Journal of Biomolecular NMR, 8 (1996) 23–35
ESCOM

J-Bio NMR 370

Solution structure of a Lewis^X analogue by off-resonance ^{1}H NMR spectroscopy without use of an internal distance reference

P. Berthault[a,*], N. Birlirakis[a], G. Rubinstenn[b], P. Sinaÿ[b] and H. Desvaux[a,c]

[a]*Molecular Chemistry Service, NMR Laboratory, C.E.A/C.E. Saclay, F-91191 Gif-sur-Yvette Cédex, France*
[b]*Department of Chemistry, Ecole Normale Supérieure, 24 Rue Lhomond, F-75231 Paris Cédex 05, France*
[c]*Center for Interdisciplinary Magnetic Resonance, National High Magnetic Field Laboratory, 1800 East Paul Dirac Drive, Tallahassee, FL 32310, U.S.A.*

Received 28 February 1996
Accepted 3 June 1996

Keywords: Lewis^X; Off-resonance ROESY; Solution structure; Internal dynamics

Summary

A recent ^{1}H NMR method has been applied to the determination of the solution structure and internal dynamics of a synthetic mixed C/O trisaccharide related to sialyl Lewis^X. Varying the rf field offset in ROESY-type experiments enabled the measurement of longitudinal and transverse dipolar cross-relaxation rates with high accuracy. Assuming that for each proton pair the motion could be represented by a single exponential autocorrelation function, it was possible to derive geometrical parameters (r) and dynamic parameters (τ_{cp}). With this assumption, 224 cross-relaxation rates have been transformed into 30 interproton distance constraints and 30 dipolar correlation times. The distance constraints have been used in a simulated-annealing procedure. This trisaccharide exhibits a structure close to the O-glycosidic analogue, but its flexibility seems highly reduced. On the basis of the determined structure and dynamics, it is shown that no conformational exchange occurs, the molecule existing in the form of a unique family in aqueous solution. In order to assess the quality of the resulting structures and to validate this new experimental procedure of distance extraction, we finally compare these solution structures to the ones obtained using three different sets of distances deduced from three choices of internal reference. It appears that this procedure allows the determination of the most precise and accurate solution.

Introduction

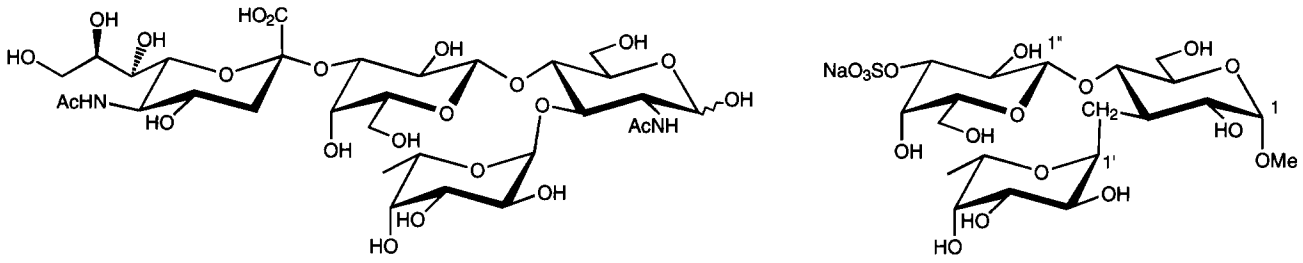
Sialyl Lewis^X (SLe^X, compound **1** in Fig. 1), a terminal tetrasaccharide of cell-surface glycoconjugates of granulocytes is a ligand for E-selectin. This endothelial glycoprotein is involved in the mechanism of recruitment of neutrophils on activated endothelia cells, followed by subsequent invasion into sites of infection or injury (Lasky, 1992; Sharon and Lis, 1993; Parekh and Edge, 1994; Varki, 1994). As a result, extensive studies are currently devoted to the possible application of this finding to the development of anti-inflammatory drugs. In particular, extensive studies on the structure–activity relationship of SLe^X derivatives have recently been achieved.

It has been shown that the glucosamine moiety is rather a 'distributor' (DeFrees et al., 1993; Giannis, 1994), presenting in space fucose on one side, and the carboxylic function of sialic acid on the other. The galactose appears more as an appropriate rigid spacer for this optimal presentation of the charged group (Parekh and Edge, 1994). We have thus recently synthesized a glycomimetic (compound **2** in Fig. 1) on the following basis:

(i) the integral structure of L-fucose is critical for the recognition process (Hasegawa et al., 1994). The usual O-glycosidic bond joining fucose to glucose has been replaced by a C-glycosidic bond (Rouzaud and Sinaÿ, 1983; Mallet et al., 1994) to increase the chemical and biochemical stability of the system;

*To whom correspondence should be addressed.

Abbreviations: COSY, correlation spectroscopy; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; rmsd, root-mean-square deviation; ROESY, rotating frame Overhauser enhancement spectroscopy; SLe^X, sialyl Lewis^X; TOCSY, total correlation spectroscopy.



Compound **1** Compound **2**

Fig. 1. Chemical structures of sialyl Lewis^X (compound **1**) and of the synthetic analogue studied in the text (compound **2**).

(ii) the D-glucosamine unit has been replaced by D-glucose (Singh et al., 1994);

(iii) it has recently been reported (Yuen et al., 1992; Feizi, 1993) that the replacement of the sialic acid residue by a sulphate group retains the affinity for selectins. For the sake of simplicity, this modification has been achieved.

Knowledge of the geometrical and motional properties of this class of compounds in aqueous solution would be beneficial for assessing the conservation of the geometrical motif and for better understanding the structure–activity relationship. However, determination of the structure of such medium-sized compounds by molecular modelling under NMR constraints commonly suffers two types of problems which are often linked: inaccuracy and lack of experimental data.

This is not so crucial for larger biomolecules, as firstly a high level of accuracy is not always required and secondly the number of constraints compensates their imprecision. By contrast, for molecules of about 0.5–2 kDa, the problem becomes important, and the inaccuracy of the NMR constraints has many causes. Among them, conformational flexibility impedes the conventional use of the nuclear Overhauser effects as distance constraints, since both dynamic and structural pieces of information are contained in the same measure and are difficult to separate. The flexibility of small peptides has been explored in various molecular modelling approaches, either to take into account the presence of different structural families (Brüschweiler et al., 1991; Mierke et al., 1994; Cicero et al., 1995; Wang et al., 1995), or to evaluate the influence of internal motions on the distance constraints obtained through NMR measurements (Brüschweiler et al., 1992; Philippopoulos and Carmay, 1994). Recently, time-averaged distance constraints, which are required to be satisfied over the course of a simulation trajectory, have been shown to give better agreement than static distances constraints. This time-averaging of the constraints provides a good representation of the conformational space covered by the molecule when the internal motions occur in the picosecond time scale (Torda et al., 1990).

The second cause of inaccuracy, lack of NMR constraints for medium-sized molecules at ambient temperature and common solvents, is essentially due to the vanishing of longitudinal dipolar cross-relaxation rates, which precludes the use of the NOESY technique (Jeener et al., 1979). An alternative consists of using ROESY (Bothner-By et al., 1984; Bax and Davis, 1985a) to extract proton–proton distances. Angular dispersion and Hartmann–Hahn coherence transfers are however present (Schleucher et al., 1995) and complicate the quantization procedure. The risk for appearance of the latter effects is frequent for oligosaccharides (Van Halbeeck, 1994) due to their narrow spectral width and inherent strong coupling which may favour Hartmann–Hahn matching when a strong rf irradiation is applied on-resonance.

In the present article, the solution structure of compound **2** is determined using a recent NMR method based on off-resonance ROESY. The procedure is based on the attempt to take into account the effects of internal dynamics before resolving the structure. As it is new, a complete validation of this procedure and a comparison of the quality of the obtained structures as a function of the experimental methods (with or without internal distance reference) are described.

Materials and Methods

Sample preparation

Compound **2** has been synthesized as described by Rubinstenn et al. (manuscript in preparation). It has been dissolved in D_2O, lyophilized and redissolved in this solvent. The final concentration was about 8 mM. The atoms are labelled as follows: classical numbering of the pyranosid units, use of the symbol prime for the fucosyl unit, and use of a double prime for the galactosyl unit; the atoms of the methylene linker are arbitrarily noted 'link' and 'link'' according to their respective chemical shifts.

