Glycoconjugate Journal (1995) **12**: 331–349

Computer simulation of histo-blood group oligosaccharides: energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens

ANNE IMBERTY[1*], EMMANUEL MIKROS[2‡], JAROSLAV KOČA[3], ROSELLA MOLLICONE[4], RAFAEL ORIOL[4] and SERGE PÉREZ[2]

[1] *Laboratoire de Synthèse Organique-CNRS, Faculté des Sciences et Techniques, 2 rue de la Houssinière, 44072 Nantes cedex 03, France*

[2] *Ingéniérie Moléculaire, INRA, BP 1627, 44316 Nantes cedex 03, France*

[3] *Department of Organic Chemistry, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republik*

[4] *INSERM U178, Hôpital Paul-Brousse, 94807 Villejuif, France*

Received 15 December 1994, Revised 26 January 1995

The three-dimensional structures of fourteen histo-blood groups carbohydrate antigens have been established through a combination of molecular mechanics and conformational searching methods. The conformational space available for each disaccharide, constituents of these determinants, has been throroughly characterized. The results have been organized in a data bank fashion. Larger relatives, i.e. 14 tri- and tetrasaccharides of histo-blood group antigens, have been modelled using a different method for exploring the complex potential energy surface. This approach is aimed at establishing all the possible families of conformations, along with the conformational pathways. Different conformational behaviours are exhibited by these oligosaccharides. Some of them, i.e. Le^x and Le^y tri and tetrasaccharides, are very rigid; 99% of their populations belong to the same conformational family. Others, like H type 1, H type 2 or H type 6 oligosaccharides, are essentially rigid, but a secondary conformational family, corresponding to 3–4% of the total population, can arise. Finally, the H types 3 and 4 trisaccharides, and the A type 1 and A type 2 tetrasaccharides are predicted to behave rather flexibly. The information gathered in the present investigation has been used to analyse the body of experimental evidence, either physical or biological, available for this series of carbohydrate antigens. Of special interest are the several different alignments that can be proposed for these molecules. They yield a realistic definition of the three-dimensional features of the epitopes thereby providing essential information about how carbohydrate antigens are recognized by proteins.

Keywords: Blood group, molecular modelling, oligosaccharide conformation

Introduction

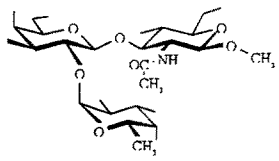
Blood group antigens on human erythrocytes are either carbohydrate-dependent (ABH, Lewis, Ii, P1, P-related, T and Tn) or protein-dependent (MNSs, Gerbich, Rh, Kell, Duffy and Cromer related). The carbohydrate determinants are covalently linked to proteins or lipids. The distribution and structure of these molecules have been recently reviewed [1] together with their possible biological functions.

The ABH(O) and Lewis blood group carbohydrate-dependent antigens have been the most widely studied in the past 30 years and are well characterized. They are now currently referred to as histo-blood groups antigens (see [2] for review) as they are expressed on the epithelia of glandular tissues, primary sensory neurones and exocrine secretions in man and other mammals (see reviews [3, 4]). In fact, only man and some anthropoids apes have ABH on erythrocytes whereas most old world monkeys have ABH only on the vascular endothelium and exocrine secretions [5]. The physiological role of the antigens on the erythrocyte cell surface remains obscure, but sialylated carbohydrates of the Le^x family (3′-sialyl-Le^x/Le^a) (review [6]) and later sulfated analogues [7] have been recognized as important ligands in cellular adhesion. At the present time, the clinical importance of histo-

* To whom all correspondence should be addressed.

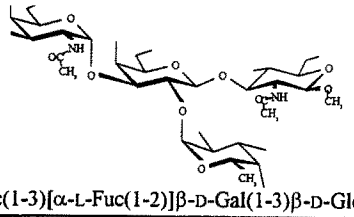
‡ Permanent address: Department of Pharmacy, University of Athens, Panespistimiopoli, Zografou, GR-15771 Athens, Greece.

H type 1



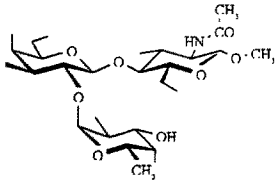
α-L-Fuc(1-2)β-D-Gal(1-3)β-D-GlcNAc

A type 1



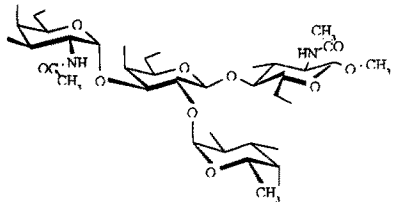
α-D-GalNAc(1-3)[α-L-Fuc(1-2)]β-D-Gal(1-3)β-D-GlcNAc

H type 2



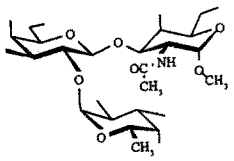
α-L-Fuc(1-2)β-D-Gal(1-4)β-D-GlcNAc

A type 2



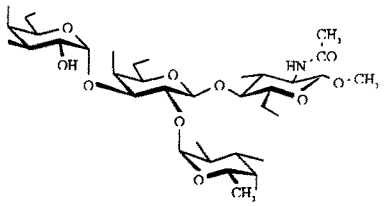
α-D-GalNAc(1-3)[α-L-Fuc(1-2)]β-D-Gal(1-4)β-D-GlcNAc

H type 3



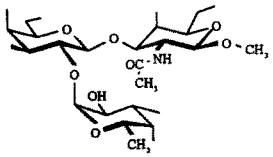
α-L-Fuc(1-2)β-D-Gal(1-3)α-D-GalNAc

B type 2



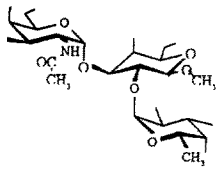
α-D-Gal(1-3)[α-L-Fuc(1-2)]β-D-Gal(1-4)β-D-GlcNAc

H type 4



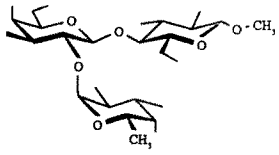
α-L-Fuc(1-2)β-D-Gal(1-3)β-D-GalNAc

A trisaccharide



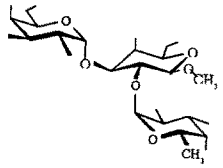
α-D-GalNAc(1-3)[α-L-Fuc(1-2)]β-D-Gal

H type 6



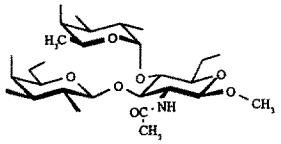
α-L-Fuc(1-2)β-D-Gal(1-4)β-D-Glc

B trisaccharide



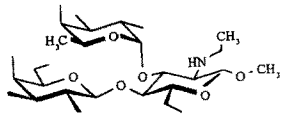
α-D-Gal(1-3)[α-L-Fuc(1-2)]β-D-Gal

Le^a



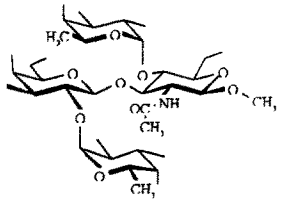
β-D-Gal(1-3)[α-L-Fuc(1-4)]β-D-GlcNAc

Le^X



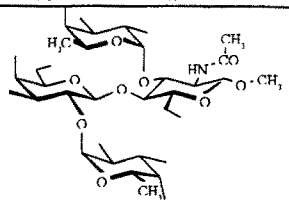
β-D-Gal(1-4)[α-L-Fuc(1-3)]β-D-GlcNAc

Le^b



α-L-Fuc(1-2)β-D-Gal(1-3)[α-L-Fuc(1-4)]β-D-GlcNAc

Le^Y



α-L-Fuc(1-2)β-D-Gal(1-4)[α-L-Fuc(1-3)]β-D-GlcNAc

Figure 1. Schematic representations of the 14 histo–blood group oligosaccharides investigated.

blood group systems depends on the specificity and frequency of the related antibodies. The histo-blood group antibodies are involved in blood group incompatibility and can also be used as indicators of tissue differentiation and of malignancy [8].

