_2' \) \( A_1' \) family are compatible with the observed NOE contacts in GXS, we have also examined the possibility that a conformational exchange between different families might account for the observed inter-residue NOE contacts. This is an important consideration since the neglect of such an exchange in the interpretation of NMR data could result in a virtual conformation for the oligosaccharide that is either unrealistic or present in the solution phase only in small populations (Cumming and Carver, 1987). Table VI shows a comparison of the calculated inter-residue proton distances for the families \( A_1 \), \( B_1 \), \( C_1 \), \( A_2 \), and \( B_2 \), together with the average torsion angles. The \( B_1 \) family predicts that both \( \beta' \) and \( \beta'' \) protons will experience a significant NOE contact with Xyl-H1' proton, whereas the \( C_1 \) family predicts a NOE contact between Xyl-H1' and Ser-H3'. These contacts were not observed in the experiment. The \( B_2 \) family for the G-X linkage predicts that both the H5' and...
the anomeric proton of one sugar may show NOE contacts with other molecules. We consider in the modeling calculations.

It is conceivable that in other, more complex oligosaccharides, in the intramolecular interactions as well as in the interactions of the polysaccharide chains with other molecules, such a conformational exchange should also be considered in the modeling calculations.

It is interesting that the galactose anomeric proton, Gal-\(H_1\)\(^\alpha\), shows NOE contacts not only to the \(H_4\) proton across the glycosidic linkage but also to the \(H_5\) proton of xylose. It is conceivable that in other, more complex oligosaccharides, the anomeric proton of one sugar may show NOE contacts only to protons not directly involved in the glycosidic linkage (Bush et al., 1986; Dua et al., 1986; Lemieux et al., 1980).

In conclusion, we have developed a general methodology suitable for modeling of complex carbohydrates. Unbranched oligosaccharides lend themselves to a simplified modular analysis by this procedure. This methodology is based on MD/EM calculations with NMR-derived constraints introduced into the calculations to generate conformations compatible with the experimental data.

Our approach differs from that of Scarsdale et al. (1988) in some essential details. These authors selected three starting conformations for the whole molecule versus nine starting conformations for each linkage in our approach. In general, it is desirable to have as many starting structures as possible so that the conformational space is adequately sampled during the molecular mechanics calculations. Second, the form of the pseudoenergy function chosen in our work to represent NOE constraints is similar to that used by Kaptein et al. (1985) in their protein modeling studies, but differs from that of Scarsdale et al. (1988). The latter authors have used a potential function that has a negative minimum at \(r = r_0\), approaches zero for \(r > r_0\), is positive for \(r < r_0\), and tends to infinity for \(r \rightarrow 0\). This function has the advantage of leading to rapid convergence, but the distance \(r_0\) has maximum weight because of the form of the function. The semiharmonic potential chosen in this work allows us to set an upper limit on \(r_0\) for the distance between any two protons between which a NOE was observed. We have selected a conservative value of 3.5 Å for \(r_0\). Further, Equation 1 gives equal weight to all distances less than \(r_0\), and hence the computed distances will be relatively free of any bias induced by the pseudoenergy function as long as the NOE distance constraint is satisfied. A third difference involves our final step in the protocol in which the NOE constraints were lifted and the conformations were further subjected to EM to relax an unreasonable steric and bond angle distortions resulting from the previous steps (MD/EM) that employed distance constraint penalty term \(E(\text{NOE})\). This final step becomes especially important in those calculations where the use of strong force constants for \(E(\text{NOE})\) can result in unacceptable bond length and bond angle distortions.