NMR data acquisition

The NMR study has been performed at 15 °C using a field of 11.7 Tesla of an AMX500 Bruker spectrometer. The ^{1}H NMR spectrum has been assigned through classical 2D methods: double-quantum filtered COSY (Shaka and Freeman, 1983), relayed coherence transfers (Wagner,

1983) and TOCSY with MLEV17 composite pulse sequence (Bax and Davis, 1985b). Cross-relaxation experiments have also been used to confirm the assignment, in particular the H_1''–H_4 proximity associated to the Gal → Glc bond.

Distances have been extracted from 1D soft off-resonance ROESY (Desvaux et al., 1994). The selective excitation was achieved by a 270° Gaussian soft pulse (Emsley and Bodenhausen, 1989). The pulse sequence previously published using adiabatic rotations (Desvaux et al., 1995a) and irradiation at two opposite offsets (Desvaux and Goldman, 1996) – to keep a maximum of sensitivity while reducing dispersion of the effective rf field orientation over the spectral width – has been used to measure cross-relaxation rates at various angles θ. Four angles θ (10°, 25°, 40° and 54.7°) were chosen for proton H_1; ten angles (0° (NOESY), 10°, 15°, 20°, 25°, 30°, 35°, 40°, 45° and 54.7°) for proton H_{link}; and six angles (10°, 20°, 30°, 40°, 47° and 54.7°) for the other protons (H_3, H_6', H_1' and $H_{link'}$). The build-up curves were defined by five points corresponding to mixing times between 30 and 150 ms, except for proton H_{link}, for which a supplementary point at 200 ms has been acquired but not used in the fitting procedure. All the spectra have been acquired under similar conditions (rf field strength ~9.7 kHz, 160 scans, cycling delay of 6.6 s, corresponding to ~18 min per spectrum).

NMR data processing

The spectra have been processed using UXNMR software. Processing was identical for all 1D spectra, with zero-filling by a factor 2, exponential filtering (line-broadening of 1 Hz), Fourier transformation, and phase and baseline correction before integration.

The dipolar cross-relaxation rates σ'_θ have been measured for each value of θ using the initial slope and two-spin approximations. The error on this slope ($\Delta\sigma'_\theta$) has been taken into account during the linear fitting procedure (Press et al., 1988) from estimated errors in the integral (percentage of error in the intensity and an identical bias representing possible errors on the baseline correction and noise effects). These values (σ'_θ and $\Delta\sigma'_\theta$) have been used for the determination of both the longitudinal (σ) and transverse (μ) relaxation rates (Desvaux et al., 1994,1995b) using a Marquard algorithm (Press et al., 1988). Uncertainty values for the longitudinal and transverse relaxation rates have been obtained using a Monte Carlo procedure which simulates new data sets from the experimental values (σ'_θ and $\Delta\sigma'_\theta$) and determines through least-square fitting the new values of σ and μ.

Extraction of ^1H-^1H distances

The cross-relaxation rates between proton i and proton j depend on the dipolar spectral densities at particular frequencies, whatever the model of motion is (Solomon, 1955; Abragam, 1961; Ernst et al., 1987; Goldman, 1988; Desvaux et al., 1994):

$$\sigma_{ij} = -J_{ij}(0) + 6J_{ij}(2\omega) \quad (1)$$

$$\mu_{ij} = 2J_{ij}(0) + 3J_{ij}(\omega) \quad (2)$$

Transforming the cross-relaxation rates in terms of distances and correlation times requires an assumption on the behaviour of the spectral density function $J_{ij}(\omega)$. As a model, we assume:

$$J_{ij}(\omega) = \frac{1}{10}\frac{\gamma^4\hbar^2}{r_{ij}^6}\frac{\tau_{cp_{ij}}}{1+\omega^2\tau_{cp_{ij}}^2} \quad (3)$$

where $\tau_{cp_{ij}}$ represents the correlation time of the proton pairs ij and γ is the proton magnetogyric ratio. The geometrical parameters (r_{ij}) and the correlation times per pair of protons (τ_{cp}) as well as their associated errors, can then be computed using the following equation (Desvaux et al., 1995b), omitting the ij indexes for clarity:

$$\tau_{cp} = \frac{1}{\omega}\sqrt{\frac{1-22\frac{\sigma}{\mu}+\sqrt{81+36\frac{\sigma}{\mu}+324\frac{\sigma^2}{\mu^2}}}{8+16\frac{\sigma}{\mu}}} \quad (4a)$$

$$r = \left(\frac{\gamma^4\hbar^2\tau_{cp}}{10\mu}\left(2+\frac{3}{1+\omega^2\tau_{cp}^2}\right)\right)^{1/6} \quad (4b)$$

We define the geometrical parameter of Eq. 4b as an NMR interproton distance, and refer to this procedure of distance extraction as 'Method A'. The second definition of NMR interproton distances (called 'Method B' in the following) is the classical procedure of distance extraction, based on the following equation:

$$r = \left(\frac{\sigma'_{\theta_{ref}}}{\sigma'_\theta}\right)^{1/6} r_{ref} \quad (5)$$

where the 'ref' subscript refers to a reference pair of protons, the internal reference. This model amounts to consider the overall isotropic Brownian motion of a rigid molecule for which the dipolar spectral density is:

$$J_{ij}(\omega) = \frac{1}{10}\frac{\gamma^4\hbar^2}{r_{ij}^6}\frac{\tau_c}{1+\omega^2\tau_c^2} \quad (6)$$

The correlation time τ_c is then the same for all proton pairs.

Molecular modelling

The distance data have been introduced as constraints in a simulated-annealing procedure (Nilges et al., 1988) using X-PLOR v. 3.1 software (Brünger, 1992). Starting from a template structure, high-temperature (1000 K)