Conformational studies of histo-blood group oligosaccharides first appeared at the beginning of the eighties [9, 10]. These studies were feasible due to both an important effort in the chemistry of synthetic oligosaccharides and the availability of conformational analysis methods in the carbohydrate field. In these early studies, the carbohydrate rings were considered as rigid, and only the absolute minimum was searched as a function of rotation about the glycosidic linkage. The main conclusions of this research underlined the similarities in the shapes of the βGal1-3βGlcNAc and βGal1-4βGlcNAc disaccharides; it was also suggested that the addition of one or two fucose residues on these disaccharides would not alter their preferred conformations (see review [11]). Most recently quantitative nOe measurements have been performed on a series of blood group determinants to complement the preliminary NMR data of Lemieux and coworkers [9]. From NMR data and calculation of molecular dynamics, it was concluded that these blood group oligosaccharides exhibit much less internal motion than do the disaccharides from which they are composed [12, 13]. Due to the biological importance of Le^x and Le^a derivatives, much effort has been made in recent years to establish their conformation by NMR and modelling [14–16]. In general, the Lewis trisaccharide core has been considered as having a well established conformation with rigid behaviour whereas the flexibility of the sialic moiety is still debated.

Even if blood group oligosaccharides are generally con sidered as rigid, secondary conformational families can occur in solution. Also it has been suggested that some of them, such as the H type 4 trisaccharide which contains the βGal1-3βGalNAc as the core disaccharide, can be more flexible [17]. Recent studies on protein/carbohydrate complexes have shown that a receptor can select or induce a ligand confor mation that is not the most frequent in solution [18]. This has been demonstrated by protein crystallography on lectin/oligosaccharide complexes [19] by transferred nOe experiments for the binding of sialyl-Lex by E-selectin [20] and by both methods for the binding of bacterial polysaccharide by antibodies [21]. It appears important to determine not only the conformation corresponding to the global energy minimum of each oligosaccharide but also all the other possible conformations and to do this in a systematic way. Therefore, we thought it was necessary to reinvestigate the conformational behaviour of ABH and Lewis oligosaccharides.

The conformational search of the 14 histo-blood group tri- and tetrasaccharides displayed in Fig. 1 has been performed in the present study. First, the relaxed potential energy surfaces of the 10 constituting disaccharide fragments were calculated with the aid of the molecular mechanics program MM3 [22, 23]. The MM3 program has been widely used in carbohydrate modelling and has been shown to give a correct description of ring geometries, anomeric equilibria and linkage torsion angles with suitable accuracy [24, 25]. Also the program allows the molecule to be optimized at each point of a conformational search, giving a more realistic approach than the 'rigid' one used in early studies.

The systematic search about the glycosidic torsion angles which is classically performed for disaccharide modelling, becomes too CPU-consuming for larger structures. Molecular modelling of oligosaccharide was therefore performed with a new method CICADA (Channels In Conformational space Analyzed by Driver Approach) [26] interfaced with the MM3 program. The CICADA approach explores the potential energy surface in an intelligent and less CPU-consuming way, therefore allowing for larger molecules to be modelled. This method has recently been proven to be of prime interest in the modelling of carbohydrates [27]: it allows for accurate determination not only of the global energy minimum and all secondary minima, but also of the low-energy conversion pathways. The results from CICADA also permit prediction of some flexibility indexes for molecules and for individual group rotations.

Methods

All the disaccharides and oligosaccharides were built using MONOBANK, a data base of three-dimensional structures for monosaccharides [28].

Nomenclature

The two torsion angles describing a glycosidic linkage are defined as

$$\Phi = \Theta(\text{O-5–C-1–O-1–C′-X}) \text{ and}$$
$$\Psi = \Theta(\text{C-1–O-1–C′-X–C′-X+1})$$

with the primed atoms belonging to the reducing side and the sign being in agreement with IUPAC nomenclature [29].

Relaxed (F,Y) maps of disaccharides

The ten disaccharides that constitute the oligosaccharides displayed in Fig. 1 have been studied by a systematic grid search method. The calculations were always performed on the disaccharides with a methyl group at the reducing end. Starting from a refined geometry, the procedure drives Φ and Ψ torsion angles in steps of 20° over the whole angular range while the MM3 program provides full geometry relaxation. The block-diagonal minimization method, with the default energy-convergence criterion (0.00008*n kcal mol^{-1} per five iteration, n = number of atom) was used for grid-point optimizations and the dielectric constant was given a value of 4. Several maps are calculated for each disaccharide in order to take into account the several possible orientations of the primary and secondary hydroxyl groups. At most, 36 starting geometries are needed to take into account the three staggered orientations of each monosaccharide hydroxymethyl group, referred as GG, GT and TG [30] and the two possible networks of secondary

hydroxyl groups, around each ring, referred as clockwise and counterclockwise. For each disaccharide, the results of these many calculations are summarized in a so-called 'adiabatic' map where only the conformer with the lowest energy for each (Φ,Ψ) value is considered. Isoenergy contours are then plotted by interpolation of 1 kcal mol^{-1} within a limit of 8 kcal mol^{-1}.

Calculation of potential energy surfaces using CICADA

The CICADA program, which is an interface to the MM3 force field, was used for the conformational analysis of tri- and tetrasaccharides. All the oligosaccharides were considered in the methyl glycoside form. Input of the CICADA program consists mainly of one or a few conformers in MM3 format and a file containing the list of torsion angles to be driven and/or monitored. During the CICADA calculations, each individual torsion angle is driven in each direction from the initial conformation at a given increment. At each step, the structure is optimized except for the driven torsion angle. When CICADA detects a minimum, the conformation is fully optimized, including the driven torsion angle. The resulting structure is compared with the previously stored ones and stored as a new geometry if it is one that has not yet been discovered. Structures corresponding to energy maxima, the 'transition states' are also saved. Calculations stop when no new conformers (local minima) are found within a desired energy window. The same criteria were used for the minimizations as for the grid search, except that the dielectric constant was set to 80, in order to lower the influence of hydrogen bonding on the potential energy surface.

For each oligosaccharide, between one and three conformers were built and fully refined to serve as starting points for the CICADA runs. The step of increment of the driven torsion angles was set at 20°. The driven torsion angles were Φ and Ψ at each linkage, the torsion angle of each hydroxymethyl group (O-5–C-5–C-6–O-6) and the first torsion angle of the *N*-acetyl groups (C-1–C-2–N–C-8). The number of driven torsion angles, and therefore the dimensionality of the potential energy surface to explore, ranged from seven or eight for the trisaccharides to 10 or 11 for the tetrasaccharides. The torsion angles of all the secondary hydroxyl groups, as well as the first torsion angle of the *O*-methyl group of the reducing end, were not driven but were monitored. Two conformers were considered different if at least one of their dihedral angles, either driven or monitored, differed by more than 30°. Two cutoffs for relative energy were applied, one for exploring the PES (50 kcal mol^{-1}), and one for considering a conformer to be a new starting geometry. This last value defines the length of the calculations, so it was set to 6 kcal mol^{-1} for the most rigid oligosaccharides but to only 3 or 4 for the flexible ones.