The conformations of the individual monosaccharide units in the connective tissue proteoglycans and the spatial relationships between adjacent units are important determinants in the intramolecular interactions as well as in the interactions of the polysaccharide chains with other molecules. We are now beginning to understand the structural basis of some of these interactions, particularly those which have readily
measurable functional consequences. The majority of the interactions examined to date involve the repeating disaccharide segments of the polysaccharide chains, e.g. the specific binding of a decasaccharide segment in hyaluronan to the core protein of the chondroitin sulfate proteoglycan of cartilage, which may be added to the examples already mentioned above (Rodén, 1980). It is presently not known whether the carbohydrate-protein linkage region of the xylose/serin-linked proteoglycans plays a role in the interaction of these macromolecules with their environment. Indeed, it is possible that the existence of the specific linkage region is a reflection of the need for metabolic regulation during the intracellular assembly of the proteoglycans. The biosynthetic pathways and various mechanisms of regulation of the assembly process have been described elsewhere (Rodén, 1980) and will not be discussed here. In closing, it should be pointed out, however, that the disaccharide component of GXS represents a unique structure in the mammalian complex carbohydrates and that both galactose and xylose, in β1,4-linkage, are required for the mechanism of action of galactosyltransferase II when the enzyme thus ensures that it operates with a high degree of fidelity and does not effect transfer to galactose residues other than the xylose-linked residue of the growing connective tissue polysaccharide chain. The basic data on the conformation of GXS reported here will aid in the understanding of the mechanism of action of galactosyltransferase 11 when the enzyme has been purified and studies of its three-dimensional structure become possible.

Acknowledgment—The use of BIOGRAF molecular modeling software supplied by BioDesign, Inc. (Pasadena, CA) through an academic grant program is gratefully acknowledged.