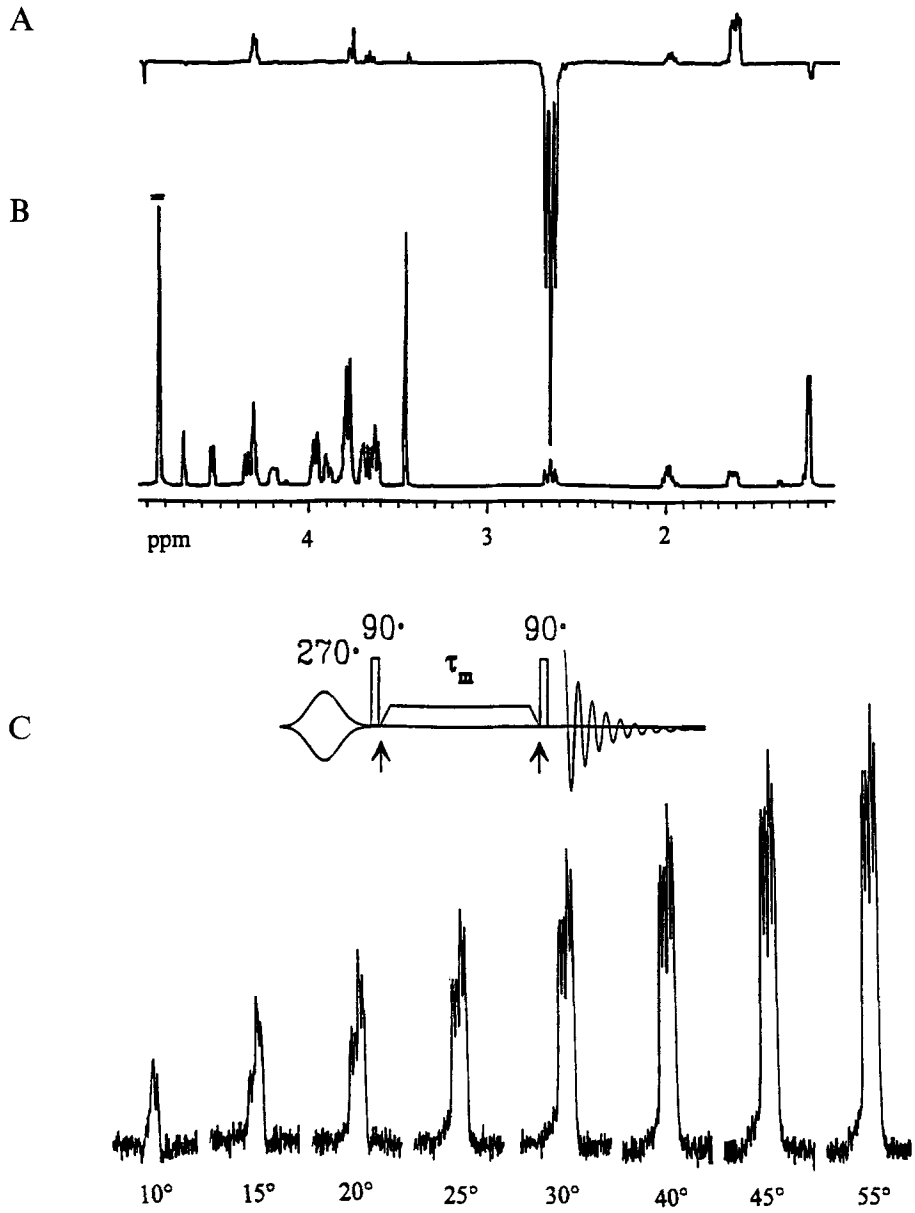


Fig. 2. (A) Subspectrum obtained by 1D off-resonance ROESY with selective excitation of H_{link} ($\tau_m = 200$ ms, $\theta = 54.7°$); (B) 1D spectrum of compound **2**; (C) variation of the H'_{link} intensity due to cross-relaxation from H_{link} at a constant mixing time $\tau_m = 200$ ms, but for various angles θ. In the insert between (B) and (C), the NMR pulse sequence used is shown (Desvaux et al., 1995a).

dynamics of 10 ps have been performed with low weight on the van der Waals terms and progressive introduction of the distance restraints. Slow cooling during 8 ps followed to restore the usual values of the covalent energy terms. The resulting structures were sorted according to the following acceptance criteria: deviation from the ideal values inferior to 0.01 Å for the bond lengths, to 5° for the angles and impropers, and no NOE violation larger than 0.2 Å. Improper angles were neither introduced for H_{link} and H'_{link}, nor for H6 of the different units. The nonequivalence and the high number of constraints for protons H_{link} and H'_{link} were sufficient to use nonequivalent protons in the simulated-annealing procedure. For protons H6, this was not the case and we have used the pseudo-atom facility of X-PLOR (with distance averaging in r^{-6}). This reduced the number of distance restraints to 30.

Solution structure of the trisaccharide

Determination of interproton distances and correlation times

Recently, we have proposed an approach for studying dipolar relaxation along an effective field generated through off-resonance irradiation and having an angle θ with the static magnetic field axis (Desvaux et al., 1994). By this method, the measurement of dipolar cross-relaxation rates σ'_θ for $0° \leq \theta \leq 60°$ enables the simultaneous

determination of the dipolar longitudinal (σ, usually measured through NOESY) and transverse (μ, theoretically determined through ROESY) cross-relaxation rates. The rates are obtained with a very high accuracy as they result from a least-square fitting procedure of an exact equation, based on several θ values:

$$\sigma'_\theta = \cos^2 \theta \, \sigma + \sin^2 \theta \, \mu \qquad (7)$$

Instead of using only one of these values (σ or μ) for the molecular structural determination, we have proposed to exploit these two values *simultaneously* in terms of a geometrical parameter (an internuclear distance) and a

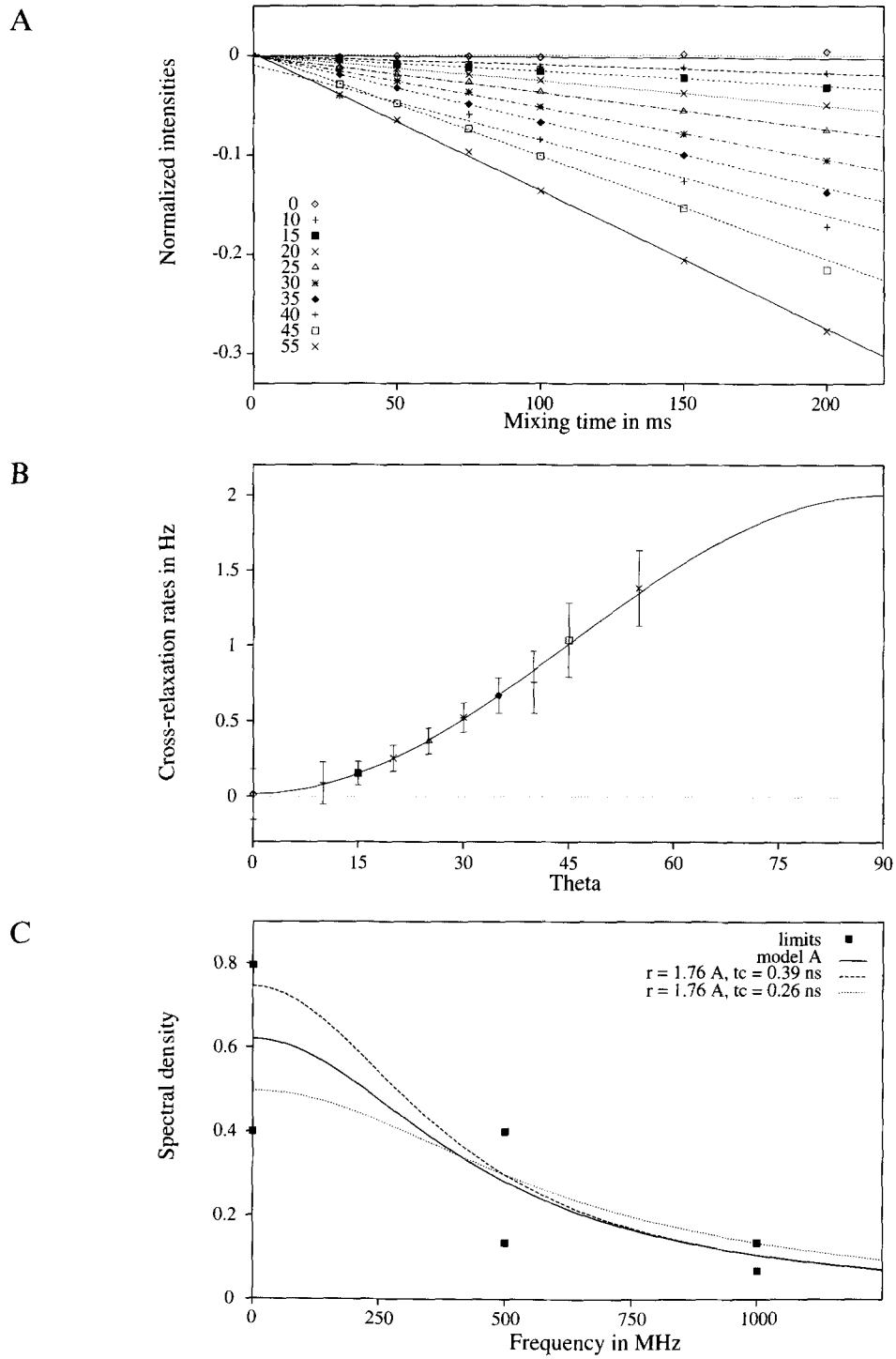


Fig. 3. (A) Build-up curves representing the magnetization transferred from H_{link} to H'_{link}. The best-fit slopes give σ'_θ; (B) cross-relaxation rates as a function of θ. The best-fit theoretical curve according to Eq. 7 is superimposed; (C) data points (■): limit values of the dipolar spectral density of the same proton pair; continuous curve (solid line): corresponding Lorentzian spectral density function (Method A). The two other curves (broken lines) correspond to Lorentzian spectral density functions for extreme values of the correlation times.