The CICADA runs required 5000 to 10 000 fully relaxed energy minimizations before the PES was thought to be correctly explored, which corresponds to 27 500 to 55 000 MM computations. There is no absolute way to determine if the sampling is correct, i.e. if all the conformational families have been explored. As a check, we used the disaccharides energy maps calculated with the grid search method to verify if all the conformations of each glycosidic linkage (and when possible all combinations) have been explored during the sampling.

Analysis of the potential energy surface

A second program, PANIC, was then used to analyse the potential energy surface. It lists the minima and transition states, and calculates some flexibility indexes [31]. Another homemade program, FAMILY [32], analysed the clustering of conformers, determining family of conformations in a given energy window. Requirements for the program are the list of torsion angles of interest and the separation value which will determine a different family. The algorithm concludes that a conformer belongs to a conformational family if at least one of its torsion angles differs by less than the separation value to at least one of the conformers of the family. In the present study, the families are defined only by the variations in the glycosidic torsion angles, and a separation value of 15° was used. Families are characterized by their conformation of lowest energy and also by their possible variations in oligosaccharide shapes within a given energy window. As for the relative importance of each family, their population is calculated using a Boltzmann type of distribution.

Alignment of conformers

The alignment of conformers was performed with the FIT procedure of the SYBYL molecular modelling software [33]. Only the ring atoms and glycosidic oxygens of the H trisaccharide cores (aFuc1-2bGal1-3bGlcNAc or aFuc1-2bGal1-4bGlcNAc or aFuc1-2bGal1-3bGalNAc) were used. In this procedure, the atoms belonging to the galactose and fucose were fitted identically (O-3 on O-3...) whereas for GlcNAc and GalNAc the inversion of the ring orientation was considered depending on the nature (1-3 or 1-4) of the linkage.

Miscellaneous

The calculations were performed on UNIX computers (Silicon Graphics R3000, R4000 and Power Challenge, IBM Risc 6000, Sun). All molecular drawings were done with SYBYL [33].

Results

Data bank of disaccharides energy maps

Ten disaccharides are needed to build the oligosaccharides of Fig. 1. For the H type trisaccharides, the constituting-disaccharides are bGal1-3bGlcNAc (or Lec), bGal1-4bGlcNAc (or lacNAc), bGal1-3aGalNAc, bGal1-3bGalNAc and bGal1-4bGlc for the core moiety and aFuc1-2bGal which is the H disaccharide. The so-called Lewis disaccharide aFuc1-4bGlcNAc is present in Le^a and Le^b whereas the disaccharide aFuc1-3bGlcNAc is present in LeX and LeY. Finally, the two disaccharides aGalNAc1-3bGal and aGal1-3bGal are the terminal determinants of blood group A and B, respectively. The conformational analysis of all these disaccharides has been

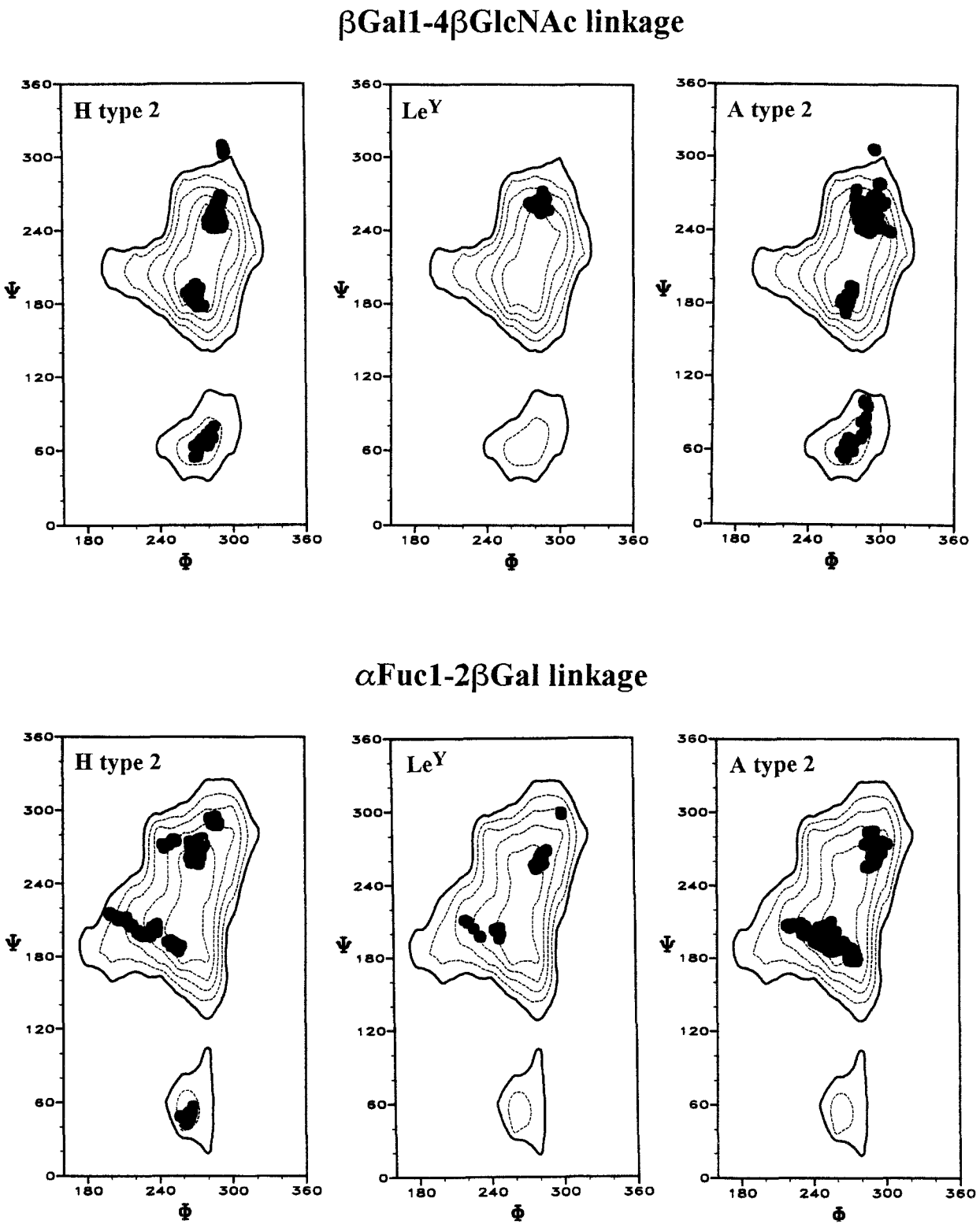


Figure 2. Projections on two (Φ,Ψ) maps of the CICADA conformational search of three oligosaccharides. The dots indicate the Φ,Ψ values of all the optimized conformations determined by CICADA for each oligosaccharide in a 5 kcal mol^{-1} energy window. For comparison, the isocontours, drawn in 1 kcal mol^{-1} steps with an outer limit of 5 kcal mol^{-1}, represent the energy levels of each disaccharide as calculated with the relaxed grid search approach (see Figs S1 and S4). The secondary low energy region of βGal(1-4)βGlcNAc, (conformation D in Fig. S1) has not been included since it does not contain any oligosaccharide structures with a relative energy of less than 5 kcal mol^{-1}.

completed by a systematic grid search approach. The resulting adiabatic maps, along with the drawings of the conformations of interest are given in Figs S1–S8 (see supplementary Figures). Only eight maps have been represented since the disaccharide bGal1-3bGalNAc displays the same energy surface as bGal1-3aGalNAc, and the lactose disaccharide (bGal1-4bGlc) has the same conformational behaviour as LacNAc (bGal1-4bGlcNAc).