REFERENCES

Ałtona, C., and Haasnoot, A. G. (1980) Org. Magnetic Res. 13, 417
Aqvist, J., van Gunsteren, W. F., Leijonmarck, M., and Tapia, O. (1985) J. Mol. Biol. 183, 461
Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) J. Chem. Phys. 64, 2229
Bociek, S., Darke, A. H., Welti, D., and Rees, D. A. (1980) Eur. J. Biochem. 109, 147
Bush, C. A., Yan, Z. Y., and Rao, B. N. N. (1986) in Abstracts: XIIIth International Carbohydrate Symposium, Cornell University, Ithaca, NY
Byrd, R. A., Egan, N., Summers, M. F., and Bax, A. (1987) Carbohydr. Res. 166, 47
Bystrøv, V. F. (1986) Prog. NMR Spectroscopy 10, 41
Casu, B. (1979) Pharmacol. Aids Commun. 78, 99
Casu, B., Ore&le, P., Torri, G., Zoppetti, G., choix, J., Lormeau, J., Bartolomé, E., and Ernst, R. R. (1989) Biochemistry 28, 2601
Comper, W. D., and Laurent, T. C. (1978) Physiol. Rev. 58, 555
Cowman, M. K., in Industrial Uses of Polysaccharides (Crescenzi, V., ed) Gordon & Breach Science Publishers Inc., New York
Cowman, M. K., Cozart, D., Nakanighi, K., and Balazs, E. A. (1984) Arch. Biochem. Biophys. 230, 203
Cowman, M. K. (1986) in Carbohydrate Chemistry of Proteoglycans (Lennarz, W. J., ed) Plenum Publishing Co., New York, pp. 227-239
Djerassi, M., van Hove, J., and Hendrickson, W. A. (1980) Acta Crystallogr. A36, 314
Ekborg, G., Cureton, T., Krishna, N. R., and Rodén, L. (1989) J. Carbohydr. Chem. 9, 15
Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Chap. 9, Clarendon Press, Oxford
Folkers, P. J. M., Rees, D. A., and Welsh, E. J. (1979) Eur. J. Biochem. 94, 261
Fransson, L. Å., Huckerby, T. N., and Nieduszynski, I. A. (1978) Biochem. J. 175, 299
Ha, S. N., Madsen, L. J., and Brady, J. W. (1988) Biopolymers 27, 1927
Heathley, F., and Scott, J. E. (1988) Biochem. J. 254, 489
Hoike, K. R., and Perlin, A. S. (1989) Carbohydr. Res. 195, 301
Höök, M., Kjellén, L., and Johansson, S. (1984) Annu. Rev. Biochem. 53, 847
Hounsell, E., Feeney, J., Scudder, P., Tang, P. W., and Feizi, T. (1986) Eur. J. Biochem. 157, 375
Kaptein, R., Zuidweg, E. R. P., Scheek, R. M., Boelens, R., and Van Gunsteren, W. W. (1985) J. Mol. Biol. 182, 179
Kessler, H., Uhlen, W., Grässinger, C., and Kolar, C. (1986) Angew. Chem. Int. Ed. Engl. 25, 342
Konnert, J. H., and Hendrickson, W. A. (1980) Acta Crystallogr. A, 314
Krishna, N. R., Choe, B.-Y., and Harvey, S. C. (1990) in Computer Modeling of Carbohydrate Molecules (French, A., and Brady, J. W., eds) American Chemical Society Symposium Series, pp. 227-239
Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) Biochem. Biophys. Res. Commun. 95, 1
Lehinger, R. U., Bock, K., Delmerre, J. T. J., Kato, S., and Rao, V. G. (1980) Can. J. Chem. 58, 651
Lerner, L., and Bax, A. (1987) Carbohydr. Res. 166, 35
Levitt, M. (1983) J. Mol. Biol. 170, 723
Lindahl, V., Bäckström, G., Thunberg, L., and Leder, I. G. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 6551
McCammmon, J., C., and Harvey, S. C. (1987) Dynamics of Proteins and Nucleic Acids, Cambridge University Press, London
Meyer, B., Thunberg, L., Lindahl, V., Larm, O., and Ledder, I. G. (1981) Carbohydr. Res. 88, 81
Mulloy, B., and Johnson, E. A. (1987) Carbohydr. Res. 170, 151
Nagayama, K., Wüthrich, K., Bachmann, P., and Ernst, R. R. (1977) Biochem. Biophys. Res. Commun. 78, 99
Prabhakaran, M., and Harvey, S. C. (1987) Biopolymers 26, 1087
Ragazzi, M., Fagg, N., P., Perly, B., Torri, G., Casu, B., Sinay, P., Petitou, M., and Choy, J. (1987) Carbohydr. Res. 165, 1
Rodén, L. (1980) in The Biochemistry of Glycoproteins and Proteoglycans (Lennarz, W. J., ed) Plenum Publishing Co., New York, pp. 267-371
Sanderson, P. N., Nieduszynski, I. A., and Huckerby, T. N. (1983) Biochem. J. 211, 677
Searsdale, J. N., Ram, P., Prestegard, J. H., and Yu, R. K. (1988) J. Comput. Chem. 9, 133
Scheraga, H. A. (1983) Biopolymers 22, 1
Scott, J. E., Heathley, F., and Hull, W. E. (1984) Biochem. J. 220, 197
Sheidr, R. H. (1989) Acta Crystallogr. B32, 1016
Sugahara, K., Yamashina, I., De Waard, P., Van Halbeek, H., and Vliegenthart, J. F. G. (1988) J. Biol. Chem. 163, 10168
Torchia, D. A., Hasson, M. A., and Hascall, V. C. (1977) J. Biol. Chem. 252, 3617
Torchia, D. A., Hasson, M. A., and Hascall, V. C. (1987) J. Biol. Chem. 252, 3617
Torchia, D. A., Hasson, M. A., and Hascall, V. C. (1987) J. Biol. Chem. 252, 3617
Torchia, D. A., Hasson, M. A., and Hascall, V. C. (1987) J. Biol. Chem. 252, 3617
Torchia, D. A., Hasson, M. A., and Hascall, V. C. (1987) J. Biol. Chem. 252, 3617
Wheeler, C. C., and Nunez, H. A. (1982) Annu. Rev. Biochem. 51, 765
Van Gunsteren, W. F., Berendsen, H. J. C., Hermans, J., Hoe, W. G. J., and Postma, J. P. M. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 4315
Van Halbeek, H., Dorland, L., Veldink, G. A., Vliegenthart, J. F. G., Garez, P. J., Norbertz, T., and Lindberg, B. (1982) Eur. J. Biochem. 127, 1
Welti, D., Rees, D. A., and Welsh, E. J. (1979) Eur. J. Biochem. 94, 505
Nuclear magnetic resonance and molecular modeling studies on O-beta-D-galactopyranosyl-(1----4)-O-beta-D-xylopyranosyl-(1----0)-L-serine, a carbohydrate-protein linkage region fragment from connective tissue proteoglycans.

N R Krishna, B Y Choe, M Prabhakaran, G C Ekborg, L Rodén and S C Harvey

*J. Biol. Chem.* 1990, 265:18256-18262.

Access the most updated version of this article at [http://www.jbc.org/content/265/30/18256](http://www.jbc.org/content/265/30/18256)

Alerts:
- When this article is cited
- When a correction for this article is posted

[Click here](http://www.jbc.org/content/265/30/18256.full.html#ref-list-1) to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at [http://www.jbc.org/content/265/30/18256.full.html#ref-list-1](http://www.jbc.org/content/265/30/18256.full.html#ref-list-1)