Data bank of three-dimensional structures of disaccharides, a tool to build 3-D structures of oligosaccharides. Part 1. Oligo-mannose type N-glycans

A. Imberty, Sophie Gerber, V. Tran, Sarah Perez

Glycoconjugate J (1990) 7:27-54

Data Bank of Three-Dimensional Structures of Disaccharides, A Tool to Build 3-D Structures of Oligosaccharides

Part I. Oligo-mannose Type *N*-Glycans

ANNE IMBERTY[1,2]*, SOPHIE GERBER[1], VINH TRAN[1] and SERGE PÉREZ[1]

[1] *Laboratoire de Physicochimie des Macromolécules, INRA, BP 527, 44026 Nantes, Cedex 03, France*
[2] *On post-doctoral leave from CERMAV-CNRS, Grenoble, France*

Received August 16/November 29, 1989.

Key words: oligosaccharide, three dimensional structure, modeling

This work presents the first part of a database of conformations for all the disaccharide fragments that are found in *N*-glycans. The conformational study of the five disaccharides found in the oligo-mannose type are presented here. For each disaccharide, several possible conformations are described. A method is presented to obtain realistic models of oligosaccharides using molecular mechanic methods. Analysis of some possible conformations of the oligo-mannose type glycan Man_6-$GlcNAc_2$ is given as an illustration of the possibilities.

The spatial structures of the glycan chains of glycoproteins are known to be intimately related to their biological functions (see reviews [1-3]) and to be implied in the regulation of the processing of their biosynthetic pathways [4-5]. The sole knowledge of the primary structure of the glycan chains of glycoproteins is not sufficient to understand and explain their functions and specificities. The goal of most molecular modeling studies is to develop a systematic understanding of the relationships between the detailed structural features of the macromolecules and their biological functions and/or physical properties.

X-Ray crystallography is the method of choice to elucidate complex macromolecular structures. It is quite unfortunate that the number of available crystallographic data on glycoproteins, glycopeptides or corresponding oligosaccharides is very limited. The difficulty of growing single crystals of glycoproteins is without any doubt responsible for this. There are also some cases of crystalline glycoproteins where the oligosaccharide part could not be determined through X-ray crystallography because of their high mobility and/or disorder. Actually, the only usable structure of a glycoprotein deposited in the crystallographic Protein Data Bank is the human IgG Fc fragment [6]. Also, difficulties of crystallization may explain why there are only three X-ray structures of oligosaccharides moieties of *N*-glycans which

* Author for correspondence.

have been reported in the literature: one trisaccharide [7] and two disaccharides [8-9]. Despite or because such a lack of available crystallographic data, structural models derived from conformational analysis calculations have been proposed. Molecular modeling has been seldom used as the only tool [10]. It is generally performed in conjunction with NMR spectroscopy data to determine the three-dimensional structure of *N*-linked oligosaccharides in solution [11-13]. Most of these studies deal with di- or trisaccharides. From this work, the concept of the occurrence of a unique and rigid conformation of oligosaccharide in solutions begins to be obsolete, and the flexibility not only of the α(1-6) arm [14] but also of the Manα(1-3)Man linkage [15, 16] has been demonstrated.

In view of these problems specific to the carbohydrate field, alternate routes have to be sought for building and analysing large oligosaccharide structures. The present work describes such an attempt, which follows the establishment of a data bank of the three-dimensional features of the most commonly found monosaccharide units [17]. As a first part, a data bank of the low energy disaccharides contained in the oligo-mannoside type of *N*-glycan (for classification see [18]) has been produced.

Carbohydrates, as a first approximation, are usually considered as rigid segments (the glycosidic rings) separated by flexible segments (the glycosidic linkages). When the ring geometry, and the glycosidic valence angle are fixed at their equilibrium values, the several conformations attainable by the macromolecule are those occurring from rotations about the glycosidic torsion angles: ϕ , ψ [and ω for (1-6)-linkages]. Therefore, computations of individual potential energy surfaces corresponding to each disaccharide fragment are a pre-requisite; they can be adequately computed in (ϕ, ψ, ω) space. Owing to the size and relative rigidity of the intervening sugar residues, the rotations at a particular glycosidic linkage could be virtually independent of the nearest neighbour torsion angles. This is the case of linear polysaccharides for which spatial representations can be described by specifying values for all torsion angles (ϕ_1, ψ_1, ω_1) taken from consecutive dimeric fragments. However, in the case of more complex carbohydrates, the whole structure may not behave as if the glycosidic linkages were independent. Interactions involving non-neighbouring residues can occur, either in the case of α(1-2) linkage, or in the case of branched structures. It has to be emphasized that the occurrence of interactions between different entities, i.e. other branches of the glycan, surface of the glycoprotein or neighbouring molecules, or solvent molecules will only reduce the conformational space available for the oligosaccharide. It is highly improbable that new linkage conformations of high energy, and none beïng predicted by a thorough conformational analysis of the disaccharide, can occur.

Treatment of glycan at that level of complexity can be undertaken using the following route: a) Identification of disaccharide segments of glycoproteins; b) Molecular modeling of each segment, and description of the potential energy surface; c) Collection of the 3-D structures of low energy conformers; d) Building realistic possible starting conformations from assembly of the low energy conformers; e) Energy minimization on the whole structure.

Steps b) and c) are a pre-requisite since they provide the basic information for generating starting conformations prior to any further computations. The information gathered from steps b), i.e potential energy surfaces, and c), i.e. description of all the low energy conformers, constitute the body of the present data base.

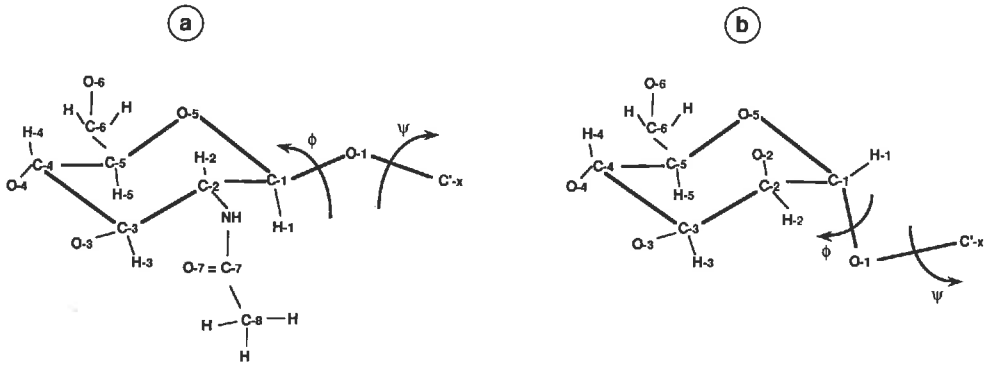

Figure 1. Schematic representation of the monosaccharide units which are found in the oligomannose type *N*-glycan : a) *N*-acetylglucosamine ; b) mannose. The labeling of the atoms, along with the torsion angles of interest are shown. The hydroxylic hydrogen atoms have been omitted.

To illustrate a possible use of this data bank, different conformations of the glycan $Man_6GlcNAc_2$ have been built. This oligo-mannose type glycan is found in various families of glycoproteins, such as hen ovalbumin [19], bovine pancreatic ribonuclease B [20], human β-hexosaminidase [21] and Taka-amylase [22].

Methods

The geometry and conformation of each sugar unit is taken from the data bank of the most commonly found monosaccharide residue. Each residue has been submitted to a complete optimization through the use of a force-field appropriated to the carbohydrate field [17].

For each disaccharide fragment studied here, a schematic drawing is given in Fig. 1. The atomic labeling and the torsional angles taken into account are reported in each scheme (the primed residue corresponding to the reducing one). For glycosidic linkages, the orientation of two linked carbohydrate residues is given by two torsional angles:

$$\phi = \theta(O\text{-}5 - C\text{-}1 - O\text{-}1 - C'\text{-}X)$$
$$\psi = \theta(C\text{-}1 - O\text{-}1 - C'\text{-}X\text{-}C'\text{-}X{+}1)$$

which correspond to the crystallography nomenclature; or by the values of

$$\phi_H = \theta(H\text{-}1 - C\text{-}1 - O\text{-}1 - C'\text{-}X)$$
$$\psi_H = \theta(C\text{-}1 - O\text{-}1 - C'\text{-}X - H'\text{-}X)$$

which correspond to the nomenclature most commonly adopted in NMR studies. In the case

of the α(1-6)-linkage, the second torsional angle is usually referred to as ψ [θ(C-1 - O-1 - C-6 - C-5)] for both crystallography and NMR studies. The orientation of the third torsional angle of the α(1-6)-linkage ω is described as:

$$\omega = \text{O-1 - C'-6 - C-5' - O'-5}$$
or $$\omega_H = \text{O-1 - C'-6 - C'-5 - H'-5.}$$

For each residue the hydroxymethyl group orientation χ is described as:

$$\chi = \text{O-5 - C-5 - C-6 - O-6}$$
or $$\chi_H = \text{H-5 - C-5 - C-6 - O-6.}$$

The conformations of the ω and χ torsional angles are referred to [23] as either *gauche-trans* (GT) (ω or $\chi = 60°$, ω_H or $\chi_H = -60°$), *gauche-gauche* (GG) (ω or $\chi = -60°$, ω_H or $\chi_H = 180°$) and *trans-gauche* (TG) (ω or $\chi = 180°$, ω_H or $\chi_H = 60°$). In this terminology the torsion angle θ (O-5 - C-5 - C-6 - O-6) is stated first and the torsion angle θ (C-4 - C-5 - C-6 - O-6) is stated second. For the disaccharides, the non-reducing end is stated first. The sign of the torsion angles is defined in agreement with the rules recommended by the IUPAC-IUB Joint Commission on Biochemical Nomenclature [24].

"Rigid" Maps Calculations

For a given disaccharide the conformational analysis was performed using the PFOS (Potential Functions for Oligosaccharide Structure) program [25]. The conformational energy is evaluated by including the contributions from van der Waals interactions [26], torsional and exo-anomeric contributions [27]. For each conformation, possible hydrogen bonds were identified using a criterion based on the distance between oxygen atoms (hydrogen donor and acceptor) [28]; each potential energy surface was computed with and without the energy contributions arising from the inter-residue hydrogen bonds. The glycosidic valence angle (C-1 - O-1 - C-X) of each disaccharide was fixed at the value observed in the corresponding crystal structure when available, or, otherwise in the value observed in a related crystal structure. The chosen values are 116.3° for GlcNAcβ(1-4)GlcNAc, 117.7° for Manβ(1-4)GlcNAc, 114.5° for Manα(1-3)Man , 116.0° for Manα(1-2)Man and 111.5° for Manα(1-6)Man. For the χ torsional angles, the only conformations observed in the solid state for carbohydrates with a *gluco* configuration are the GG and GT ones [23]. Recent NMR studies on the coupling constants of selective deuterated mono- and disaccharides confirmed the very low occurrence of the TG conformation in the solution [14, 29] for the *gluco* configuration. The hydroxymethyl group was then fixed in the two allowed conformations for each monosaccharide and then for the disaccharides the four combinations, i.e. GG-GG, GG-GT, GT-GG and GT-GT were studied. The energy maps were calculated as a function of ϕ and ψ at intervals of 5° over the whole angular range. Iso-energy contours are drawn at intervals of 1 kcal/mol with respect to the observed energy minimum of each map.