The analysis of the energy maps demonstrates that these disaccharides can access a large number of conformational states and have therefore high potential flexibility. The lowest energy conformations of each region are represented in Figs S1–S8. Hydrogen bonding stabilizes some of these conformations, but this is not a general feature. The linkages with the α orientation are quite restricted for the rotations about the Φ torsion angle with values centred at about –80° for the L configuration and about 80° for the D configuration as dictated by the exo-anomeric effect [34]. Much more flexibility is exhibited by the Ψ torsion angles which can adopt two or three different low energy values and which can, in general, span the entire angular range within an energy barrier of 8 kcal mol^{-1}. The energy maps of the β linkages show almost the same flexibility for the Ψ torsion angle. For the Φ torsion angle, in addition to the main low energy value of around –80° that corresponds to the exo-anomeric effect, a secondary minimum is predicted for Φ of about 100°. Even if this secondary minimum is 3–5 kcal mol^{-1} above the global one, it should not be *a priori* discarded. In the crystalline complex between a plant lectin and an oligosaccharide, it has been observed that the βGlcNAc1-2αMan moiety has such a conformation [19, 32].

Potential energy surfaces of histo-blood group oligosaccharides

The conformational behaviour of the 14 oligosaccharides of Fig. 1 were fully investigated. Since the conformational searches were performed in conformational spaces having from seven to eleven dimensions, it is not straightforward to describe the results in a simple representation. One possibility would be to project the results onto each two-dimensional (F, Y) map. This will allow visualization of the reduction in flexibility of each linkage when comparing the disaccharide alone or as part of an oligosaccharide. Figure 2 displays such an analysis for two linkages, bGal1-4bGlcNAc and aFuc1-2bGal, when they are part of the H type 2 trisaccharide and also of the Leb and A type 2 tetrasaccharides. For each oligosaccharide, all the CICADA conformations with relative energy of less than 5 kcal mol–1 have been projected onto the corresponding disaccharide energy maps. (741 conformers for an H type 2, 271 for LeY and 905 for A type 2). The same linkage has a quite different flexibility in different oligosaccharides. The lactosamine linkage can belong to the three low energy regions A, B and C (see Fig. S1) of the map when it is part of an H type 2 trisaccharide and an A type 2 tetrasaccharide. In constrast, it is confined to the A conformation when it is part of the more rigid Le^y tetrasaccharide. Other conformations are explored by the sampling, but their energy is significantly higher because of steric conflicts involving the two other residues of the tetrasaccharide.

For H type 2 and A type 2 oligosaccharides, some conformations (with a relative energy between 4 and 5 kcal mol^{-1}) can occur outside the 5 kcal mol^{-1} limit of the disaccharide. This originates from the 'stacking-type' of stabilizing interaction occurring between the apolar face of the fucose at position 2 of galactose (created by methine hydrogens H-3, H-4 and H-5) and the apolar face of *N*-acetyl-glucosamine (created by the hydrogens H-1, H-3 and H-4). Such a long range interaction could not be predicted from the lactosamine energy map alone and results from the 1–2 type of glycosidic linkage.

Another representation of the results requires a prior characterization of the conformational families resulting from the potential energy surfaces. A clustering method, or family analysis [32] has been applied to conformational data of the 14 oligosaccharides.

For the H trisaccharides, the characteristics of the conformational families having a population of more than 1% at room temperature are listed in Table 1. Two very different conformational behaviours are present among the five trisaccharides studied here. For three of them (H type 1, H type 2 and H type 6), one family of conformation overwhelms the others, representing more than 90% of the whole population. In all three cases, a second family of conformations appears but it represents only 3.5–5% of the population. The occurrence of this less populated family does not depend on the conformational change of one glycosidic linkage, but on the concerted conformational changes of both. Although the Φ torsion angles always display the same value centred at about –80° for Fam. I and Fam. II, both Ψ torsion angles rotate when changing of conformational family. For these three trisaccharides, Fam. I contains the B conformation of the αFuc1-2βGal disaccharide (Ψ close to –95°) and the A conformation of the core disaccharides (see Figs S1–S5); whereas Fam. II contains the A conformation of the αFuc1-2βGal disaccharide (Ψ close to –165°) and the B conformation of the core disaccharides. Several other families of conformations can be described for the oligosaccharides but only two of them (Fam. III for H type 1 and H type 6) have a population of more than 1%.

Two trisaccharides (H type 3 and H type 4) behave in quite a different manner. Their populations divide into two conformational families of almost equivalent importance. The local minima of each family have similar energies. In both cases, the absolute minimum does not belong to the most populated family (Fam. I) but to the second one (Fam. II). This can be correlated to the extended lowest energy region of the adiabatic energy map of the core disaccharide βGal1-3α/βGalNAc (see Fig. S3). The two lowest energy conformations A and B have almost no difference in energy; they belong to a very flat plateau which allows facile interconversions. For both disaccharides, Fam. I contains conformation B of the αFuc1-2βGal

Table 1. Characteristics of H oligosaccharide conformational families. Only the families having an energy-weighted population of more than 1% (at 296 K) have been listed. For each oligosaccharide and each family, E_{min} indicates the glycosidic linkages torsion angles and relative energy of the lowest energy conformation, whereas min and max are the limit values for each torsion angle within an energy window of 5 kcal mol^{-1}

		Fam. I						*Fam II*						*Fam. III*					
		Φ	Ψ	Φ	Ψ	E_{rel}	%	Φ	Ψ	Φ	Ψ	E_{rel}	%	Φ	Ψ	Φ	Ψ	E_{rel}	%
		Fuc12Gal		Gal13GlcNAc				Fuc12Gal		Gal13GlcNAc				Fuc12Gal		Gal13GlcNAc			
H	E_{min}	−81.1	−95.5	−71.8	138.8	0.00		−97.3	−168.1	−87.2	67.2	1.38		−109.1	−159.1	−66.8	157.9	2.62	
Type	min	−90	−109	−78	127		**94.5**	−104	−175	−90	57		**4.1**	−151	−173	−79	131		**1.1**
1	max	−71	−89	−62	171			−92	−167	−86	68			−100	−145	−55	191		
		Fuc12Gal		Gal14GlcNAc				Fuc12Gal		Gal14GlcNAc									
H	E_{min}	−79.0	−93.5	−76.9	−113.6	0.00		−96.0	−170.8	−87.5	−179.8	1.82							
Type	min	−86	−103	−79	−122		**94.6**	−104	−174	−91	175		**3.6**						
2	max	−73	−87	−67	−103			−92	−164	−86	−177								
		Fuc12Gal		Gal13αGalNAc				Fuc12Gal		Gαl13αGalNAc									
H	E_{min}	−79.8	−95.3	−72.2	129.6	0.11		−96.6	−168.1	−87.9	66.4	0.00							
Type	min	−82	−96	−78	118		**51.4**	−99	−178	−90	57		**46.7**						
3	max	−76	−93	−69	142			−92	−167	−84	68								
		Fuc12Gal		Gal13βGalNAc				Fuc12Gal		Gal13βGalNAc									
H	E_{min}	−79.2	−95.4	−71.5	127.1	0.05		−98.0	−164.6	−90.0	61.7	0.00							
Type	min	−90	−104	−79	74		**68.5**	−105	−177	−91	57		**29.7**						
4	max	−73	−84	−60	152			−93	−164	−84	70								
		Fuc12Gal		Gal14Glc				Fuc12Gal		Gal14Glc				Fuc12Gal		Gal14Glc			
H	E_{min}	−82.0	−97.3	−71.4	−113.0	0.00		−100.2	−165.9	−88.0	178.8	1.73		−64.5	−67.7	−91.3	−172.8	1.99	
Type	min	−87	−101	−80	−121		**91.3**	−105	−178	−90	176		**4.8**	−66	−72	−95	−178		**1.5**
6	max	−72	−83	−67	−103			−90	−160	−86	−178			−61	−58	−87	−171		

and conformation B of βGal1-3α/βGalNAc, whereas Fam. II contains conformation A of both linkages.