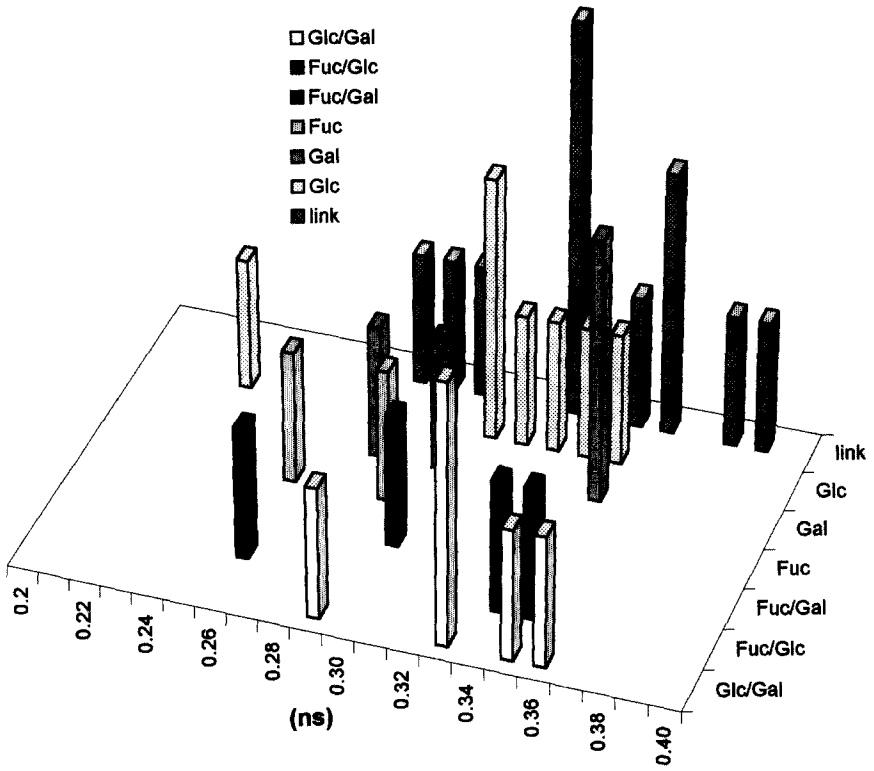


Fig. 4. 3D histogram representing the correlation times of the proton pairs, τ_{cp}, as a function of their occurrences (height) and of the source of the two protons: Glc = glucose, Gal = galactose, Fuc = fucose, link = methylene linker.

dynamic parameter (a dipolar correlation time per pair of protons τ_{cp}) (Desvaux et al., 1995b). This is realized assuming that the spectral density function of the considered proton pair is Lorentzian, which does not prevent the correlation time τ_{cp} to be different for each pair of protons and differs strongly from the classical procedure (Method B), in which an internal distance reference r_{ref} is needed. In Method B, the interproton distances r are given by Eq. 5, assuming that all proton pairs experience the same dynamics and considering that the molecule is *rigid* and that its motion is *isotropically* Brownian. These two assumptions have many times been invalidated by experiments (Krishnan et al., 1991; Peng and Wagner, 1992; Li et al., 1995) or numerical simulations (Brüschweiler et al., 1992; Philippopoulos and Carmay, 1994) and it has been shown that the solution structures depend strongly on the choice of internal reference (Brüschweiler et al., 1992; Philippopoulos and Carmay, 1994).

To measure a maximum number of constraints with very high accuracy, 1D soft off-resonance ROESY experiments (Desvaux et al., 1994) have been performed for each sufficiently isolated resonance, namely H_1, H_3, H_{link}, H'_{link}, H'_6 and H''_1, at various angles θ and for various mixing times τ_m. From these sets of experiments, 224 cross-relaxation rates σ'_θ have been measured. As an illustrative example, the selective excitation of proton H_{link} is presented in Figs. 2 and 3. In total, 37 cross-relaxation peaks (30 independent) have been recorded and transformed into proton–proton distance constraints; 10 of them concern interactions between pyranosid units, while

TABLE 1
ϕ AND ψ AVERAGE ANGLES OF THE GLYCOMIMETIC STRUCTURES FOR DIFFERENT SETS OF DISTANCE CONSTRAINTS

Set of distance constraints	Galactose-glucose		Fucose-galactose	
	ϕ_1 (°)	ψ_1 (°)	ϕ_2 (°)	ψ_2 (°)
A	−91.4 ± 6.3	134.1 ± 5.5	−59.9 ± 1.1	−100.4 ± 2.4
B1	–	–	–	–
B2	−91.8 ± 6.6	143.0 ± 5.0	−61.8 ± 0.7	−98.2 ± 2.7
B3	−91.7 ± 6.2	140.0 ± 3.4	−61.3 ± 1.0	−99.0 ± 1.7

Only the structures accepted following the criteria given in the paragraph Materials and Methods are considered. Set A denotes the constraints obtained with the method described in this paper; sets B are obtained with different internal distance references; B1: pair H_{link}-H'_{link}, B2: pair H1"-H2", B3: H1-H2. The interglycodisic angles are defined in accordance with IUPAC recommendations: ϕ_1: (O5"-C1"-O4-C4); ψ_1: (C1"-O4-C4-C3); ϕ_2: (O5'-C1'-C_{link}-C3); ψ_2: (C1'-C_{link}-C3-C2).

10 others involve the CH_2 linker. One supplementary (intra-unit) interaction has been observed by a 2D off-resonance ROESY experiment. However, it has not been used as a distance constraint, since no build-up procedure at different angles θ was performed.

The number of distance constraints obtained by off-resonance ROESY is much larger than what is usually derived by classical procedures for such molecules (Ball et al., 1992; Lin et al., 1992; Miller et al., 1992; Mukhopadhyay et al., 1994). As stated above and as can be seen in Figs. 2C, 3A and 3B, this arises from the quasi-vanishing longitudinal cross-relaxation rates for the relevant motions of this molecule. Indeed, the average value of $|\sigma/\mu|$ is 7.5×10^{-2}. Now, assuming a Lorentzian dipolar spectral density function (Eq. 3), the average value found for the dipolar correlation times per proton pair τ_{cp} is 0.32 ± 0.03 ns. The dispersion of dipolar correlation times over all proton pairs is relatively small (see Fig. 4), but is not negligible, since using Method B it can induce an error in the distances of 4% with μ and more than 50% with σ. Note, however, that no difference in correlation times involving protons of the same unit and inter-unit pairs can be found, which indicates that the trisaccharide is relatively rigid.

Error estimation

Intrinsic uncertainties in the geometrical (r) and dynamic (τ_{cp}) parameters have been determined through Monte Carlo simulations (Press et al., 1988). Further estimation of the experimental reproducibility is given by results obtained when the two reciprocal cross-relaxation rates (from proton i to proton j and reciprocally) can be derived. The resulting errors are less than 0.07 Å for distances smaller than 3.2 Å, and about 0.13 Å for the larger distances. Dispersion of these values depends on the number of angles θ acquired, the quality of the fit and the estimation of the integral errors (see Materials and Methods). Several biases (two-spin and initial rate approximations) are not taken into account in these estimations. However, their influence for short distances seems to be relatively limited, since the results come from a fitting procedure using nearly vanishing cross-relaxation rates σ'_θ with nonvanishing direct relaxation rates ρ'_θ. Spin-diffusion processes are not efficient in this case and the two-spin approximation can be safely considered to be valid.

The last source of error is the assumption that the geometrical parameter (r obtained from Eq. 4b) can be considered as a good representation of the internuclear distance for the modelling procedure, i.e. that the spectral density function can be adequately represented by a Lorentzian function where radial and angular correlation functions are separated. Several arguments obtained through numerical simulations (Brüschweiler et al., 1992; Philippopoulos and Carmay, 1994; Abseher et al., 1995) tend to prove that this separation is often justified. Whatever the definite reply to these questions will be, since the described method takes at least partially into account the internal motions through simultaneous exploitation of σ and μ, the bias is smaller than the one associated with the internal distance reference method. Considering this, an experimental uncertainty of 0.2 Å has been assumed for all the NMR distances used in the simulation, a value much smaller than what is usually chosen with the second type of method.

Molecular modelling

The interproton distance data have been introduced in an 'ab initio' simulated-annealing procedure repeated 50 times with random initial velocities in the dynamics. Among the resulting structures, 29 have been accepted according to criteria on the constraints and covalent geometry terms (no NOE violation larger than 0.2 Å). Recording all the interproton distances smaller than 4 Å allowed us to check that no nuclear Overhauser effects were missing in the off-resonance ROESY 2D spectra (some cannot be observed due to overlap of the involved signals).