"Relaxed" Maps Calculations

Starting from the study of the "rigid" potential surfaces another type of calculation was performed, with a relaxation of all other internal coordinates for each conformation. This

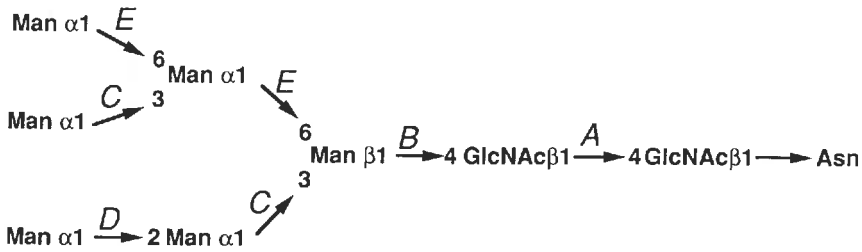

Figure 2. Labelling of the disaccharide fragments constituting the Man_6-$GlcNAc_2$ oligosaccharide.

was realized using the MM2CARB program [30] which is a version of the MM2 force-field program [31] modified with the acetal fragment parameters [32]. This version has been shown to reproduce in an adequate fashion the variation of the carbohydrate ring with the glycosidic linkage conformations observed in the solid state [30]. The energy terms taken into account in the optimization procedure are the stretching, bending, Stretch-bending, torsional, dipolar and van der Waals contributions. The procedure used to calculate the "relaxed" maps of disaccharide has been reported previously [33, 34] and will not be described extensively here. First, each of the low energy conformers detected on the "rigid" potential energy surface is fully optimized. Then, they are used as starting points to explore a 20° by 20° grid of the (ϕ, ψ) space in a progressive way, using the DRIVER option of the MM2 program. At each of these points, the torsion angles ϕ and ψ are fixed, and all the other parameters are refined. Testing different orientations of the hydroxyl hydrogen atoms, and checking for continuity are done thoroughly to avoid being trapped in false minima.

Building Oligosaccharide Models

Different conformations of the *N*-linked oligosaccharide of the oligo-mannose type can be built using this disaccharide data bank. For the model studied here (Fig. 2) different conformations were built in order to show the great range of flexibility. The construction of the models and the orientation of the glycosidic linkage to the desired values was done using the SYBYL software (Tripos Associates Inc, a subsidiary of Evans & Sutherland, St Louis, MO, USA). To test the validity of the models, the total energy of the molecules is computed and the optimization is performed using the MMP2(85) program (Quantum Chemistry Program Exchange, Indiana University, Department of Chemistry, USA). This new version of the MM2 program is not especially modified for saccharides but new functions have been included to reproduce the anomeric effect.

Miscellaneous

All the energy calculations were performed on Dec Work-Stations running under VMS. The model building and graphic representations were realized on an Evans and Sutherland PS-350 graphic station, using the SYBYL software.

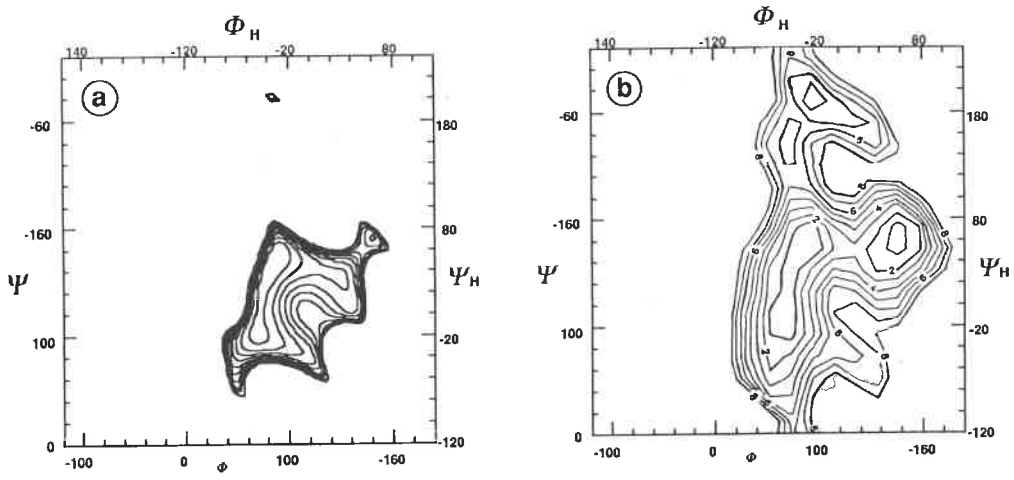

Figure 3. Iso-energy contour maps of the Manα(1-3)Man disaccharide as a function of the torsion angle ϕ and ψ; a) "rigid residue" map calculated with the PFOS program; b) "relaxed residue" map calculated with the MM2CARB program. The hydroxymethyl groups of the two mannose residues are fixed in a GT orientation. The iso-energy contours are drawn by extrapolation of 1 kcal/mol with respect to the absolute minimum of each map. The external contour is 10 kcal/mol for a) and 8 kcal/mol for b).

Results and Discussion

The five disaccharides studied here are the building units of the oligo-mannose type *N*-linked glycans like the one shown in Fig. 2. They are A: GlcNAcβ(1-4)GlcNAc, B: Manβ(1-4)GlcNAc, C: Manα(1-3)Man, D: Manα(1-2)Man and E: Manα(1-6)Man.

"Rigid" Versus "Relaxed" Potential Energy Surfaces

In this section, the approximation which makes use of a "rigid residue" approach, instead of the "flexible" one, is considered. The case of the disaccharide Manα(1-3)Manα(1-*O*)-Me is examined. The "rigid" potential surfaces calculated for a GT-GT orientation of the hydroxymethyl groups are reported in Fig. 3a, whereas the one obtained with the "relaxed" residue approach is shown in Fig. 3b. At a first glance, introducing flexibility in the sugar geometry seems to modify the shape of the potential energy surface. In fact, it can be observed that the locations of the low energy regions remain identical in the "rigid" and in the "relaxed" maps. In the case studied here, the only difference is that the "relaxed" potential energy surface map is enlarged compared to the "rigid" one. The relaxation relieves an important number of small steric conflicts, allowing more conformational states to be reached. The more noticeable expansion is along the ψ axis. This expansion creates a pathway between the main low energy domain and the conformation which is located in

Table 1. Geometrical features, calculated energies (kcal/mol) and possible hydrogen bonds for the GlcNAcβ(1-4)GlcNAc linkage as calculated by the PFOS program (τ = 116.3°). Torsional angles are given in both the crystallography and the NMR nomenclatures, the two sets being presented, respectively, with and without parentheses.

Name	ϕ, ψ	(ϕ_H, ψ_H)	χ, χ'	(χ_H, χ'_H)	Relative energy	Possible hydrogen bond
A1	-80, -110	(40, 10)	-60, -60	(180,180)	0.13	O-5.....O-3'
			60,-60	(-60,180)	0.07	O-5.....O-3'
	-85, -105	(35, 15)	-60, 60	(180,-60)	0.05	O-5.....O-3'
			60, 60	(-60,-60)	0.00	O-5.....O-3'
A2	-90, -175	(30, -55)	-60, -60	(180, 180)	0.52	O-5.....O-3'
			60, -60	(-60, 180)	0.35	O-5.....O-3'
			-60, 60	(180, -60)	0.74	O-5.....O-3'
			60, 60	(-60, -60)	0.57	O-5.....O-3'
A3	-95, 55	(25, 175)	-60, -60	(180, 180)	1.61	
			60, -60	(-60, 180)	1.45	
			-60, 60	(180, -60)	1.52	
			60, 60	(-60, -60)	1.30	O-6.....O-6'
A4	-50, 90	(70, -150)	-60, -60	(180, 180)	1.98	
			60, -60	(-60, 180)	1.65	
			-60, 60	(180, -60)	2.05	
			60, 60	(-60, -60)	1.61	
A5	25, -110	(145, 10)	-60, -60	(180, 180)	2.80	O-7.....O-3'
						O-5.....O-6'
	25, -115	(145, 5)	60, -60	(-60, 180)	3.13	O-7.....O-3'
						O-5.....O-6'
						O-6.....O-6'
	25, -105	(145, 15)	-60, 60	(180, -60)	2.47	O-7.....O-3'
			60, 60	(-60, -60)	2.27	O-7.....O-3'
A6	25, -170	(145, -50)	-60, -60	(180, 180)	3.22	O-5.....O-3'
			60, -60	(-60, 180)	3.39	O-5.....O-3'
			-60, 60	(180, -60)	3.41	O-5.....O-3'
			60, 60	(-60, -60)	3.49	O-5.....O-3'

an island in the "rigid" potential energy surface. Ideally, the computation of such a "relaxed" map should be done for each of the disaccharides studied here. This is impossible because of the computation time needed for such a study. Nevertheless, the approximation done by using "rigid" maps for this data bank can be justified. We are mainly interested here in the conformation of the energy minima and, from this study as well as from other comparisons between "relaxed" MM2 maps and "rigid" PFOS maps of disaccharides [33, 34], it appears that both methods give very comparable conformations of the low energy minima; using the two methods, differences in conformations appear mostly in the region of relatively higher energy. It has to be emphasized that such approximation must be limited, at the moment,

Table 2. Geometrical features, calculated energies (kcal/mol) and possible hydrogen bonds for the Manβ(1-4)GlcNAc linkage as calculated by the PFOS program ($\tau = 117.7°$). Torsional angles are given in both the crystallography and the NMR nomenclatures, the two sets being presented, respectively, with and without parentheses.

Name	ϕ, ψ	(ϕ_H, ψ_H)	χ, χ'	(χ_H, χ'_H)	Relative energy	Possible hydrogen bond
B1	-75, -110	(45, 10)	-60, -60	(180, 180)	0.06	O-5.....O-3'
			60, -60	(-60, 180)	0.00	O-5.....O-3'
	-75, -105	(45, 15)	-60, 60	(180, -60)	0.16	O-5.....O-3'
			60, 60	(-60, -60)	0.12	O-5.....O-3'
B2	-100, -165	(20, -45)	-60, -60	(180, 180)	0.39	O-5.....O-3'
			60, -60	(-60, 180)	0.24	O-5.....O-3'
			-60, 60	(180, -60)	0.43	O-5.....O-3'
			60, 60	(-60, -60)	0.29	O-5.....O-3'
B3	-150, -140	(-30, -20)	-60, -60	(180, 180)	0.77	
			60, -60	(-60, 180)	0.76	
			-60, 60	(180, -60)	0.91	
			60, 60	(-60, -60)	0.90	
B4	-95, 50	(25, 170)	-60, -60	(180, 180)	1.17	
			60, -60	(-60, 180)	0.96	
			-60, 60	(180, -60)	1.15	
			60, 60	(-60, -60)	0.92	O-6.....O-6'
B5	35, -110	(155, 10)	-60, -60	(180, 180)	3.78	O-2.....O-3'
						O-5.....O-6'
						O-6.....O-6'
			60, -60	(-60, 180)	3.82	O-2.....O-3'
						O-5.....O-6'
						O-6.....O-6'
	30, -105	(150, 15)	-60, 60	(180, -60)	3.31	O-2.....O-3'
			60, 60	(-60, -60)	3.22	O-2.....O-3'
B6	-50, 85	(70, -155)	-60, -60	(180, 180)	4.21	
			60, -60	(-60, 180)	3.90	
			-60, 60	(180, -60)	4.17	
			60, 60	(-60, -60)	3.87	
B7	175, 180	(-65, -60)	-60, -60	(180, 180)	6.12	
			60, -60	(-60, 180)	6.12	
			-60, 60	(180, -60)	6.30	
			60, 60	(-60, -60)	6.30	

to the case of hexopyranose rings. Therefore, for this database, it is possible to use the disaccharide conformations detected on the "rigid" maps for further model building, providing, 1) that all the low energy conformations, even the weak and/or isolated ones, are taken into consideration, and 2) that an optimization of the whole macromolecule is performed after its synthesis.

Table 3. Geometrical features, calculated energies (kcal/mol) and possible hydrogen bonds for the Manα(1-3)Man linkage as calculated by the PFOS program (τ = 114.5°). Torsional angles are given in both the crystallography and the NMR nomenclatures, the two sets being presented, respectively, with and without parentheses.