Figure 3 displays several conformations for each family: the lowest energy conformation and also the ones with the largest variations for the linkage torsion angles in an energy window of 5 kcal mol^{-1}. These variations are also listed in Table 1. For all trisaccharides, Fam. I and Fam. II do not exhibit large variations around the absolute minimum. In general, each torsion angle can have variation amplitudes of about 20°. The exception is the H type 4 trisaccharide where the Ψ torsion angle of the βGal1-3βGalNAc disaccharide can vary by almost 90° in the large conformational plateau (see Fig. S3).

Another striking feature appears when taking a closer look at Fig. 3. The shapes of all the oligosaccharides of Fam. I are almost identical, irrespective of the subtypes. This can be correlated to the previously reported close shapes of the global minima of the βGal1-4βGlcNAc, βGal1-3βGlcNAc and lactose disaccharides [9, 10] (A conformations of the present study). The novel feature is that the second low energy conformation of the βGal1-3α/βGalNAc disaccharide (B conformations) also displayed the same shape. Since this conformation generates Fam. I, all the most populated conformational families have the same shape. The only difference arises from the position of the *N*-acetyl group, which is either on one side or the other side of the neighbouring glycosidic linkage depending on the 1–3 or 1–4 nature of the latter, and the axial or equatorial position of the O-4 atom of the reducing ring depending on the *galacto* or *gluco* configuration. The second family of conformations (Fam. II), whether heavily populated or not, also displays the same overall shape for all the H oligosaccharides (but obviously different from Fam. I). The slight differences between oligosaccharides noted in Fam. I also apply to this family.

The same family analysis has been applied to the Lewis oligosaccharides (Table 2). The conformational behaviours of these tri- and tetra-saccharides appear to be more rigid than those of the H oligosaccharides. Le^a and Le^b oligosaccharides have more than 95% of their conformational population in one family, whereas this overpasses 99% in the case of Le^X and Le^Y oligosaccharides. Nevertheless, the existence of a second conformational family, even if not energetically favoured, cannot be discarded, specially in the case of the Le^a trisaccharide (see Fig. 4). When this conformational family occurs (see Table 2), it is correlated to a conformational change in the αFuc1-4βGlcNAc linkage which switches from to conformation A (with Ψ value close to −170°).

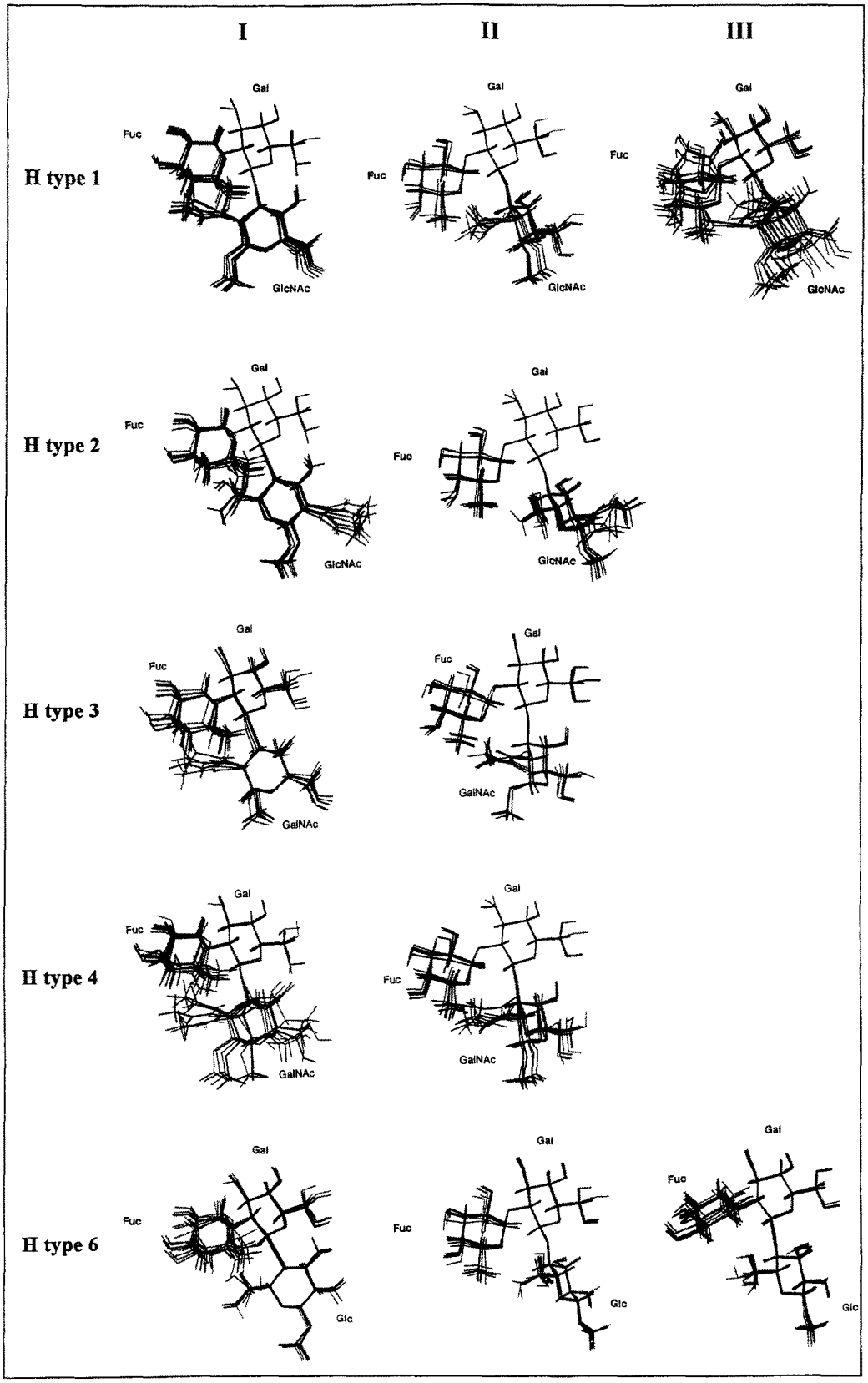


Figure 3. Graphic representation of all the conformational families of H type trisaccharides having a population of more than 1%. For each family, the lowest energy conformation has been drawn in black. The conformations with the largest difference in Φ and Ψ torsion angles within an energy window of 5 kcal mol^{-1} are drawn in grey.

Table 2. Characteristics of Lewis type oligosaccharide conformational families. Only the families having a energy-weighted population of more than 1% (at 296 K) have been listed. For each oligosaccharide and each family, E_{min} indicates the glycosidic linkages torsion angles and relative energy of the lowest energy conformation, whereas min and max are the limit values for each torsion angle within an energy window of 5 kcal mol^{-1}.