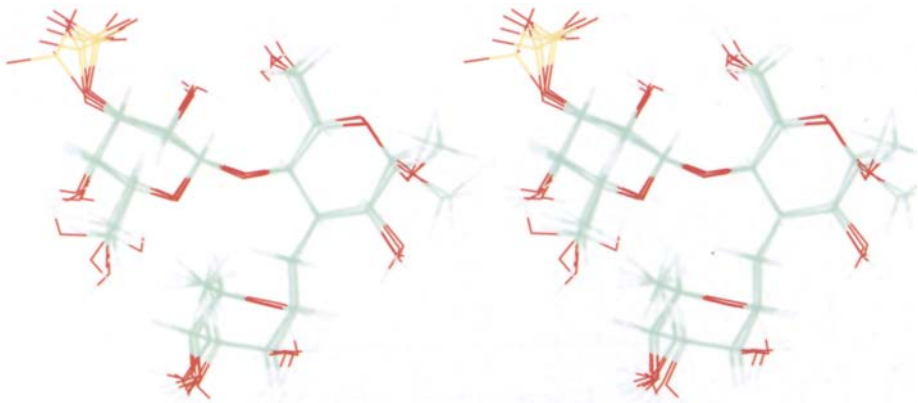


Fig. 5. Stereoview of the superposition of 10 out of 29 accepted structures obtained with Method A. The rmsd value between the coordinates of the heavy atoms is 0.58 Å and falls to 0.21 Å when only the backbone is taken into account. Red = oxygen; yellow = sulfur; and white = hydrogen.

A pictorial superposition of 10 of these structures is given in Fig. 5. The rmsd on the coordinates is 0.58 Å considering the heavy atoms and falls to 0.21 Å when the sulphate and hydroxyl groups are removed from the calculation subset (we will refer to this second subset of heavy atoms as the 'backbone' in the rest of the text). These structures have been further minimized in CHARMm22 (Brooks et al., 1993) by 500 steps of Adopted Basis Newton–Raphson with parameters derived from the work of Brady (Ha et al., 1988), in order to check their consistence with a more standard force field. The pairwise rmsd between the X-PLOR and CHARMm structures is less than 0.5 Å considering the heavy atoms, and is principally due to the treatment of the sulphate group in vacuum.

To enable a simple description of the solution structures and a comparison with molecular modelling results on Lewis^X-type molecules, the ϕ and ψ dihedral angles concerning the geometry around the interglycosidic bonds have been measured on each of the accepted structures. Their average values are given in Table 1. Due to the high number of NMR constraints available around the Fuc–Glc link, the associated angle values are precisely defined. For both glycosidic bonds, all the angles are fully consistent with the calculated maps of simulated NOEs versus ϕ–ψ published by Bush et al. (Miller et al., 1992; Mukhopadhyay et al., 1994), falling in the most probable region for total agreement between simulation and experiment.

Number of structural families

The question of the number of structural families for Lewis^X analogues has often be discussed and the answer is not well established. Although most of the studies in the last years tend to conclude that the trisaccharide fragment is rigid, in agreement with the pioneering work of Lemieux's group (Lemieux et al., 1980; Thorgersen et al., 1982), other authors have proposed conformational equilibrium (Ejchart et al., 1992; Ejchart and Dabrowski, 1992). The principal reason for this controversy is the lack of data allowing definition of the conformation of carbohydrates in water solution (French and Brady, 1990). Nevertheless, answering this question is crucial for understanding the biological activity of Lewis^X-type compounds, which is supposed to involve bimolecular binding phenomena. Investigation of this question in the light of the present NMR results has been performed.

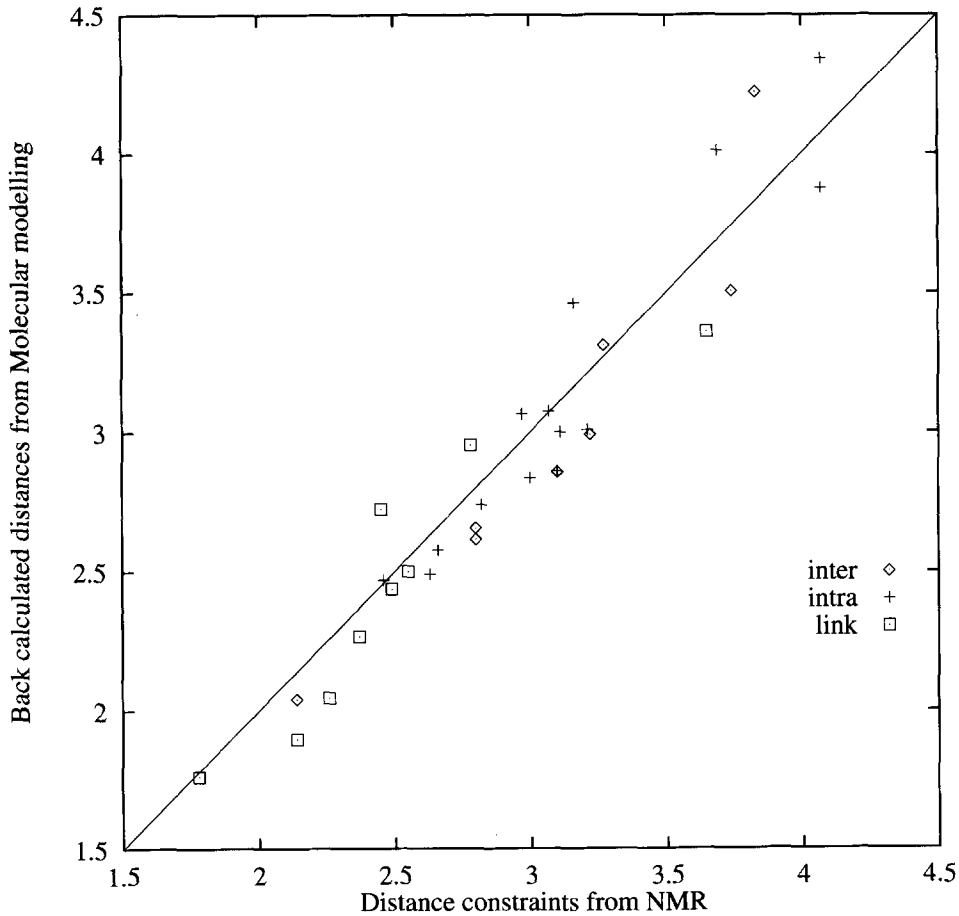


Fig. 6. Experimental interproton distances determined by method A versus back-calculated distances after averaging in r^{-6} over all the structures accepted after simulated annealing.

TABLE 2
PARAMETERS OF THE GLYCOMIMETIC STRUCTURES FOR DIFFERENT SETS OF DISTANCES

Set of distance constraints	Number of accepted structures[a]	Total number of violations	Rmsd between accepted structures[b] (Å)
A	29	58	0.21
B1	0	80	–
B2	16	72	0.19
B3	16	75	0.21

[a] Following the acceptance criteria given in Materials and Methods.
[b] Taking into account the backbone atoms.

Measurement of the ^3J couplings around the C-linkage in the proton spectrum gives the values: $J_{H'_1,H_{link}} = 12.6$ Hz; $J_{H'_1,H_{link}} = 2.8$ Hz; $J_{H_3,H_{link}} < 1$ Hz; and $J_{H_3,H'_{link}} = 7.5$ Hz. Each pair corresponds to a maximum and a minimum in the Karplus law, parametrized by Haasnoot et al. (1980), which prevents the trisaccharide to exhibit fast exchange (< 1 μs) between conformations with different couplings. In this case, the measured couplings are averaged and cannot correspond to local extrema of the trigonometric function. Back-calculation of these four dihedral angles over all the 29 accepted simulated structures is in full agreement with the experimental values. Through these results we were able to identify H_{link} as the pro-S proton.

For all protons except H'_6, the cross-over direct relaxation-rate ratios measured by off-resonance ROESY at $\theta = 54.7°$ are consistent with the theoretical value of 1/2, characteristic of purely dipolar interactions (Desvaux et al., 1994). According to the usually accepted relaxation mechanism for proton relaxation, this indicates that no fast chemical exchange processes between 1 μs and 10 ms are detected (Desvaux et al., 1995c). Moreover, in contrast to other studies (Desvaux, unpublished results), such a result shows that the cross-relaxation rates are correctly measured, i.e., if a large conformational sampling is present for compound **2**, the sum of all not effectively detected cross-relaxation rate contributions must be very small. This point is confirmed by the rather small and homogeneous distribution of the τ_{cp} values for the various sets of pairs (intraresidues or interresidues), as seen in Fig. 4. This indicates that there are no parts of the molecule which exhibit large internal motions in the nanosecond range different from other parts.