Name	ϕ, ψ	(ϕ_H, ψ_H)	χ, χ'	(χ_H, χ'_H)	Relative energy	Possible hydrogen bond
C1	85, 175	(-35, 55)	-60, -60	(180, 180)	0.17	O-5.....O-2'
			60, -60	(-60, 180)	0.10	O-5.....O-2'
						O-6.....O-2'
			-60, 60	(180, -60)	0.18	O-5.....O-2'
			60, 60	(-60, -60)	0.00	O-5.....O-2'
						O-6.....O-2'
C2	70, 100	(-50, -20)	-60, -60	(180, 180)	0.82	
			60, -60	(-60, 180)	0.68	
			-60, 60	(180, -60)	0.83	
			60, 60	(-60, -60)	0.68	
C3	150, 150	(30, 30)	-60, -60	(180, 180)	2.82	
			60, -60	(-60, 180)	2.80	
			-60, 60	(180, -60)	2.82	
			60, 60	(-60, -60)	2.80	
C4	125, 70	(5, -50)	-60, -60	(180, 180)	4.33	
			60, -60	(-60, 180)	4.31	
			-60, 60	(180, -60)	4.35	
			60, 60	(-60, -60)	4.33	
C5	180, -170	(60, 70)	-60, -60	(180, 180)	5.72	
			60, -60	(-60, 180)	5.71	
			-60, 60	(180, -60)	5.72	
			60, 60	(-60, -60)	5.71	
C6	85, -40	(-35, -160)	-60, -60	(180, 180)	8.39	O-5.....O-4'
			60, -60	(-60, 180)	8.76	O-5.....O-4'
						O-6.....O-2'
			-60, 60	(180, -60)	8.44	O-5.....O-4'
			60, 60	(-60, -60)	8.81	O-5.....O-4'
						O-6.....O-2'

The Data Bank

The body of the data bank is constituted by the potential energy surfaces, along with the description of the three-dimensional features of all the local minima which are located on these maps. All the potential energy surfaces were calculated using the PFOS program for the four possible orientations of the hydroxymethyl groups (GG-GG, GT-GG, GG-GT and GT-GT). Then, all the local energy minima were looked for on these surfaces. The characteristics of all these minima are listed in Tables 1-5, corresponding respectively to

Table 4. Geometrical features, calculated energies (kcal/mol) and possible hydrogen bonds for the Manα(1-2)Man linkage as calculated by the PFOS program (τ = 116.0°). Torsional angles are given in both the crystallography and the NMR nomenclatures, the two sets being presented, respectively, with and without parentheses.

Name	ϕ, ψ	(ϕ_H, ψ_H)	χ, χ'	(χ_H, χ'_H)	Relative energy	Possible hydrogen bond
D1	90, 180	(-30, 60)	-60, -60	(180, 180)	0.35	
			60, -60	(-60, 180)	0.00	O-6.....O-5'
						O-6.....O-6'
			-60, 60	(180,-60)	0.61	
			60, 60	(-60, -60)	0.18	O-6.....O-5'
D2	70, 100	(-50, -20)	-60, -60	(180, 180)	1.56	
			60, -60	(-60, 180)	1.45	
			-60, 60	(180, -60)	1.60	
			60, 60	(-60, -60)	1.49	
D3	150, 150	(30, 30)	-60, -60	(180, 180)	3.54	
			60, -60	(-60, 180)	3.50	
			-60, 60	(180, -60)	3.59	
			60, 60	(-60, -60)	3.55	
D4	120, 70	(0, -50)	-60, -60	(180, 180)	4.70	
			60, -60	(-60, 180)	4.68	
			-60, 60	(180, -60)	4.80	
			60, 60	(-60, -60)	4.77	
D5	-175, -170	(65, 70)	-60, -60	(180, 180)	7.79	
			60, -60	(-60, 180)	7.77	
			-60, 60	(180, -60)	7.83	
			60, 60	(-60, -60)	7.81	

GlcNAcβ(1-4)GlcNAc, Manβ(1-4)GlcNAc, Manα(1-3)Man, Manα(1-2)Man and Manα(1-6)Man. The iso-energy maps which are represented in Figs. 4-9 are calculated for a GT-GT orientation of the hydroxymethyl group. As the orientation of the O-6 atom does not affect either the overall shapes of the potential surfaces or the occurrence of the low energy conformations, the maps corresponding to other orientations (GG-GG, GT-GG and GG-GT) are not represented here.The only differences seen in these other orientations are very small variations in the ϕ, ψ and energy values as can be observed in Tables 1-4. For the Manα(1-6)Man disaccharide, the two orientations of the ω angle at the glycosidic linkage were considered (Fig. 8), since in this case, the potential energy surfaces are different. All the (ϕ, ψ) maps were calculated with and without the contributions of the hydrogen bonds. The ones shown here are calculated without the hydrogen bond contribution, since it has been verified that the inclusion of this contribution does not alter the shape of the energy surfaces. Furthermore, no change in the locations of the low energy minima was detected. Only the magnitude of the energy of some of them varied. It has been checked that no new minima appear. Calculations of the potential energy surface without the hydrogen bond contribution

Table 5. Geometrical features, calculated energies (kcal/mol) and possible hydrogen bonds for the Manα(1-6)Man linkage as calculated by the PFOS program (τ = 111.5°). Torsional angles are given in both the crystallography and the NMR nomenclatures, the two sets being presented, respectively, with and without parentheses.

Name	ϕ, ψ	(ϕ_H, ψ)	χ, ω	(χ_H, ω_H)	Relative energy	Possible hydrogen bond
Egg1	70, -170	(-50, -170)	-60, -60	(180, 180)	0.18	
			60, -60	(-60, 180)	0.00	
Egg2	75, 100	(-45, 100)	-60, -60	(180, 180)	2.44	
			60, -60	(-60, 180)	2.39	
Egg3	150, -105	(30, -105)	-60, -60	(180, 180)	4.36	
	150, -115	(30, -115)	60, -60	(-60, 180)	4.29	
Egg4	160, 120	(40, 120)	-60, -60	(180, 180)	5.10	
			60, -60	(-60, 180)	5.09	
Egt1	70, -170	(-50, -170)	-60, 60	(180, -60)	1.00	
			60, 60	(-60, -60)	0.87	
Egt2	75, 90	(-45, 90)	-60, 60	(180, -60)	2.04	
			60, 60	(-60, -60)	1.98	
Egt3	150, -100	(30, -100)	-60, 60	(180, -60)	4.90	
			60, 60	(-60, -60)	4.87	
Egt4	165, 115	(45, 115)	-60, 60	(180, -60)	4.39	
			60, 60	(-60, -60)	4.38	
Egt5	105, -80	(-15, -80)	-60, 60	(180, -60)	5.72	

seem to be an appropriate way to model solution behaviour of oligosaccharides. It may be advocated that an aqueous surrounding medium will favor hydrogen bonds between the molecule and the solvent instead of the formation of intramolecular hydrogen bonding. This statement may not always be true, since in particular cases intramolecular hydrogen bonds may be formed as a further stabilization of an already existing low energy conformation. For example, this is frequently the case for crystalline compounds. Therefore, in order to present the most exhaustive description, all the possible hydrogen bonds which can occur for the different low energy conformations have been searched; they are included in Tables 1-5 and the potential maps calculated with the energy contribution arising from hydrogen bonding have been deposited as supplementary material. To evaluate the validity of the calculations, the conformations observed in the solid state have been reported in the corresponding figures. It can be shown that none of them lies outside of the calculated low energy regions (less than 10 kcal/mol above the minimum). The body of the results gathered in the present

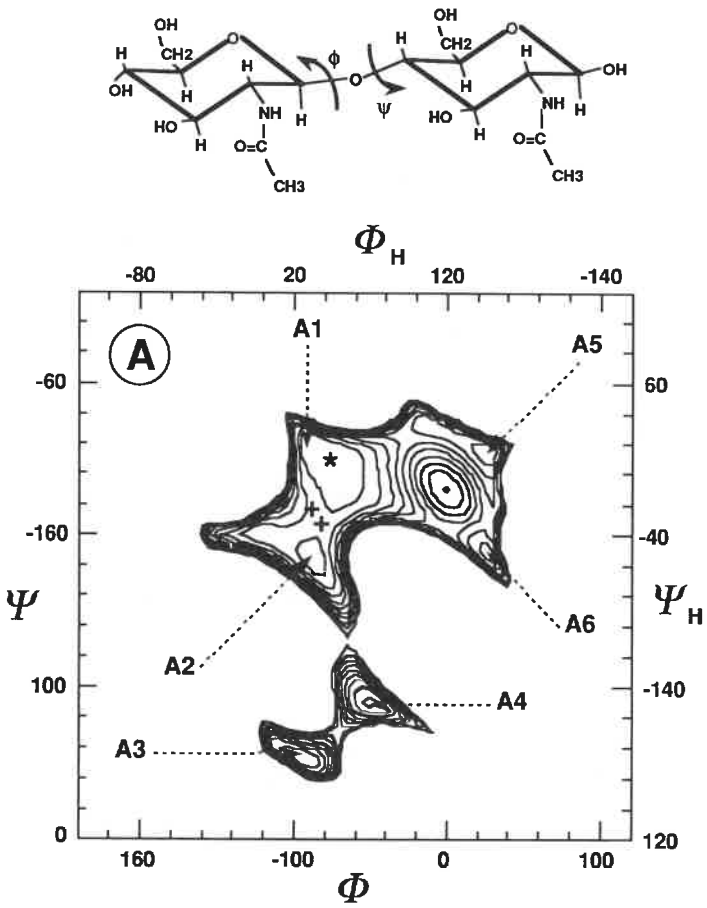

Figure 4. Iso-energy map of the GlcNAcβ(1-4)GlcNAc disaccharide as a function of the ϕ and ψ torsion angles. The two hydroxymethyl groups are fixed in a GT orientation. Iso-energy contours are drawn by extrapolation of 1 kcal/mol with respect to the absolute minimum (A1). All the energy minima, A1 to A6, have been located on the surface. The naming of the axes at the top and bottom are different since they correspond to the crystallographic and NMR nomenclature, respectively, for torsional angles (see the Methods section for definition). The same conventions are used for the left and right axes. The conformations of the reported crystal structures are shown on the map by the signs * [9] and + [6].

work is in itself self-explanatory. In all the cases, the low energy domains, as delineated by the 10 kcal/mol iso-energy contours, encompass between 10 to 40% of the total available ϕ, ψ space. Very often, distinct regions define the low-energy domains. For each disaccharide, in the low energy domains, between four and seven conformers corresponding to local