		Fam. I								*Fam. II*							
		Φ	Ψ	Φ	Ψ	Φ	Ψ	E_{rel}	%	Φ	Ψ	Φ	Ψ	Φ	Ψ	E_{rel}	%
				Gal14GlcNAc		Fuc13GlcNAc											
Le^X	E_{min}			−74.9	−103.6	−81.3	150.7	0.00									
	max			−81	−107	−86	146		**99.5**								
	min			−71	−99	−75	155										
		Fuc12Gal		Gal14GlcNAc		Fuc13GlcNAc											
Le^Y	E_{min}	−78.3	−99.0	−73.0	−102.2	−79.4	146.6	0.00									
	min	−85	−105	−81	−106	−88	145		**99.2**								
	max	−77	−91	−65	−97	−66	159										
				Gal13GlcNAc		Fuc14GlcNAc						Gal13GlcNAc		Fuc14GlcNAc			
Le^a	E_{min}			−75.4	144.1	−79.4	−97.7	0.00				−63.4	178.1	−96.5	−170.7	1.8	
	min			−78	140	−80	−99		**95.8**			−65	177	−103	−171		**3.8**
	max			−75	146	−76	−96					−62	−178	−96	−161		
		Fuc12Gal		Gal13GlcNAc		Fuc14GlcNAc				Fuc12Gal		Gal13GlcNAc		Fuc14GlcNAc			
Le^b	E_{min}	−79.0	−97.6	−74.0	145.2	−78.3	−97.0	0.00		−86.7	−95.6	−63.7	178.1	−96.2	−170.9	2.57	
	min	−88	−106	−88	139	−87	−100		**97.9**	−88	−98	−88	151	−138	−172		**1.4**
	max	−72	−85	−70	160	−76	−95			−76	−84	−63	180	−92	−154		

Comparison of Fam I in Figs. 3 and 4 shows the similarities of shape in the common part of Le^Y and its precursor H type 2, and between Le^b and its precursor H type 1. It appears therefore, as already mentioned [9, 10], that the addition of a fucose moiety on the GlcNAc does not alter the preferred conformations of these trisaccharides.

When applied to the A and B blood group determinants, the family analysis demonstrates a much higher flexibility. As displayed in Table 3 and in Fig. 5, tetrasaccharide determinants of blood group A type 1, A type 2 and B type 2 can occupy three families of conformations, two of them being highly populated. The most populated family of conformations (Fam. I) closely corresponds to the most populated conformation of the corresponding trisaccharide precursor (H type 1 or H type 2) of Table 1. However, this family contains only between 55 and 73% of the tetrasaccharides conformers. The second family (Fam. II) contains a quite important population (between 21 and 39%); it corresponds to the second conformational family of the H trisaccharides. As in the case of the H oligosaccharides, the differences between the two families consist of large modifications in both the αFuc1-2βGal and βGal1-4βGlcNAc (or βGal1-3βGlcNAc) linkages. The terminal linkage αGalNAc1-3βGal or αGal1-3βGal always keeps the same orientation (centred at about $\Phi = 70°$ and $\Psi = 65°$) and is neither influenced by the other linkage orientations nor by the presence of the *N*-acetyl substituent. A third conformational family, which contains about 5% of the population, can also be identified. In this family, the terminal A or B disaccha ride determinants also maintained the same orientation.

If the reducing residue is not taken in account, i.e. when considering the A and B trisaccharide determinants, the conformational behaviour is totally different. The first conformational family of the A and B tetrasaccharides hardly exists anymore and the second family of conformations becomes the dominant one.

Discussion

Comparison with X-ray data

Due to the difficulties of crystallizing such compounds, no single crystals of histo-blood group oligosaccharides have been grown at the present time. The only crystallographic structure obtained of such oligosaccharides has been by X-ray investigations of the complex between a plant lectin, *Griffonia simplicifolia* isolectin IV (GSIV) and the Le^b tetrasaccharide [35]. The conformation adopted in the complex displays the values of $\Phi = -65°$, $\Psi = -100°$ for the αFuc1–2βGal linkage, $\Phi = -64°$, $\Psi = 138°$ for the βGal1-3βGlcNAc, and $\Phi = -63°$, $\Psi = -88°$ for the αFuc1-4βGlcNAc linkage. This conformation corresponds closely to the family predicted to be the most populated (Fam. I in Table 2). When fitting the ring atoms and glycosidic oxygen atoms of the lowest energy conformation to the X-ray structure, they superpose well and the resulting RMS deviation is only 0.23 Å. The only disagreement is a systematic shift in the value of the Φ angle. For the α-L and for the β-D linkages of this tetrasaccharide the MM3 calculations predict an optimal value around −80° whereas the protein crys-