Finally, the possibility to resolve the structure without any violation or missing proximity and with strong distance constraints (as the distance bounds used are rather narrow), proves that there is essentially one family of structures for compound **2** in solution and justifies the a posteriori choice of the modelling procedure without any time averaging (Torda et al., 1990) or ensemble averaging (Brüschweiler et al., 1991; Mierke et al., 1994; Cicero et al., 1995; Wang et al., 1995).

Validation of the experimental procedure

Since this work represents the first attempt to resolve a structure using the new definition of NMR distances (Eq. 4b), it seems relevant to validate this definition and to compare the effects of choice of NMR distances (Eqs. 4b and 5) on the solution structures.

Assessment of the quality of the structures

Two tests have been performed to check the validity of the assumption that the geometrical parameter (r) of Method A is a realistic representation of internuclear distances:

(i) the measured distance between the two methylene protons of the linker is 1.79 Å, which is very close to the separation of glycine α-protons (1.76 Å) often used as an internal reference in peptides. The same excellent agreement is found for all distances involving proton H_1 with the protons of the glucose unit (H_2 to H_6 and methyl). This point is worth noting, as slight variations in the dynamics should be present: the pyranosid ring is generally considered to be rigid (Hajduk et al., 1993) but the time modulation of dipolar interactions between proton H_1 and proton H_6 on one hand, or between H_1 and the methyl on the other hand, should be different. This is confirmed by the study of the correlation time per pair of protons (Fig. 4), which in the last two cases was found to be smaller than the average value observed for the pyranosid ring. A simple interpretation of the same data using a distance reference would not be so accurate;

(ii) Figure 6 reveals an excellent agreement between the interproton distances back-calculated from the accepted structures and the experimental ones (root-mean-square deviation from the central values of the constraints: 0.19 Å). The best-fit slope is equal to 0.98 ± 0.03, assuming that the straight line ordinate is zero. Without this assumption, the fit is not highly improved and the ordinate constant is zero (within experimental error). Over the whole range of internuclear distances, no large bias appears to exist between experimental distances and simulated ones. Finally, no distortion in the distance determination is found for the case of relaxation rates involving methyl-type groups, although they represent seven out of the 30 restraints. Again, these results differ strongly from what has been obtained in other studies using Method B for which a large discrepancy has been pointed out (Post, 1992). These results prove that our method allows one to take, at least partially, into account the internal dynamics.

Comparison with the internal distance-reference method

In order to evaluate the quality of the solution structures obtained by Method A, a comparison with the classical distance-reference procedure (Method B) has been performed. For Method B, the cross-relaxation rates

σ'_θ were measured at $\theta = 54.7°$, where the dipolar cross-relaxation reaches its largest value.

The same molecular modelling procedure has been applied to three sets of NMR constraints obtained from three internal references, representing three possible choices: $H_{link} \leftrightarrow H'_{link}$ (set B1), H1" → H2" (set B2) and H1 → H2 (set B3). These three pairs exhibit different dynamic behaviour, since their correlation times are respectively larger than, smaller than and nearly identical to the average value. The problems of internal motion and of the choice of the reference pair are illustrated in Fig. 3C. Lorentzian spectral density functions are drawn for the extreme values of the correlation time distribution in comparison to the result obtained by Method A, and limit values of the dipolar spectral density (Desvaux et al., 1995b). It clearly appears that although the correla-

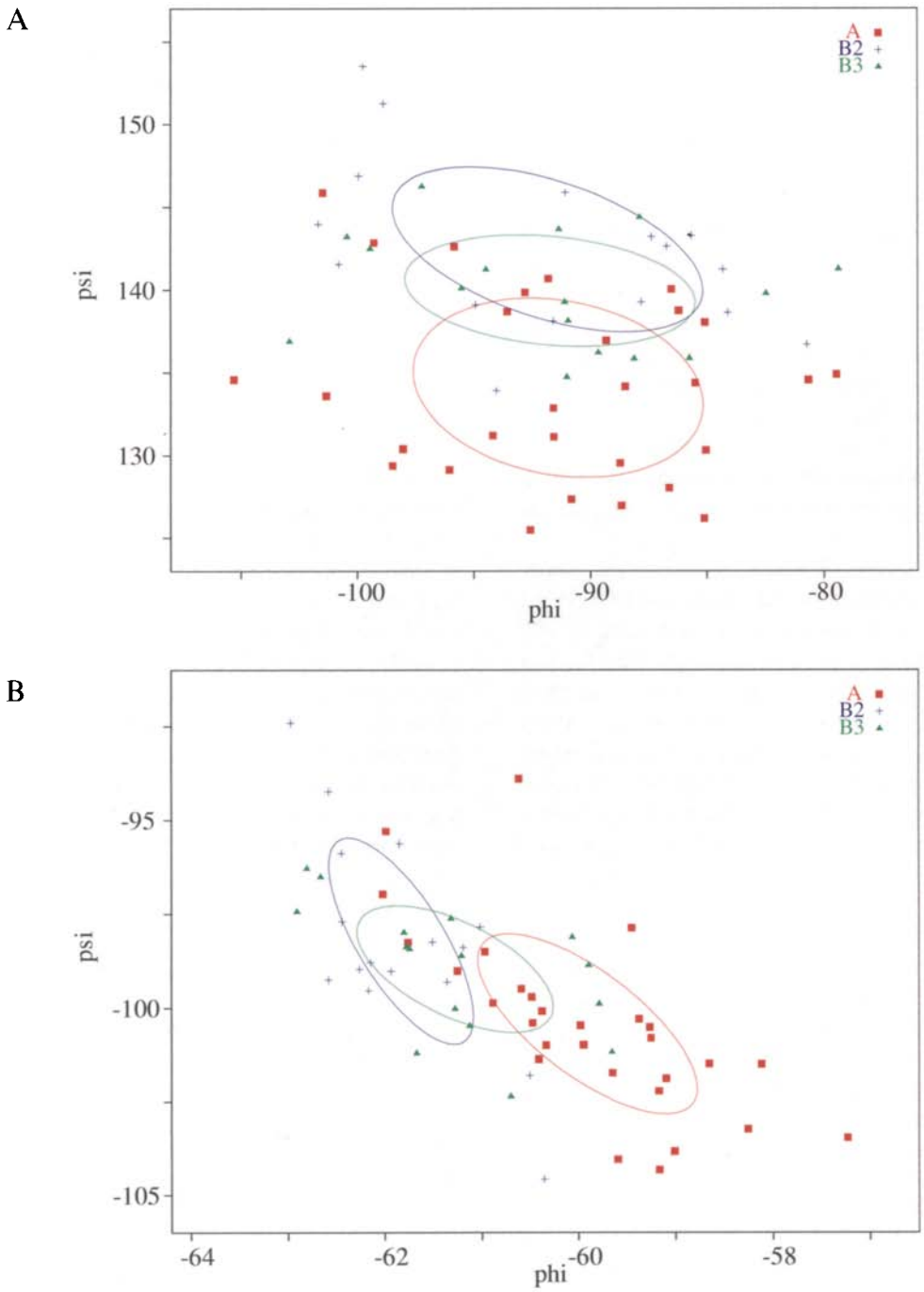


Fig. 7. ϕ versus ψ maps of the galactose-glucose (A) and fucose-glucose (B) glycosidic links. Each point describes one structure resulting from the molecular simulation. Red squares correspond to structures obtained using the A set of distances, blue crosses and green triangles represent structures obtained using $H''_1 \rightarrow H''_2$ (B2 set) of $H_1 \rightarrow H_2$ (B3 set) as internal reference, respectively. The superimposed ellipses show the average distribution of ϕ-ψ values for each set at one standard deviation.

tion time distribution is restricted, the behaviour of the spectral density function varies strongly. The molecular modelling results are summarized in Tables 1 and 2. From Table 2, it is clear that the set of constraints A gives better results than the others, both for the average number of accepted structures and for the total number of NOE violations. The observed variations are much larger than a statistical deviation: there are 1.8 times more accepted structures using the procedure A than when using the sets B2 or B3. The discrepancy between B1 and A is even more striking, as no structure has been accepted in the former case. This does not prove that this set is totally inconsistent, (in fact 44 structures have been rejected due to 1 violation), but it demonstrates that it is less consistent than sets A, B2 and B3. These observations (first columns in Table 2) are however moderated by the last column: the rmsd between the accepted structures display no significant differences and are very low. This shows that the number and the repartition of NMR restraints are good enough to allow a precise definition of the structure in all three cases. Comparison of the rmsd values of experimental and back-calculated H–H distances (averaged in r^{-6} over all the accepted structures) as well as energies, do not reveal significant differences either. These results are nevertheless not surprising, as they show that no fundamental errors in the constraints are present, and that precision and accuracy are different.