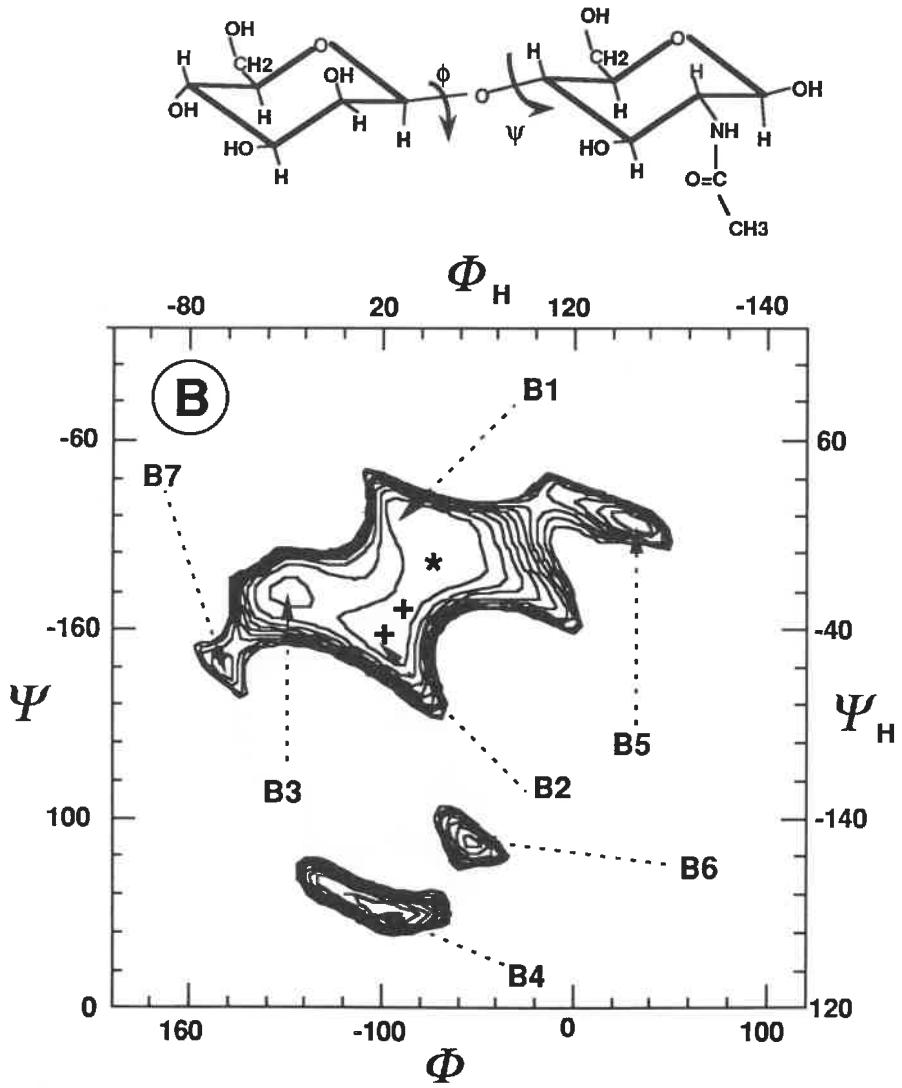

Figure 5. Iso-energy map of the Manβ(1-4)GlcNAc disaccharide as a function of the ϕ and ψ torsion angles. The legend is the same as in Fig. 4, the energy minima being labeled from B1 to B7. The reported conformations derived from crystal structures are shown on the map by the signs * [7] and + [6].

minima can be identified. All of them can be considered as valid candidates in generating starting conformations for the *N*-glycan moiety. Drawings of the conformation of all the local minima located in the energy maps have been deposited as supplementary material.

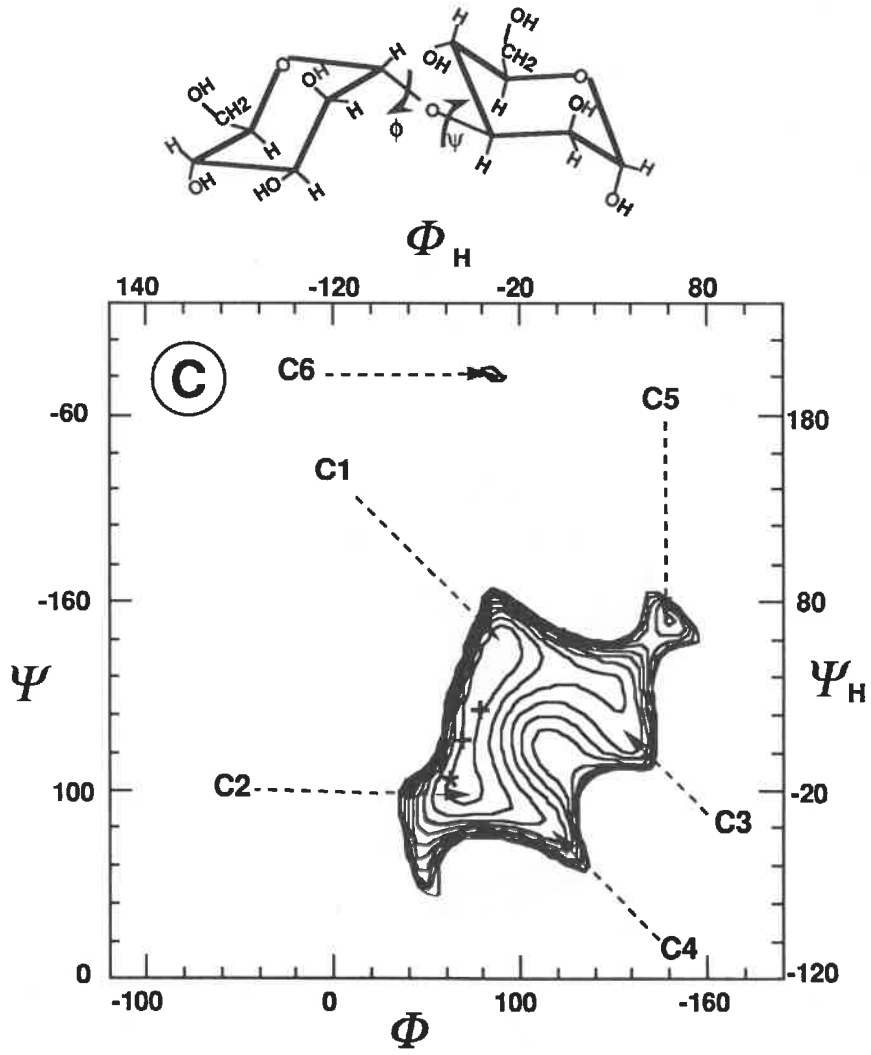

Figure 6. Iso-energy map of the Manα(1-3)Man disaccharide as a function of the ϕ and ψ torsion angles. The legend is the same as in Fig. 4, the energy minima being labeled from C1 to C6. The reported conformations derived from crystal structures are shown on the map by the signs * [7] and + [6].

Model Building and Optimization of the Oligosaccharide Conformations

Combining the different low energy conformers would give 756,000 possible starting models for the $Man_6GlcNAc_2$ oligosaccharide shown in Fig. 1. Even though most of these combinations would not correspond to stable arrangements, such a number gives a good

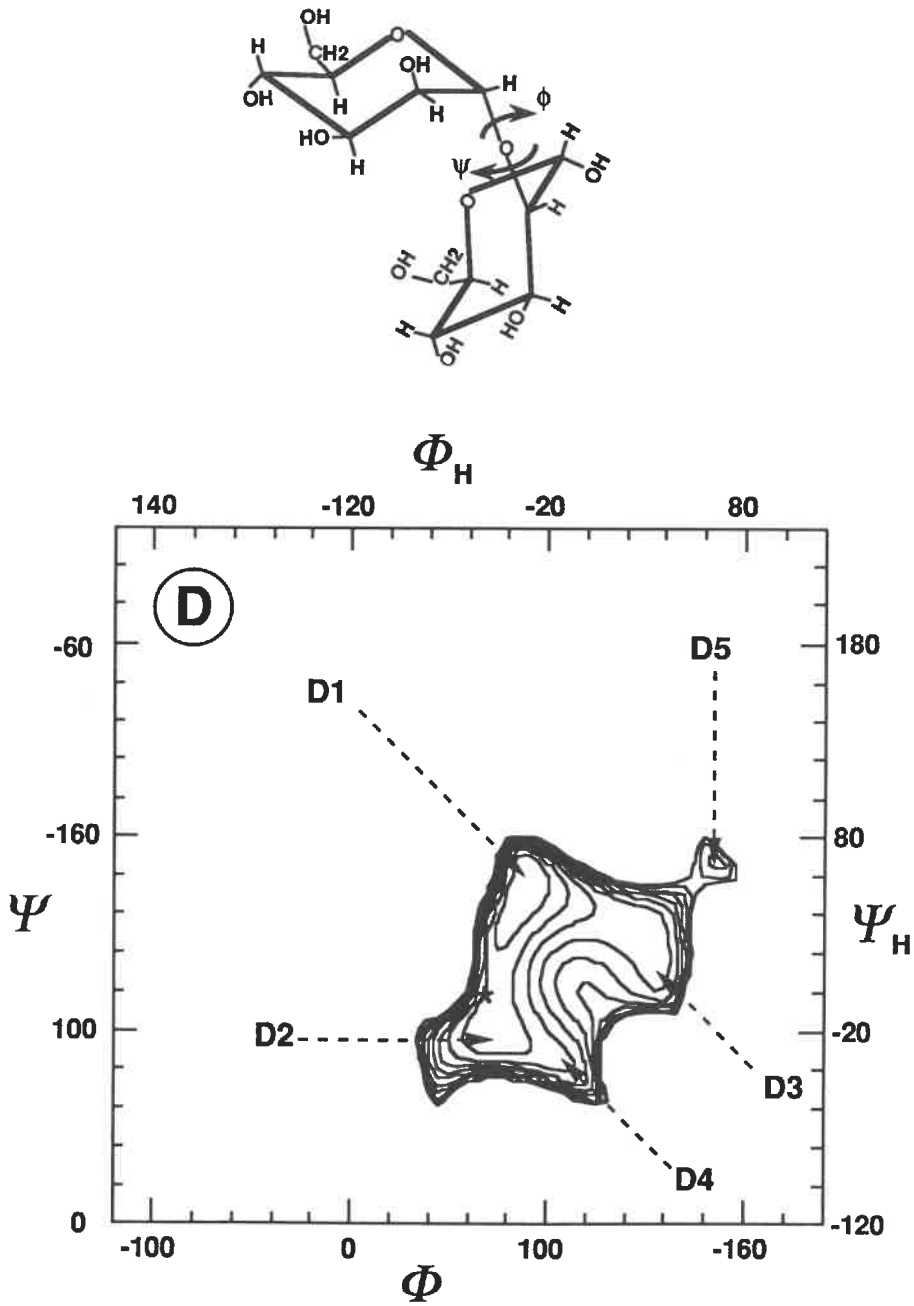

Figure 7. Iso-energy map of the Manα(1-2)Man disaccharide as a function of the ϕ and ψ torsion angles. The legend is the same as in Fig. 4, the energy minima being labeled from D1 to D5. The reported conformation derived from one crystal structure is shown on the map by the sign * [8].