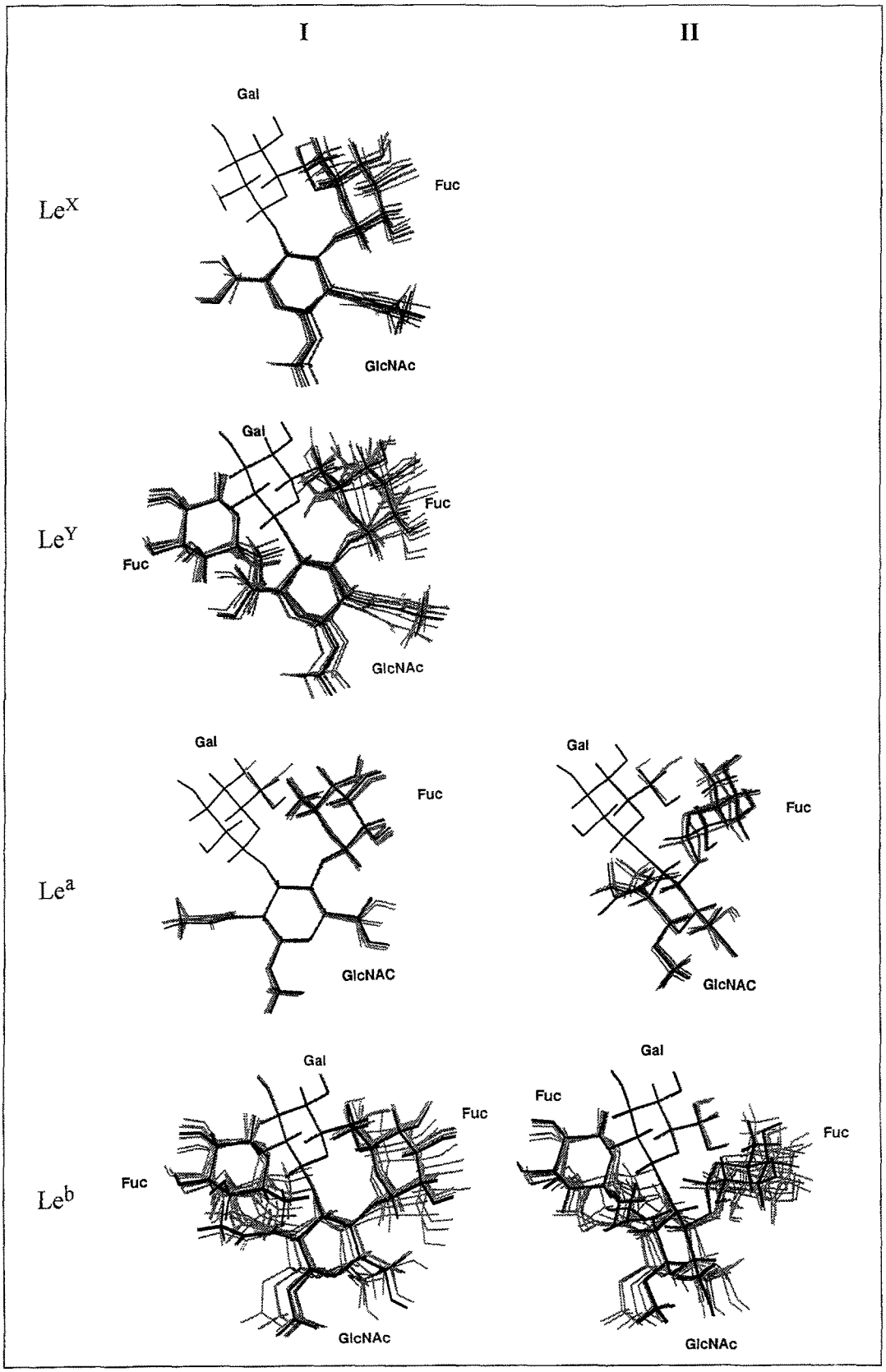


Figure 4. Graphic representation of all the conformational families of Lewis type oligosaccharides having a population of more than 1%. For each family, the lowest energy conformation has been drawn in black. The conformations with the largest difference in Φ and Ψ torsion angles within an energy window of 5 kcal mol^{-1} are drawn in grey.

Table 3. Characteristics of A and B oligosaccharide conformational families. Only the families having an energy-weighted population of more than 1% (at 296 K) have been listed. For each oligosaccharide and each family, E_{min} indicates the glycosidic linkages torsion angles and relative energy of the lowest energy conformation, whereas min and max are the limit values for each torsion angle within an energy window of 5 kcal mol^{-1}.

	Fam. I								*Fam. II*								*Fam. III*							
	Φ	Ψ	Φ	Ψ	Φ	Ψ	E_{rel}	%	Φ	Ψ	Φ	Ψ	Φ	Ψ	E_{rel}	%	Φ	Ψ	Φ	Ψ	Φ	Ψ	E_{rel}	%
A–1	Fuc12Gal		Gal13GlcNAc		GalNAc13Gal				Fuc12Gal		Gal13GlcNAc		GalNAc13Gal				Fuc12Gal		Gal13GlcNAc		GalNAc13Gal			
E_{min}	−71.4	−96.7	−75.1	136.7	65.3	64.8	0.11		−96.4	−169	−88.5	67	75.2	77.8	0.00		−101.4	−166.4	−60.3	171.4	73.4	77.6	1.24	
min	−75	−98	−81	128	59	59		**55.3**	−98	−174	−90	55	66	73		**38.4**	−111	178	−66	154	66	71		**5.5**
max	−67	−88	−66	151	82	74			−90	−166	−84	69	81	81			−97	−150	−55	166	79	80		
A–2	Fuc12Gal		Gal14GlcNAc		GalNAc13Gal				Fuc12Gal		Gal14GlcNAc		GalNAc13Gal				Fuc12Gal		Gal14GlcNAc		GalNAc13Gal			
E_{min}	−69.5	−93.5	−75.9	−115.4	65.8	67.9	0.00		−94.9	−169.5	−87.8	−179.1	76.7	78.0	0.29		−110.9	−155.2	−65	−99	74.1	78.4	1.64	
min	−80	−104	−81	−125	59	52		**58.1**	−106	176	−92	175	69	66		**35.2**	−145	179	−73	−151	64	58		**6.0**
max	−66	−87	−67	−90	89	74			−85	−164	−83	−175	82	81			−90	−148	−14	−85	83	79		
A_{tri}	Fuc12Gal				GalNAc13Gal				Fuc12Gal				GalNAc13Gal											
E_{min}	−68.2	−88.3			71.7	68.7	1.72		−94.8	−168.7			73.6	78	0.00									
max	−77	−109			62	61		**7.0**	−102	−173			64	69		**92.5**								
min	−67	−86			82	74			−91	−164			79	81										
B–2	Fuc12Gal		Gal14GlcNAc		Gal13Gal				Fuc12Gal		Gal14GlcNAc		Gal13Gal				Fuc12Gal		Gal14GlcNAc		Gal13Gal			
E_{min}	71.4	−94.9	−74.8	−114.7	71.4	66.9	0.00		−100.2	−166.9	−87.6	−176.1	73.5	75.9	0.38		−129.5	−157.4	−69.2	−111.9	70.5	70.8	1.48	
min	−74	−101	−77	−123	66	61		**72.6**	−105	−172	−89	178	72	75		**21.0**	147	−173	−74	−126	68	63		**5.5**
max	−69	−90	−68	−102	101	73			−95	−166	−85	−176	78	78			−101	−151	−52	−85	84	78		
B_{tri}	Fuc12Gal				Gal13Gal				Fuc12Gal				Gal13Gal											
E_{min}	−68.5	−90.1			73.6	67.2	1.62		−98.8	−168.5			77.0	78.4	0.00									
min	−78	−110			63	60		**5.4**	−101	−173			68	65		**93.6**								
max	−65	−86			88	72			−91	−164			89	81										