Whatever the sets of distances taken into account, the solution structures seem to be relatively similar. This is illustrated by the ϕ-ψ angles of the two glycosidic bonds (Table 1). The points displayed in these ϕ versus ψ maps characterize the relative position of the three osidic units. A detailed analysis of the maps reveals however that the geometrical areas of the structures arising from the A and B sets weakly overlap. This is clearly apparent in Fig. 7, which shows for each set the domains where the probability of finding an accepted structure is maximal: the distance from the centre of the A distribution to B2 (or B3) is larger than one standard deviation. Between A and B2 the two distributions are almost fully separated. This shows that the three average structure sets (A, B2, B3) can be said to be equivalent but not identical. This experimental evidence is not so surprising, since many numerical simulations have proved that the solution structures depend on the choice of the internal distance reference (Brüschweiler et al., 1992; Philippopoulos et al., 1994).

Cross-validation

In regard to the additional experimental time required for our approach, when compared to the internal distance-reference method, and considering the small difference between the resulting structures, one can wonder what the relevance of our approach is. This is related to two other important questions, namely whether the accuracy of the cross-relaxation rate values is the only cause of the higher precision, and if this gain is related to accuracy or precision. To start answering these questions without entering further in the debate of accuracy versus precision (Clore et al., 1993; Zhao and Jardetzky, 1994), a recently proposed statistical method, complete cross-validation (Brünger et al., 1993), has been applied for both procedures.

As the number of NMR restraints is too small to separate them in different subgroups (intra-unit, inter-unit) and still yield reliable structures (Brünger et al., 1993), we have adapted this method. 'Working sets' are generated by random removal of a fixed number of constraints – the 'rest set' – from the complete data set. Each working set is then submitted twenty times to the simulated-annealing procedure, giving therefore 20×20 structures. Keeping only the accepted structures according to the criteria described in Material and Methods, average structures were calculated. On the one hand, the rmsd between the average structures is an indication of the completeness of the experimental constraints – the level of compatibility between them – independently of the force field. On the other hand, the evolution of this rmsd with the size of the rest set gives an idea of the richness of the original data set. Cross-validation has been performed comparatively for the A set and for the B3 set, the latter using as reference the proton pair whose correlation time τ_{cp} is the closest to the average value. A similar cross-validation procedure between B1, B2 and B3 sets (results not presented) showed that the richness and the completeness of the B3 set are the best. To enable a meaningful comparison between the A and B3 sets, whatever the statistical fluctuations, for each 'working set' of A, a corresponding 'working set' of B3 was created, containing the same remaining NMR constraints. The results presented in Table 3 show that, whatever the number of NMR restraints kept, the results obtained with the A set are always better than those with the B3 set. This is the case both for the rmsd between the average structures and for the average number of accepted structures, and the results prove that the A set is more complete than the B3 set. The behaviour

TABLE 3
RESULTS OF THE CROSS-VALIDATION PROCEDURE

Number of kept constraints	Average number of accepted structures[a]		Rmsd between average structures[b] (Å)	
	A set	B3 set	A set	B3 set
27	18.6	16.1	0.10	0.11
25	18.9	15.9	0.18	0.32
22	15.9	15.7	0.32	0.35
20	15.8	15.2	0.36	0.39
15	15.8	14.9	0.69	0.72
10	17.2	16.8	0.78	0.80

[a] Following the acceptance criteria given in Materials and Methods.
[b] Taking into account the backbone atoms.

of the rmsd for the A set is a slowly decreasing function of the number of restraints kept with no sudden variation. On the other hand, for the B3 set nearly identical rmsd values are observed for 20, 22 and 25 restraints, which are much larger than the rmsd value for 27 restraints. The most interesting result is the relative behaviour of the A and B3 rmsd values when less than 10 restraints are suppressed. For 25 restraints, the rmsd of the A set is much better than the one obtained for the B3 set (0.18 Å versus 0.32 Å). This last value is equal to the one obtained with set A, but for 22 restraints. Consequently, for the domain in which the behaviour of the A and B3 sets shows the largest differences (between 20 and 25 restraints kept), it is necessary to have three more restraints in the B3 'working set' than with A to get the same rmsd values between the average structures. This proves clearly the higher richness of the A set. From these comments, it results that the A set is the most consistent one. The solution structure of molecule **2** obtained with this set of distances is then at the same time more accurate and more precise.

Conclusions

We have firstly shown that off-resonance ROESY at a sole spin-lock rf offset (one angle θ) is sufficient to enable the measurement of effective cross-relaxation rates. Thus, reliable interproton distances can be obtained by using the internal distance-reference method, even when the longitudinal cross-relaxation rates are nearly vanishing (results of the B1, B2 and B3 sets). This is of high interest for small oligosaccharides, for which this drawback is further complicated by the problem of the narrow spectral bandwidth favoring unwanted Hartmann–Hahn coherence transfer in on-resonance ROESY. However, this procedure does not solve problems inherent to internal dynamics, whose consequences may mask, for example, the presence of several structural families in solution. In contrast, the accurate measurement of pure longitudinal *and* transverse cross-relaxation rates determined from the effective cross-relaxation rates at various angles θ, enables separation of the data into structural and dynamic parameters. Although there are many ways to define these parameters, we suggest that for each pair of protons (i,j), a simple separation into a geometrical r_{ij} and a dynamic τ_{cpij} parameters should be done by assuming a dipolar Lorentzian spectral density function. This may be seen as a new definition of NMR distances, similarly to the classical definition which uses a pair of protons as internal distance reference. Of course, by assuming other models of motions (other expressions of the spectral density function) or other exploitation procedures, one could define other NMR distances. The large advantage of this point of view is the fact that our inability to measure the dipolar spectral density function completely, i.e., to fully describe the relative motion of two protons, is no longer relevant, since it has been replaced by a definition which does not need a physical justification.

In this paper, we have thus applied this principle to a small biomimetic oligosaccharide. The use of off-resonance ROESY has allowed the extraction of many more NMR restraints than can be obtained from NOESY or classical ROESY for this type of molecules. The longitudinal and transverse cross-relaxation rates have been determined for 30 proton pairs and transformed into geometrical and dynamic parameters under the assumption that the motion of each pair can be described by an independent Lorentzian spectral density function. Validity of the structures resulting from a simulated-annealing procedure using these constraints revealed that this assumption was satisfying. We could thus determine the solution structure of this mixed C/O trisaccharide with an excellent resolution and settle on its conformational rigidity and on the absence of different structural families. The solution structure of this analogue of $Lewis^x$ is similar to the calculated one described for O-glycosidic bond (Lemieux et al., 1980), indicating that, in this case, the C-glycosidic bond does not induce large conformational modifications, in accordance with usual expectations (Haneda et al., 1992; Wei et al., 1995). The application of this NMR procedure to other analogues could allow a detailed comparative study of the structure and dynamics of the C and O trisaccharides, allowing the investigation of conformational modifications or differences in flexibility, which may be related to biological activity. Such a task is currently in progress in our laboratory.

Comparison of the solution structures obtained by this method with those obtained using various internal distance references has shown that the resulting solution structures are similar, but not identical. This illustrates the well-known problem of the choice of the reference pair (Brüschweiler et al., 1992; Philippopoulos et al., 1994) and the difficulty of comparing solution structures, in relation with the precision versus accuracy debate (Clore et al., 1993; Zhao and Jardetzky, 1994). The cross-validation procedure has demonstrated that for this given oligosaccharide, the set of distances obtained by the method exploiting simultaneously σ and μ seems the most consistent one among the four sets tested. It also demonstrates that a part of the difficult task of taking into account the effect of internal motions was achieved. In particular, we have shown that this method allows the cancellation of the usual problem of dynamics of proton pairs involving freely rotating methyl groups. Even if the conclusions have to be tempered since the studied molecule certainly does not exhibit all kinds of internal dynamics, this method represents a powerful tool for determination of structure and dynamics of such medium-sized molecules. Moreover, its capability to deal with local motions can be exploited for cases where no internal distance reference is available, in particular for interaction studies between molecules.