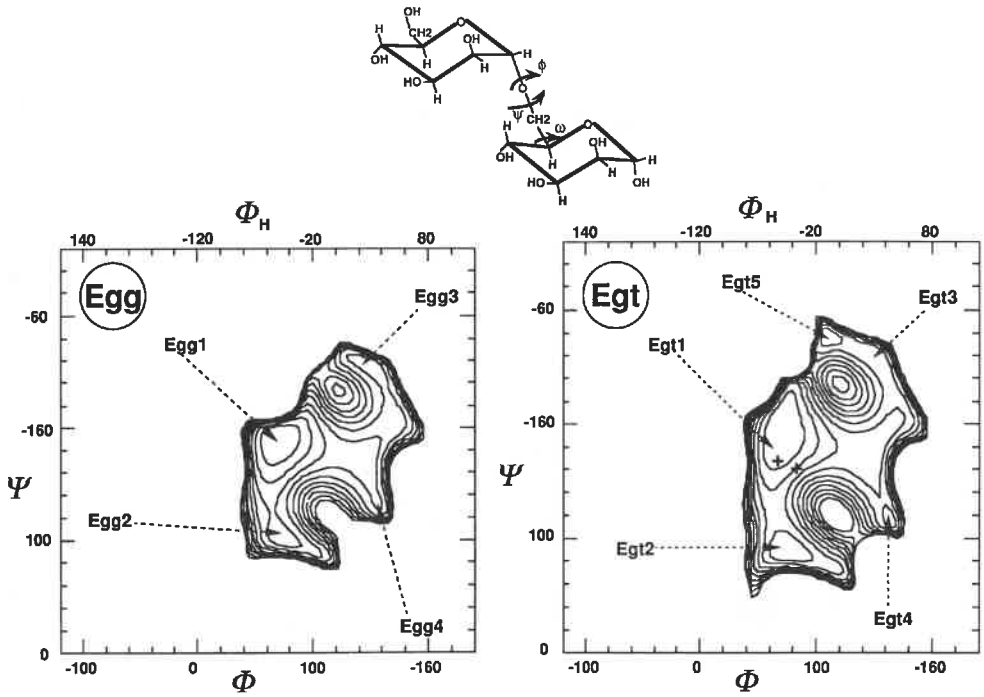

Figure 8. Iso-energy maps of the Manα(1-6)Man disaccharide as a function of the ϕ and ψ torsion angles. The third torsion angle of the linkage is fixed in the GG orientation. (ω = -60°, ω_H = 180°) in the left panel and in the GT orientation (ω = +60°, ω_H = -60°) in the right panel. The legend is the same as in Fig. 4, the energy minima being labeled from Egg1 to Egg4 and from Egt1 to Egt5. The reported conformation derived from one crystal structure is shown on the corresponding map by the + signs [6].

idea of the flexibility which may occur in these oligosaccharide structures. In order to illustrate such a flexibility, six starting conformations of the $Man_6GlcNAc_2$ oligosaccharide (Fig. 1) have been generated. The linkage conformations have been chosen among the ones with the lowest energies. The first two models have been built with all their linkages in the lowest minima: GlcNAcβ(1-4)GlcNAc (A1), Manβ(1-4)GlcNAc (B1), Manα(1-3)Man (C1), Manα(1-2)Man (D1), for Manα(1-6)Man (E1) the orientations of ω or χ in the two α(1-6) linkages were fixed both in a GG orientation (model GG1) or both in a GT orientation (model GT1). In the same way, for the two following models, all the conformations of the glycosidic linkages were fixed in the second lower energy minima of the disaccharide maps (A2, B2, C2, D2, and E2) creating the GG2 and GT2 oligosaccharides. The last two models (GG3 and GT3) were built from the third lower conformations (A3, B3, C3, D3 and E3).

These six different conformations, arbitrarily selected from the 756,000 possibilities, are displayed in Fig. 9. For the sake of comparison, the projection of these drawings is such that the orientation of the *N*-acetylglucosamine residue linked to the asparagine is the same for

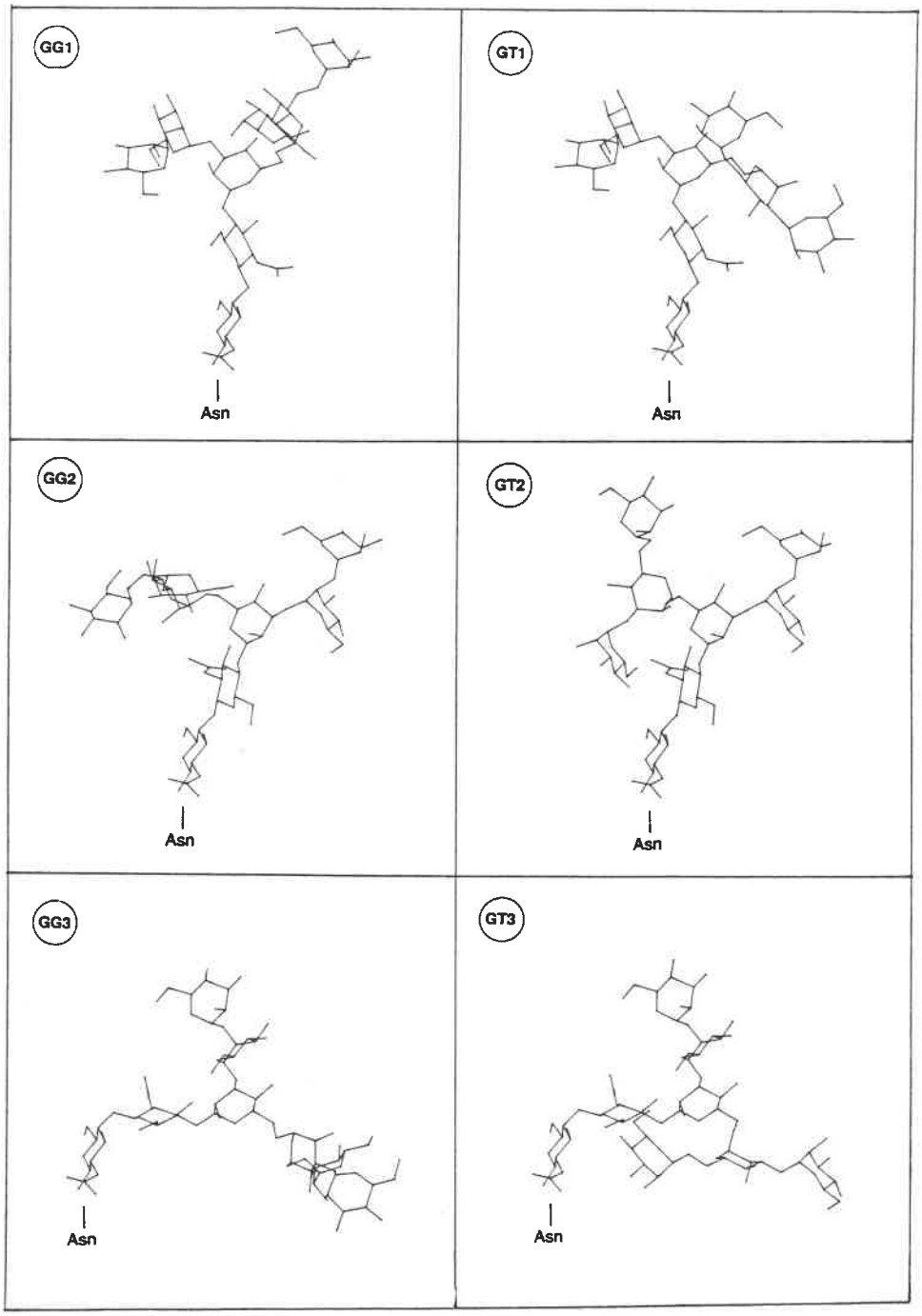

Figure 9. Six different models of the Man_6-$GlcNAc_2$ oligosaccharide built from the calculated low energy minima of each disaccharide.

Table 6. Geometrical features of six different conformations of Man_6-$GlcNAc_2$ before and after minimization by the MM2P85 program. The energy values (kcal/mol from MM2P85) are relative values with respect to the absolute minima (GT3). Non-optimized conformation (*).

	GG1		GT1		GG2		GT2		GG3		GT3	
	PFOS	**MM2P85**	PFOS	**MM2P85**	PFOS	**MM2P85**	PFOS	**MM2P85**	PFOS	**MM2P85**	PFOS	**MM2P85**
Linkage A GlcNAcβ(1-4)GlcNAc												
τ	116.3	*	116.3	*	116.3	*	116.3	*	116.3	*	116.3	*
ϕ	-85	*	-85	*	-90	*	-90	*	-95	*	-95	*
ψ	-105	*	-105	*	-175	*	-175	*	55	*	55	*
Linkage B Manβ(1-4)GlcNAc												
τ	117.7	**114.0**	117.7	**114.0**	117.7	**115.6**	117.7	**115.4**	117.7	**113.9**	117.7	**117.7**
ϕ	-70	**-75.0**	-70	**-81.1**	-100	**-103.4**	-100	**-108.4**	-150	**-152.2**	-150	**179.6**
ψ	-110	**-111.6**	-110	**-109.9**	-165	**-170.5**	-165	**-172.7**	-140	**-149.2**	-140	**178.4**
Linkage Ca Manα(1-3)Man												
τ	114.5	**115.6**	114.5	**115.7**	114.5	**113.2**	114.5	**113.2**	114.5	**113.4**	114.5	**115.3**
ϕ	85	**89.6**	85	**89.5**	70	**73.0**	70	**69.5**	150	**145.2**	150	**100.4**
ψ	175	**174.7**	175	**175.5**	100	**104.9**	100	**111.2**	150	**153.1**	150	**169.4**
Linkage D Manα(1-2)Man												
τ	116.0	**117.2**	116.0	**117.3**	116.0	**116.7**	116.0	**117.0**	116.0	**113.8**	116.0	**117.1**
ϕ	90	**94.7**	90	**84.8**	70	**68.8**	70	**66.3**	150	**139.6**	150	**86.3**
ψ	180	**162.8**	180	**163.2**	100	**97.7**	100	**96.8**	150	**146.5**	150	**166.5**
Linkage Ea Manα(1-6)Man												
τ	111.5	**113.5**	111.5	**114.3**	111.5	**114.4**	111.5	**115.1**	111.5	**113.9**	111.5	**113.1**
ϕ	70	**70.9**	70	**68.1**	75	**74.8**	75	**71.6**	150	**152.2**	150	**160.8**
ψ	-170	**-179.5**	-170	**179.9**	100	**93.8**	90	**74.3**	-105	**-113.7**	-100	**-164.3**
ω	-60	**-66.2**	60	**58.7**	-60	**-70.1**	60	**51.6**	-60	**-72.8**	60	**69.6**
Linkage Cb Manα(1-3)Man												
τ	114.5	**116.5**	114.5	**116.3**	114.5	**115.8**	114.5	**117.8**	114.5	**114.2**	114.5	**116.3**
ϕ	85	**85.2**	85	**83.6**	70	**66.0**	70	**64.2**	150	**138.7**	150	**144.2**
ψ	175	**163.4**	175	**161.8**	100	**102.6**	100	**83.5**	150	**148.1**	150	**115.2**
Linkage Eb Manα(1-6)Man												
τ	111.5	**114.7**	111.5	**115.1**	111.5	**115.2**	111.5	**115.8**	111.5	**114.6**	111.5	**115.7**
ϕ	70	**67.4**	70	**65.5**	75	**72.1**	75	**68.1**	150	**147.5**	150	**126.6**
ψ	-170	**179.2**	-170	**177.3**	90	**93.4**	90	**80.8**	-100	**-115.6**	-100	**-104.4**
ω	-60	**-66.6**	60	**58.8**	-60	**-68.6**	60	**54.9**	-60	**-65.4**	60	**80.1**
Energy (kcal/mol) from MM2P85												
		8.9		**0.0**		**16.7**		**15.5**		**34.6**		**11.7**

all the models. It appears that five of them are devoid of steric conflicts, whereas the sixth one (GT3) exhibits some serious steric conflicts between the second *N*-acetylglucosamine residue and the terminal mannose residue of the (1-6) arm. The two models built from the lowest energy conformations (GG1 and GT1) correspond closely to the classical models of *N*-linked glycan: the "Y" model (GG1) and the "broken-wing" model (GT1) [1]. However, these two models represent only two points of the hyperspace of possible conformations. The other models exhibit rather different features. With respect to the GG1 and GT1 models, the directions of GG2 and GT2 arms have undergone a rotation of about 180°. The GG3 and GT3 models are characterized by a curvature of the oligosaccharide which, in the case of a glycoprotein would probably generate interactions, either favorable or unfavorable, with the protein surface.

These models, generated from the disaccharide potential energy surfaces, need to be further optimized through molecular mechanic calculations. This was achieved through MMP2(85) calculations. If the starting models do not exhibit steric conflict between non-adjacent residues, the optimization will only adjust the linkages and ring conformations to the optimum geometry, and the linkages will remain in the region of the starting low energy minima. In the case where steric conflicts occur, the optimization will give different results depending upon the severity of the van der Waals surfaces interpenetrations, and upon the position of the atoms involved. In some cases, the steric conflicts will be relieved only by small adjustments of some side groups, with no concomitant change of the linkages conformations. Alternatively, in the case of severe steric conflicts involving numerous atoms, the whole geometry and conformation of the oligosaccharide will be altered and new minima will be reached for some linkages conformations.

The optimization was performed on the six starting models. The initial and optimized conformations are listed in Table 6. Since the present version of MMP2(85) cannot handle more than 255 atoms (including hydroxylic hydrogens and lone pairs), the first *N*-acetylglucosamine residue linked to the asparagine was not included in the optimization. This residue does not contribute to through-space interactions in these instances, and its omission does not influence the overall results. It appears from Table 6, that for the first five models (GG1, GT1, GG2,GT2 and GG3), the optimizations yield only small changes from the initial conformations. The most important variation for the τ angle is 4.3° which is reasonable, and in most cases it is less than 2° . The ϕ and ψ torsion angles always remain in the low energy region corresponding to the starting points, the largest observed variations being less than 20°. The torsion angles, ω and χ remain in the starting orientation (either GG or GT) with a maximum variation amplitude of 15°. This behaviour is different for the model GT3, which exhibited large steric conflicts.

The geometries of the GT3 model before and after optimization are displayed in Fig. 10a and 10b, respectively. In Fig. 10b, the *N*-acetylglucosamine residue linked to asparagine has been added in its not optimized orientation with the aim of comparing with Fig. 10a. It can be observed that the optimization relieved all the steric conflicts by driving the oligosaccharide in a fairly different conformation. As can be seen in Table 6, two linkages are mainly responsible for the observed modifications. The Manα(1-2)Man linkage switches from the minimum D3 to the vicinity of the minimum D1 (the one with the lowest energy). For the Manβ(1-4)GlcNAc linkage, the conformation ends up in a small region of the potential

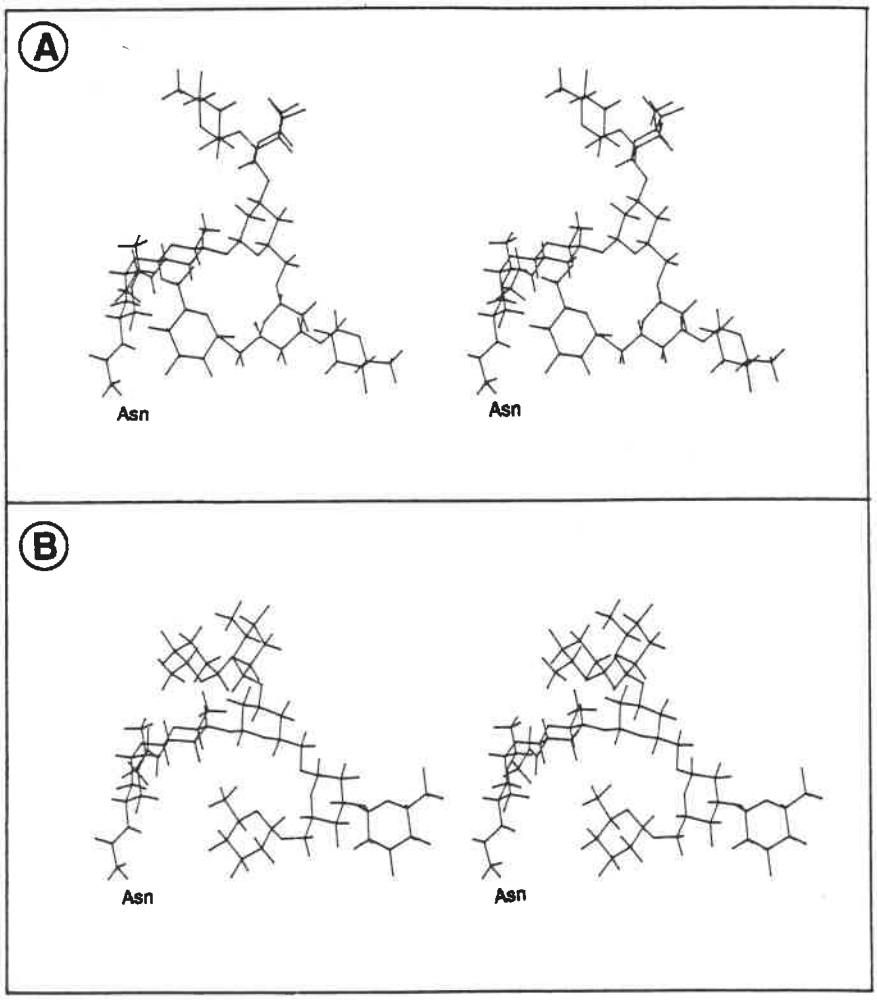

Figure 10. Stereoscopic representations of the GT3 model of the Man_6-$GlcNAc_2$ oligosaccharide: a) built from the third low energy minima of each disaccharide (A3, B3, C3, D3, Egt3); b) same oligosaccharide after minimization through MMP2(85) calculations.

surface, centered around $\phi = 180°$, $\psi = 180°$, corresponding to the B7 minimum. Therefore, the optimization procedure drove the oligosaccharide into another conformation which is one among the numerous possibilities accessible from the data bank of disaccharides.

The relative energy of each of the six optimized models is given in Table 6. Within the range of calculated energy difference, all the models which have been considered may be viable.

It appears that the oligosaccharides built using the linkage conformations of lowest energy, i.e. GG1 and GT1, present the lowest energy. This remains true as long as no steric conflict appears in the molecule. For the GT3 model, the optimization yielded a different combination of low energy minima with a very favorable global energy. It has to be emphasized that many other low energy combinations could be built, and the GT1 model, which has the lowest energy in our limited sample of combinations, may not have the lowest energy among the 756,000 possibilities.

Conclusion

The conformational study of the five disaccharide fragments which constitute the oligomannosidic type of *N*-glycan has been presented as part of a data bank on the three-dimensional structures of disaccharides. Precise information about the low energy domains, along with the description of the features displayed by all the local minima have been provided. A net consequence of these results is the fairly high number of stable conformers for each disaccharide. This clearly indicates that depending upon conditions imposed to the macromolecule, i.e., solvation, temperature, interactions with a neighbouring protein surface, etc., conformational transitions are expected to occur. As an example, analysis of some possible conformations of the oligomannose type glycan $Man_6GlcNAc_2$ has been performed.

The data bank of disaccharide conformations presented here, along with the procedures for building and optimizing oligosaccharide structures, would be helpful for different research areas, and will allow some new approaches in the field of three dimensional structures of *N*-linked glycans. Among the on-going research work, protein crystallographers, when involved in the refinement of a glycoprotein structure, may find it interesting to check whether the model they proposed for the glycan part is compatible with the stereochemical properties of carbohydrates.

Other fields of structural investigations, such as high resolution NMR spectroscopy, require the knowledge of all the possible conformations occurring in solution in order to calculate an averaged value of the data of interest (nuclear Overhauser effect, relaxation time, coupling constants, etc.) for comparison with the observed one. The present work provides the basic set of information for either molecular dynamics simulations, or Monte Carlo calculations. Oligosaccharide model building using a reasonable approach is also of interest for different studies on glycoproteins. For example, it would be possible to evaluate the surfaces of the protein which can be protected by the *N*-glycan against proteolysis, or to check the possible masking of epitopes or of active sites. The extension of this data base to all the disaccharides involved in *N*-glycans has been done in our laboratory and will be published soon.

Acknowledgements

Part of this work was supported by the French Ministère de la Recherche et de l'Enseignement Supérieur, Grant No. MRES87 T O 261.

References

1 Montreuil J (1984) Biol Cell 51:115-31.
2 Montreuil J (1984) Pure Appl Chem 56:859-77.
3 Rademacher TW, Parekh RB, Dwek RA (1988) Ann Rev Biochem 57:785-838.
4 Brisson JR, Carver JP (1983) Can J Biochem Cell Biol 61:1067-78.
5 Carver JP, Cumming DA (1987) Pure Appl Chem 59:1465-76.
6 Deisenhofer J (1981) Biochemistry 20:2361-70.
7 Warin W, Baert F, Fouret R, Strecker G, Spik G, Fournet B, Montreuil J (1979) Carbohydr Res 76:11-22.
8 Srikrishnan T, Chowdhary MS, Matta KL (1989) Carbohydr Res 186:167-75.
9 Mo F, Jensen LH (1978) Acta Crystallogr B34:1562-69.
10 Biswas M, Sekharudu YC, Rao VSR (1988) Int J Biol Macromol 8:2-8.
11 Cumming DA, Carver JP (1987) Biochemistry 26:6664-76.
12 Homans SW, Dwek RA, Rademacher TW (1987) Biochemistry 26:6571-78.
13 Breg J, Kroon-Batenburg LMJ, Strecker G, Montreuil J, Vliegenthart JFG (1989) Eur J Biochem 178:727-39.
14 Cumming DA, Carver JP (1987) Biochemistry 26:6676-83.
15 Carver JP, Michnick SW, Imberty A, Cumming DA (1989) in Oligosaccharide/Protein Interactions : A Three-Dimensional View, eds. Bock G, Harnett S, John Wiley & Sons, England.
16 Imberty A, Tran V, Pérez S (1989) J Comp Chem, in press.
17 Data base on three-dimensional structures of monosaccharides. Information available from Dr. S. Pérez & M.M. Delage, LPCM-INRA, BP527, 44026 Nantes, France.
18 Montreuil J (1980) Adv Carbohydr Chem Biochem 37:157-223.
19 Carver JP, Grey AA, Winnik FM, Hakimi J, Ceccarini C, Atkinson PH (1981) Biochemistry 20:6600-6.
20 Berman E, Walters DE, Allerhand A (1981) J Biol Chem 256:3853-57.
21 O'Dowd BP, Cumming DA, Gravel RA, Mahuran D (1988) Biochemistry 27:5216-26.
22 Yamagushi H, Ikenaka T, Matsushima Y (1971) J Biochem (Tokyo) 70:587-94.
23 Marchessault RH, Pérez S (1979) Biopolymers 18:2369-74.
24 IUPAC-IUB Commission on Biochemical Nomenclature (1971) Arch Biochem Biophys 145:405-21.
25 Pérez S (1978) DSc Thesis, Grenoble, France.
26 Scott RA, Scheraga HA (1965) J Chem Phys 42:2209-15.
27 Tvaroska I (1984) Carbohydr Res 125:155-60.
28 Pérez S, Vergelati C. (1987) Polymer Bull 17:141-48.
29 Nishida Y, Ohrui H, Meguro H (1984) Tetrahedron Lett 25:1575-78.
30 Tvaroska I, Pérez S (1986) Carbohydr Res 149:389-410.
31 Allinger NL (1977) J Amer Chem Soc 99:8127-34.
32 Jeffrey GA, Taylor R (1980) J Comp Chem 1:99-109.
33 Tran V, Buléon A, Imberty A, Pérez S (1989) Biopolymers 28:579-690.
34 French AD (1989) Carbohydr Res 188:206-11.