References

Abragam, A. (1961) *Principles of Nuclear Magnetism*, Clarendon Press, Oxford, U.K.

Abseher, R., Lüdemann, S., Schreiber, H. and Steinhauser, O. (1995) *J. Mol. Biol.*, **249**, 604–624.

Ball, G.E., O'Neill, R.A., Schultz, J.E., Lowe, J.B., Weston, B.W., Nagy, J.O., Brown, E.G., Hobbs, C.J. and Bednarski, M.D. (1992) *J. Am. Chem. Soc.*, **114**, 5449–5454.

Bax, A. and Davis, D.G. (1985a) *J. Magn. Reson.*, **63**, 207–213.

Bax, A. and Davis, D.G. (1985b) *J. Magn. Reson.*, **65**, 355–360.

Bothner-By, A.A., Stephens, R.L., Lee, J., Warren, C.D. and Jeanloz, R.W. (1984) *J. Am. Chem. Soc.*, **106**, 811–813.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1993) *J. Comput. Chem.*, **4**, 187–217.

Brünger, A.T. (1992) *X-PLOR v. 3.1, a System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT and London, U.K.

Brünger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R. and Nilges, M. (1993) *Science*, **261**, 328–331.

Brüschweiler, R., Blackledge, M. and Ernst, R.R. (1991) *J. Biomol. NMR*, **1**, 3–12.

Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M. and Ernst, R.R. (1992) *J. Am. Chem. Soc.*, **114**, 2289–2302.

Cicero, D.O., Barbato, G. and Bazzo, R. (1995) *J. Am. Chem. Soc.*, **117**, 1027–1033.

Clore, G.M., Robien, M.A. and Gronenborn, A.M. (1993) *J. Mol. Biol.*, **231**, 82–102.

DeFrees, S.A., Gaeta, F.C.A., Lin, Y.-C., Ichikawa, Y. and Wong, C.-Y. (1993) *J. Am. Chem. Soc.*, **115**, 7549–7550.

Desvaux, H., Berthault, P., Birlirakis, N. and Goldman, M. (1994) *J. Magn. Reson.*, **A108**, 219–229.

Desvaux, H., Berthault, P., Birlirakis, N., Goldman, M. and Piotto, M. (1995a) *J. Magn. Reson.*, **A113**, 47–52.

Desvaux, H., Berthault, P. and Birlirakis, N. (1995b) *Chem. Phys. Lett.*, **233**, 545–549.

Desvaux, H., Birlirakis, N., Wary, C. and Berthault, P. (1995c) *Mol. Phys.*, **86**, 1059–1073.

Desvaux, H. and Goldman, M. (1996) *J. Magn. Reson.*, **B110**, 198–201.

Ejchart, A., Dabrowski, J. and Von der Lieth, C.-W. (1992) *Magn. Reson. Chem.*, **30**, 105–S114.

Ejchart, A. and Dabrowski, J. (1992) *Magn. Reson. Chem.*, **30**, 115–124.

Emsley, L. and Bodenhausen, G. (1989) *J. Magn. Reson.*, **82**, 211–221.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, U.K.

Feizi, T. (1993) *Curr. Opin. Struct. Biol.*, **3**, 701–710.

French, A.D. and Brady, J.W. (1990) *Computer Modeling of Carbohydrate Molecules*, ACS Symposium Series 430, American Chemical Society, Washington, DC.

Giannis, A. (1994) *Angew. Chem. Int. Ed. Engl.*, **33**, 178–180.

Goldman, M. (1988) *Quantum Description of High-Resolution NMR in Liquids*, Clarendon Press, Oxford, U.K.

Ha, S.N., Giammona, A., Field, M. and Brady, J.W. (1988) *Carbohydr. Res.*, **180**, 207–221.

Haasnoot, C.A.G., De Leeuw, F.A.A.M. and Altona, C. (1980) *Tetrahedron*, **36**, 2783–2792.

Hajduk, P.J., Horita, D.A. and Lerner, L.E. (1993) *J. Am. Chem. Soc.*, **115**, 9196–9201.

Haneda, T., Goekjian, P.G., Kim, S.H. and Kishi, Y. (1992) *J. Org. Chem.*, **57**, 490–498.

Hasegawa, A., Ando, T., Kato, M., Ishida, H. and Kiso, M. (1994) *Carbohydr. Res.*, **257**, 67–80.

Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.

Krishnan, V.V., Shekar, S.C. and Kumar, A. (1991) *J. Am. Chem. Soc.*, **113**, 7542–7550.

Lasky, L.A. (1992) *Science*, **258**, 964–969.

Lemieux, R.U., Bock, K., Delbaere, L.T.J., Koto, S. and Rao, V.S. (1980) *Can. J. Chem.*, **58**, 631–653.

Li, S., Swindle, S., Smith, S., Nieman, R., Moore, A., Moore, T. and Devens, G. (1995) *J. Phys. Chem.*, **99**, 3371–3378.

Lin, Y.-C., Hummel, W., Huang, D.-H., Ichikawa, Y., Nicolaou, K.C. and Wong, C.-Y. (1992) *J. Am. Chem. Soc.*, **114**, 5452–5454.

Mallet, A., Mallet, J.-M. and Sinaÿ, P. (1994) *Tetrahedron Asymmetry*, **5**, 2593–2599.

Mierke, D.F., Kurz, M. and Kessler, H. (1994) *J. Am. Chem. Soc.*, **116**, 1042–1049.

Miller, K.E., Mukhopadhyay, C., Cagas, P. and Bush, C.A. (1992) *Biochemistry*, **31**, 6703–6709.

Mukhopadhyay, C., Miller, K.E. and Bush, C.A. (1994) *Biopolymers*, **34**, 21–29.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **239**, 129–136.

Parekh, R.R. and Edge, C.J. (1994) *Trends Biotechnol.*, **12**, 339–345.

Peng, J.W. and Wagner, G. (1992) *J. Magn. Reson.*, **98**, 308–332.

Philippopoulos, M. and Carmay, L. (1994) *J. Phys. Chem.*, **98**, 8264–8273.

Post, C.B. (1992) *J. Mol. Biol.*, **224**, 1087–1101.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988) *Numerical Recipes in C. The art of scientific programming*, Cambridge University Press, Cambridge, U.K.

Rouzaud, D. and Sinaÿ, P. (1983) *J. Chem. Soc. Ser. Chem. Commun.*, 1353–1354.

Rubinstenn, G., Bodenmüller, A., Mallet, J.-M. and Sinaÿ, P. (1996) manuscript in preparation.

Schleucher, J., Quant, J., Glaser, S.J. and Griesinger, C. (1995) *J. Magn. Reson.*, **A112**, 144–151.

Shaka, A.J. and Freeman, R. (1983) *J. Magn. Reson.*, **51**, 169–173.

Sharon, N. and Lis, H. (1993) *Scientific American*, **268**, 74–81.

Singh, K., Fernández-Mayoralas, A. and Martin Lomas, M. (1994) *J. Chem. Soc. Ser. Chem. Commun.*, 775–776.

Solomon, I. (1955) *Phys. Rev.*, **99**, 559–565.

Thorgersen, H., Lemieux, R.U., Bock, K. and Meyer, B. (1982) *Can. J. Chem.*, **60**, 44–57.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Van Halbeeck, H. (1994) *Curr. Opin. Struct. Biol.*, **4**, 697–709.

Varki, A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 7390–7397.

Wang, J., Hodges, R.S. and Sykes, B.D. (1995) *J. Am. Chem. Soc.*, **117**, 8627–8634.

Wagner, G. (1983) *J. Magn. Reson.*, **55**, 151–156.

Wei, A., Boy, K.M. and Kishi, Y. (1995) *J. Am. Chem. Soc.*, **117**, 9432–9436.

Yuen, C.-T., Lawson, A.M., Chai, W., Larkin, M., Stoll, M.S., Stuart, A.C., Sullivan, F.X., Ahern, T.J. and Feizi, T. (1992) *Biochemistry*, **31**, 9126–9131.

Zhao, D. and Jardetzky, O. (1994) *J. Mol. Biol.*, **239**, 601–607.