Supplementary Material

GlcNAcβ (1-4)GlcNAc

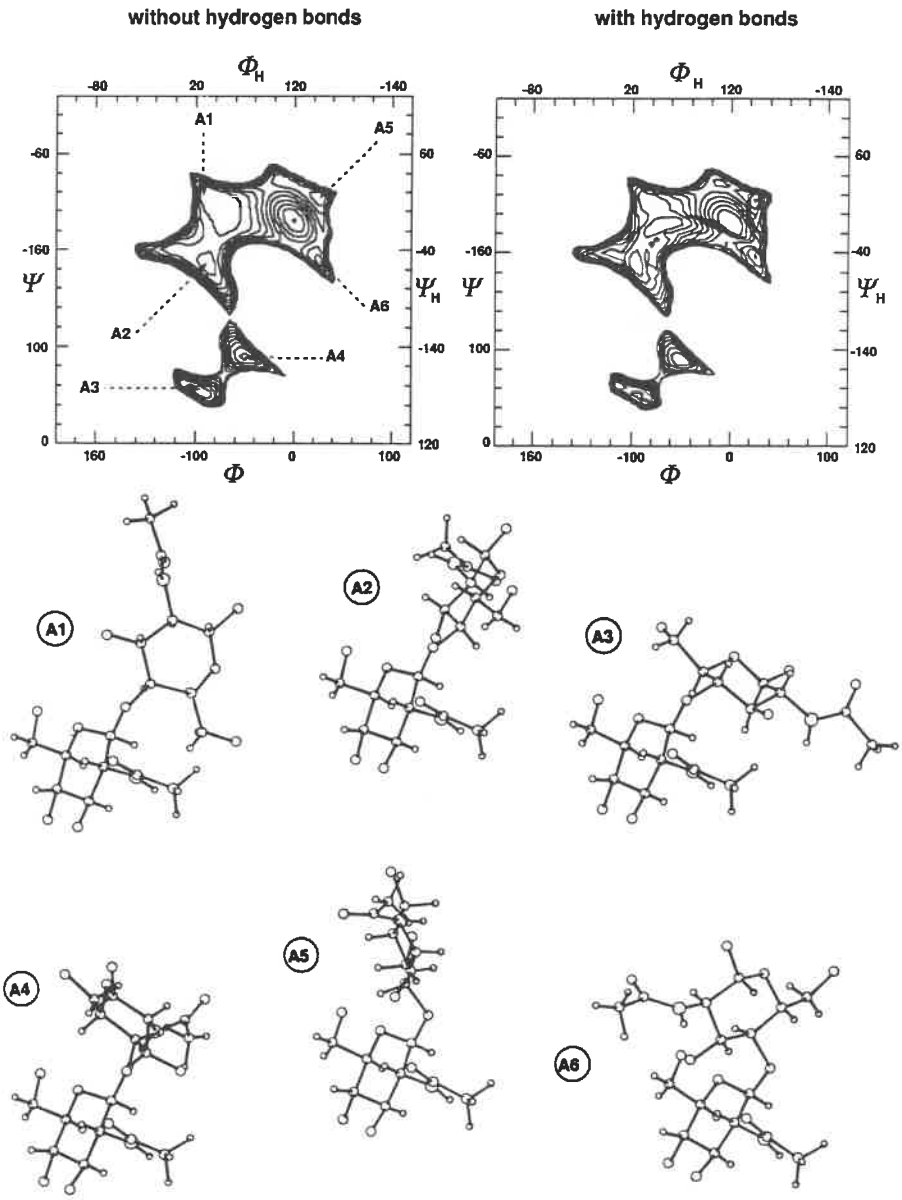

Figure 1s. Iso-energy map of the GlcNAcβ(1-4)GlcNAc disaccharide computed without (left panel) and with (right panel) the contribution arising from hydrogen bonding. The two hydroxymethyl groups are fixed in a GT orientation. The axes at the top and bottom are differently named since they correspond respectively to the crystallography and NMR nomenclature for torsional angles (see the Methods section for definitions). The same conventions are used for the left and right axes. Iso-energy contours are drawn by extrapolation of 1 kcal/mol with respect to the absolute minimum (A1). All the minima located on the iso-energy map (A1 to A6) are represented by ball and stick drawings (ALCHEMY software, Tripos).

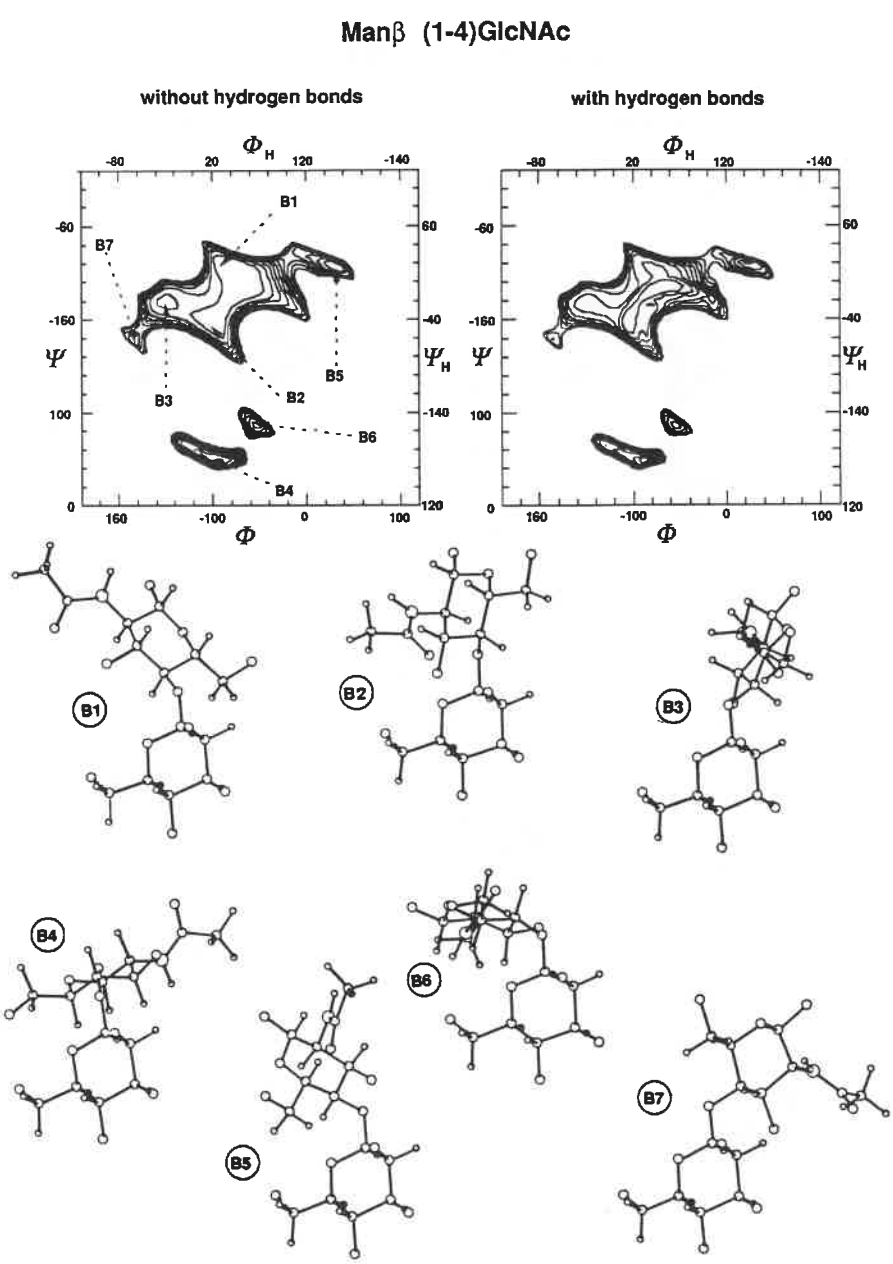

Figure 2s. Iso-energy map of the Manβ(1-4)GlcNAc disaccharide computed without (left panel) and with (right panel) the contribution arising from hydrogen bonding. The legend is the same as in Fig. 1s. The drawings represent the energy minima B1 to B7.

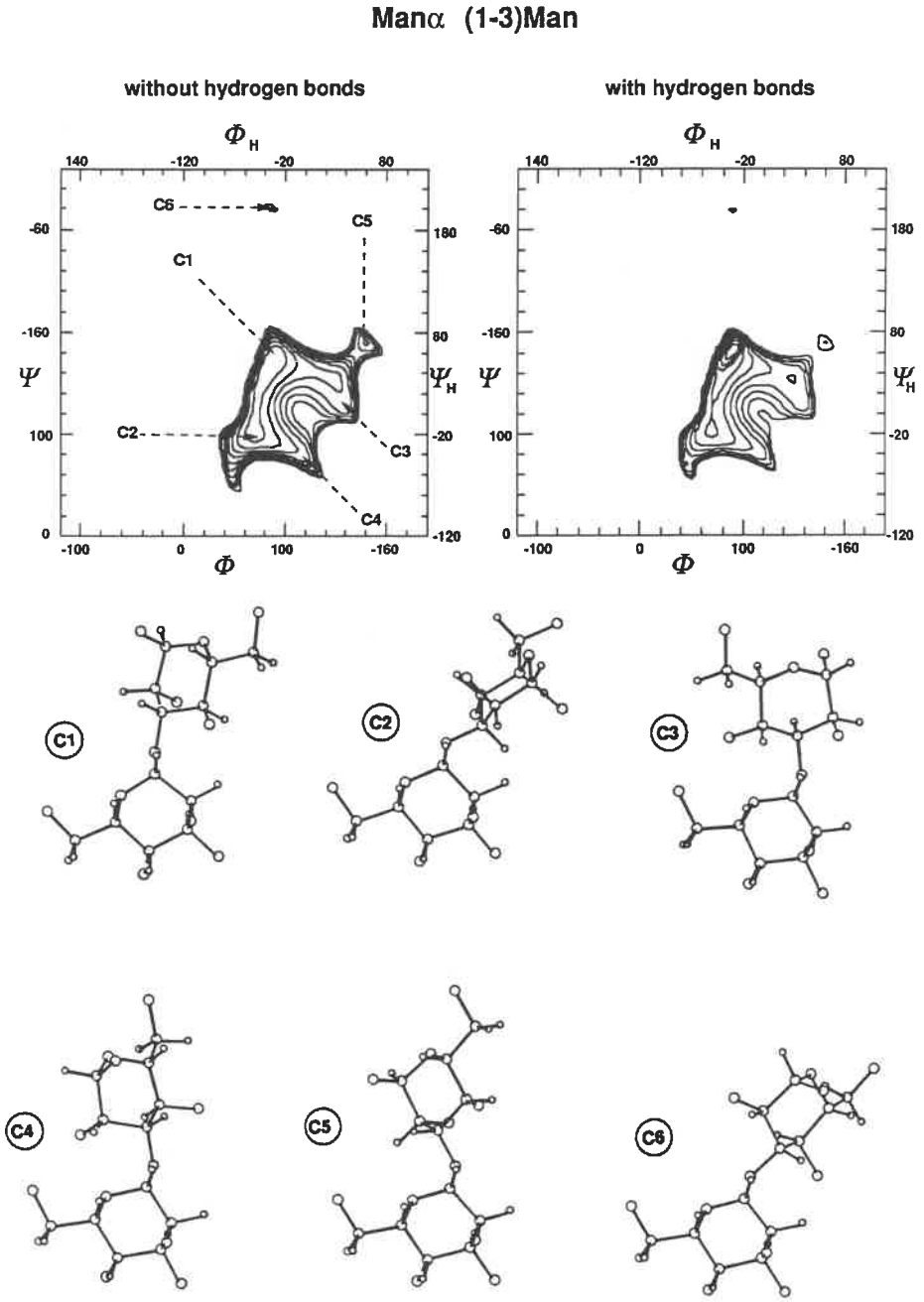

Figure 3s. Iso-energy map of the Manα(1-3)Man disaccharide computed without (left panel) and with (right panel) the contribution arising from hydrogen bonding. The legend is the same as in Fig. 1s. The drawings represent the energy minima C1 to C6.

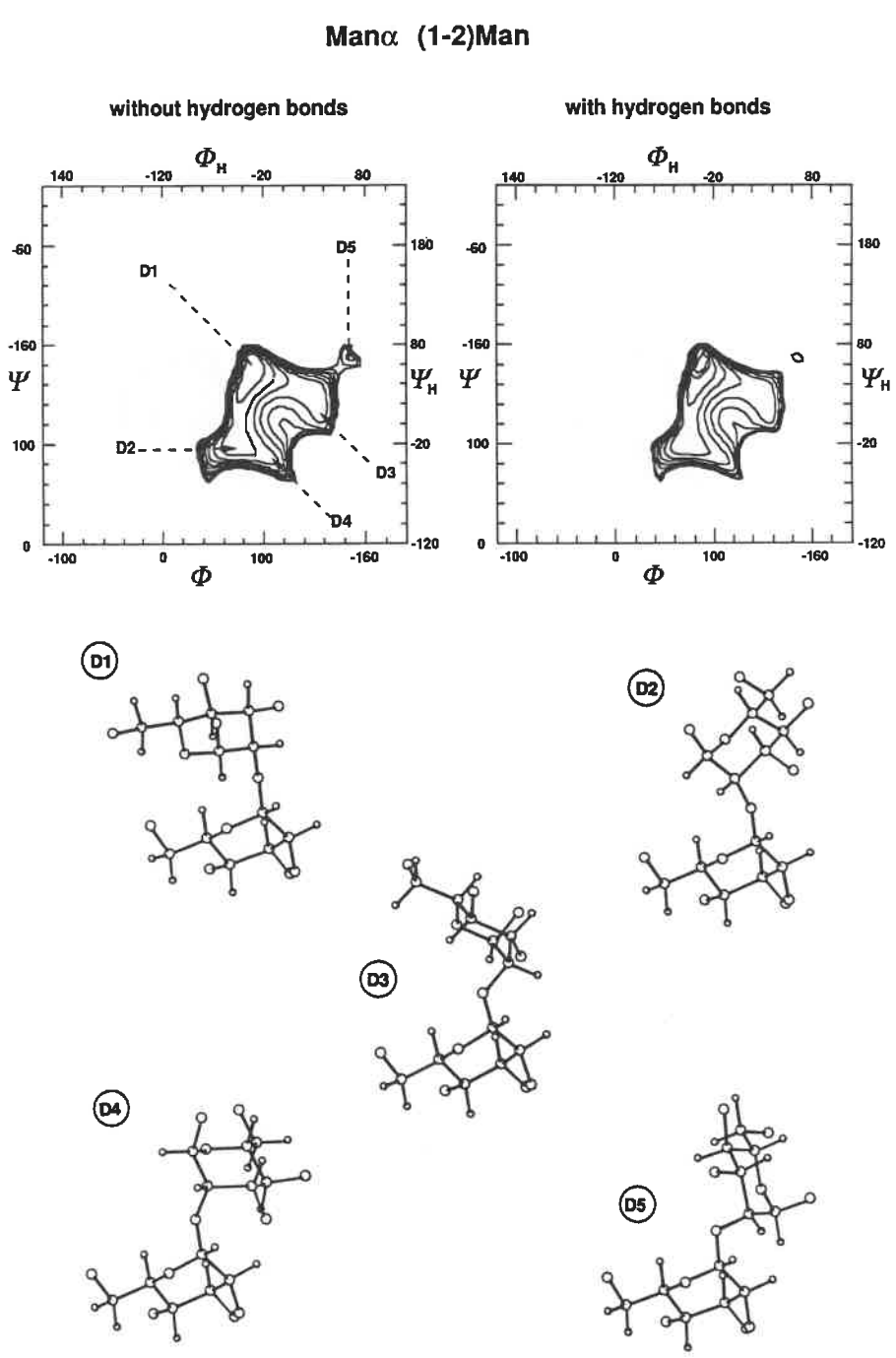

Figure 4s. Iso-energy map of the Manα(1-2)Man disaccharide computed without (left panel) and with (right panel) the contribution arising from hydrogen bonding. The legend is the same as in Fig. 1s. The drawings represent the energy minima D1 to D5.

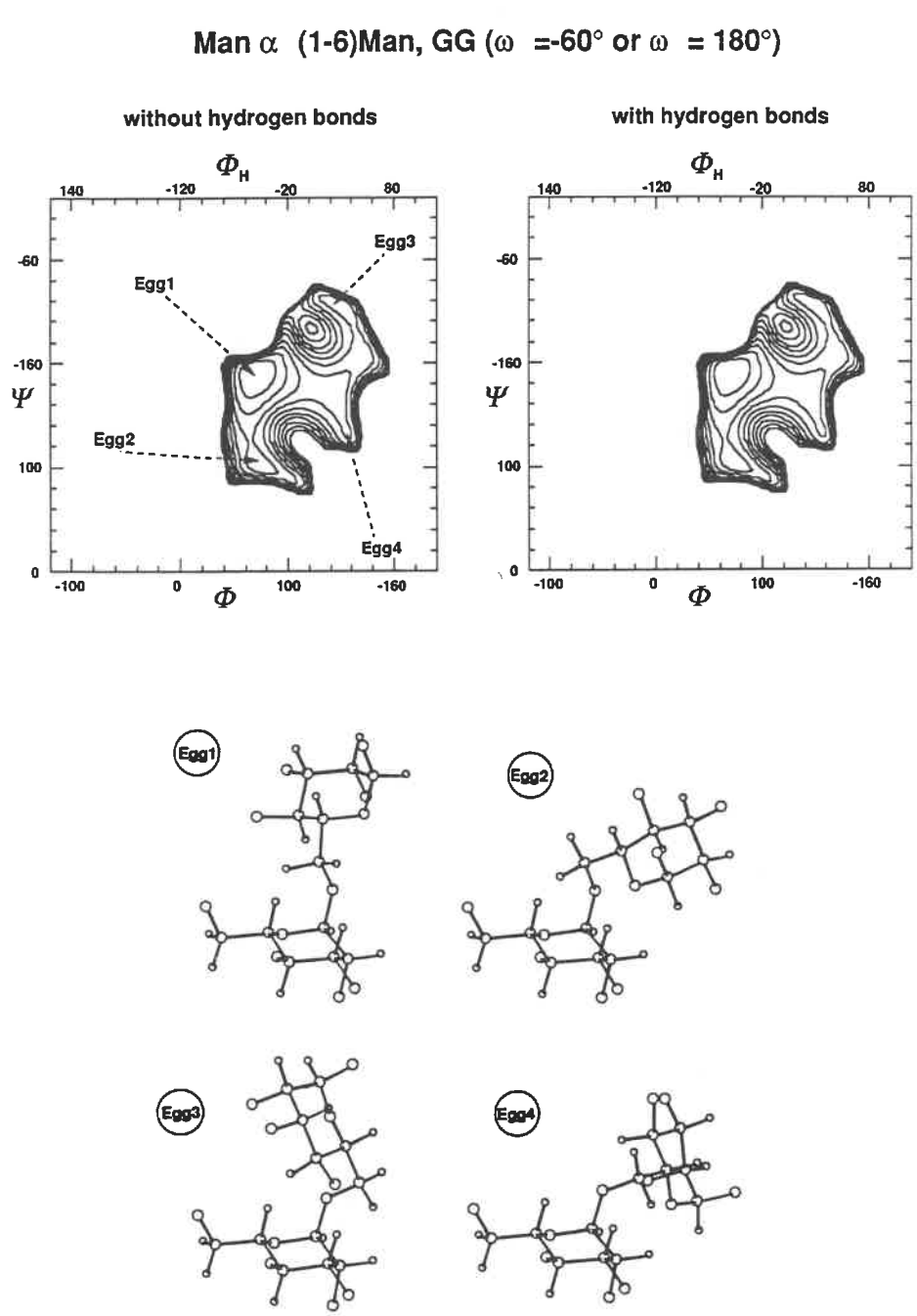

Figure 5s. Iso-energy map of the Manα(1-6)Man disaccharide computed without (left panel) and with (right panel) the contribution arising from hydrogen bonding. The ω angle at the glycosidic linkage is fixed in a GG orientation. The legend is the same as in Fig. 1s. The drawings represent the energy minima Egg1 to Egg4.

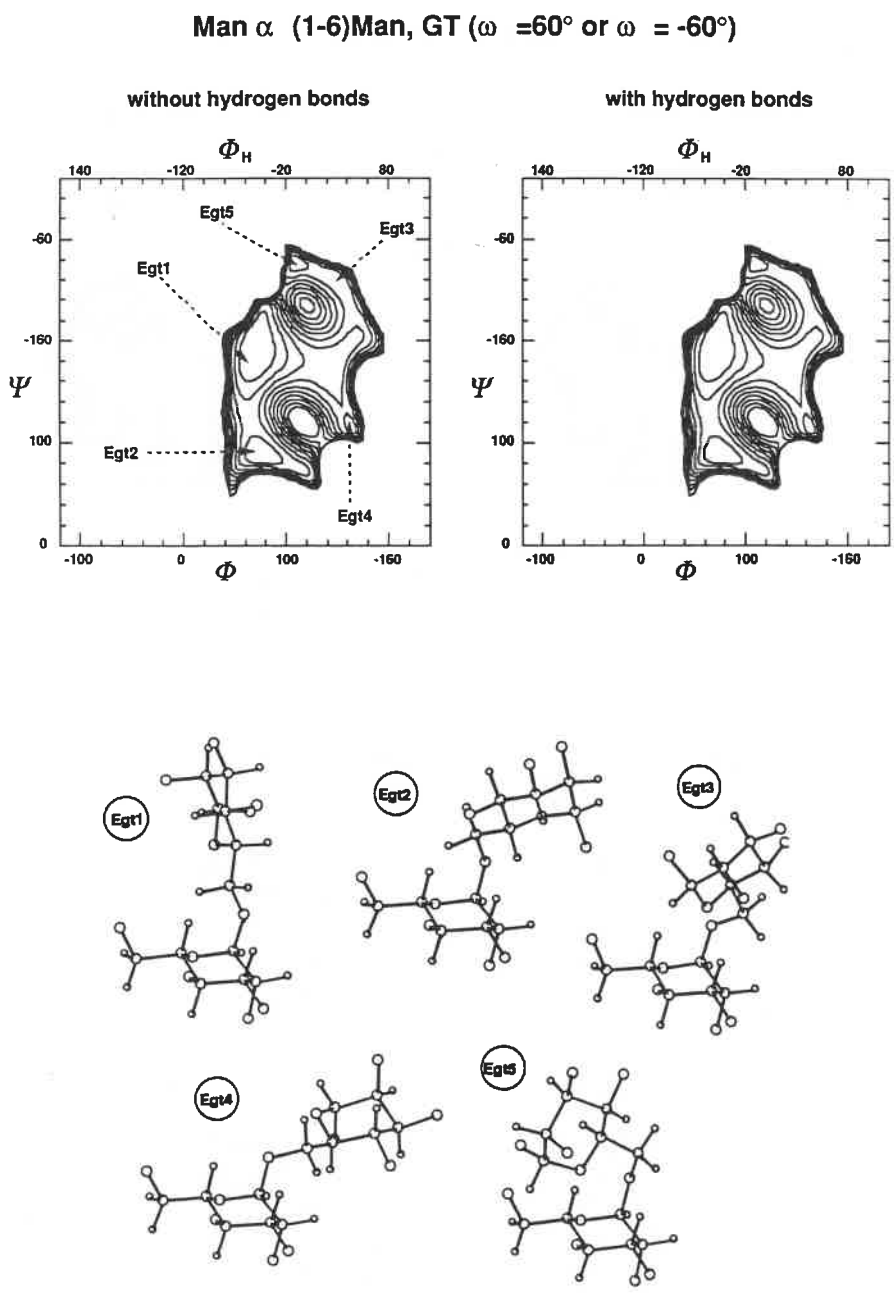

Figure 6s. Iso-energy map of the Manα(1-6)Man disaccharide computed without (left panel) and with (right panel) the contribution arising from hydrogen bonding. The ω angle at the glycosidic linkage is fixed in a GT orientation. The legend is the same as in Fig. 1s. The drawings represent the energy minima Egt1 to Egt4.