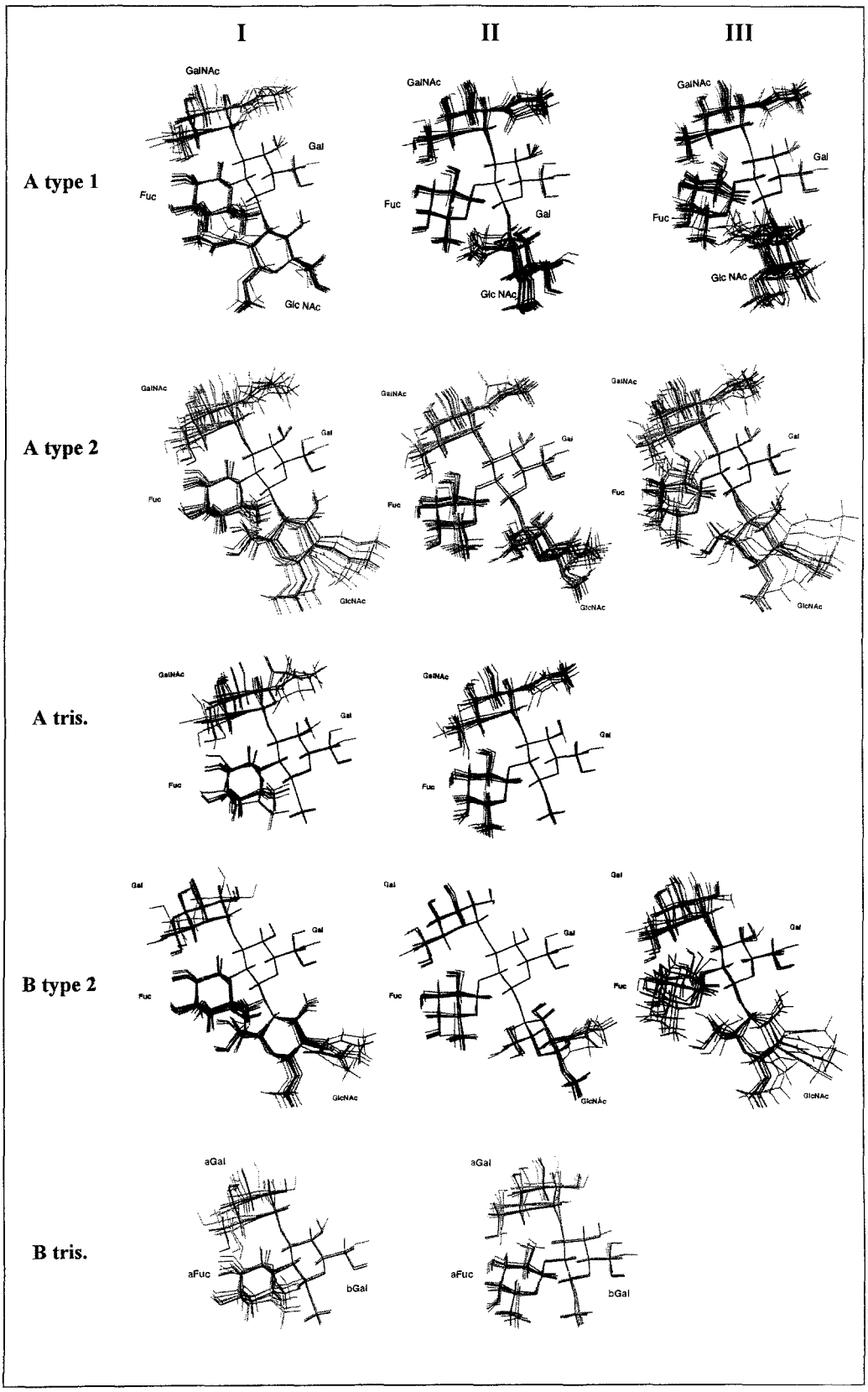


Figure 5. Graphic representation of all the conformational families of A and B type oligosaccharides having a population of more than 1%. For each family, the lowest energy conformation has been drawn in black. The conformations with the largest difference in Φ and Ψ torsion angles within an energy window of 5 kcal mol^{-1} are drawn in grey.

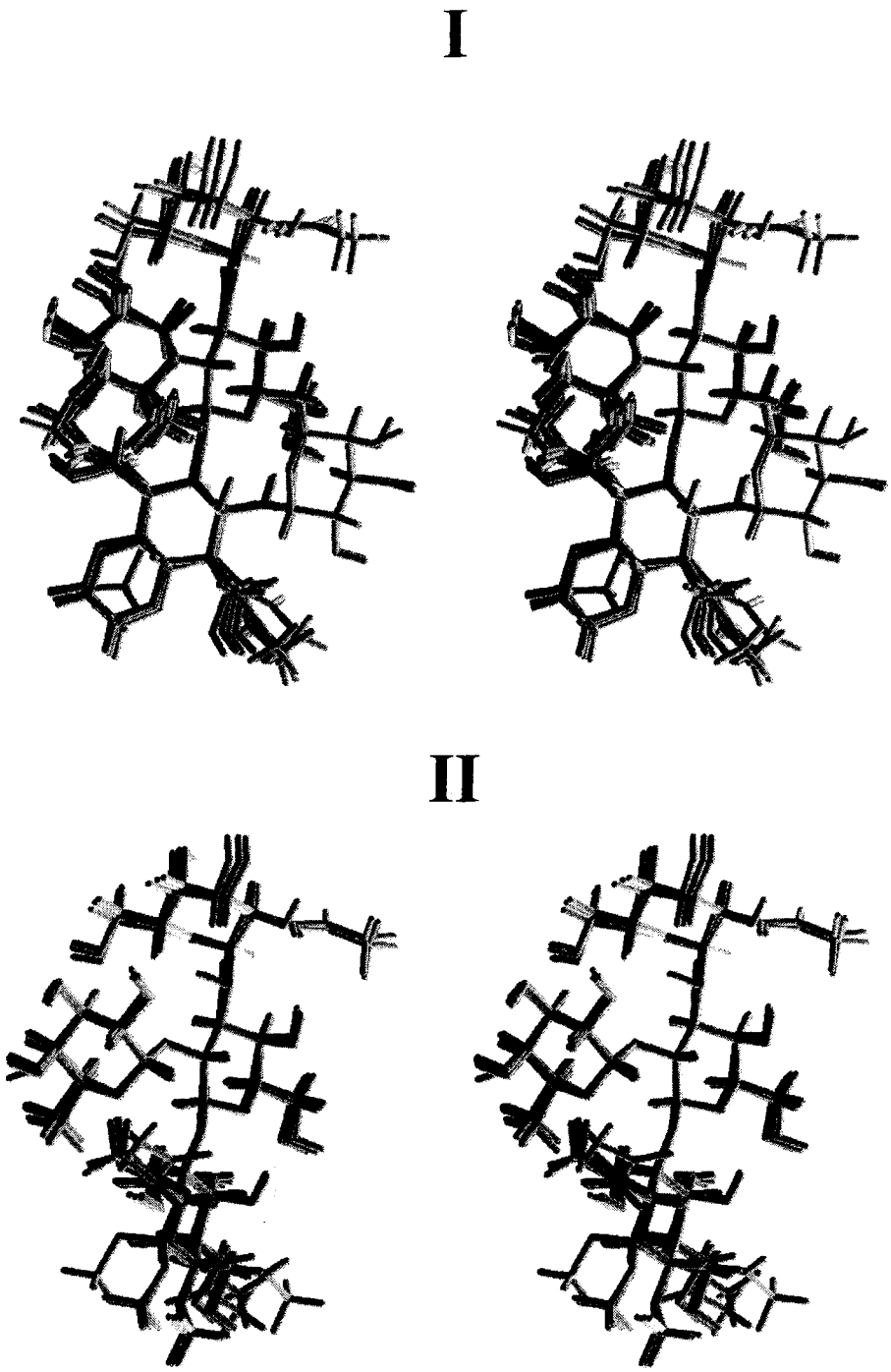


Figure 6. Stereo plots of two models for the alignment of histo-blood group oligosaccharides based on a fit of the ring and glycosidic atoms of the H trisaccharides core. I alignment of the lowest energy conformers of the conformational family Fam. I. II Alignment of the lowest energy conformers of the conformational family Fam. II.

tallography has determined a value of –60°. From a survey of carbohydrate crystal structures [36], it has been shown than 80% of this type of linkage have an exo-anomeric conformation in the range –70° to –90°. Since the protein crystal structure has been refined with a version of X-PLOR [37] which has not been modified for carbohydrate stereochemical specificities, this could explain why the value of the Φ angle differs slightly from the optimal one.

Comparison with high resolution NMR data

Most experimental data has arisen from high resolution NMR spectroscopy, with the use of nOe measurements. Since the

Table 4. Structural variations between the lowest energy conformation of the most populated conformational family (Fam. I) of the 10 histo-blood group oligosaccharides which contains an H trisaccharide moiety. The fit procedure has been performed for the 20 atoms constituting the three carbohydrate rings and the two glycosidic oxygens and the root mean squares (Å) have been calculated from the same atoms.

	H type 2	*H type 3*	*H type 4*	*H type 6*	*Le^Y*	*Le^b*	*A type 1*	*A type 2*	*B type 2*
H type 1	0.11	0.09	0.10	0.10	0.12	0.13	0.18	0.22	0.18
H type 2		0.08	0.08	0.08	0.14	0.16	0.17	0.17	0.12
H type 3			0.02	0.07	0.18	0.19	0.20	0.20	0.17
H type 4				0.07	0.18	0.19	0.20	0.20	0.17
H type 6					0.17	0.19	0.23	0.24	0.20
Le^Y						0.05	0.14	0.21	0.18
Le^b							0.13	0.22	0.19
A type 1								0.12	0.05
A type 2									0.05

Table 5. Structural variations between the lowest energy conformation of the second most populated conformational family (Fam. II) of eight histo-blood group oligosaccharides. See Table 4 for details.

	H type 2	*H type 3*	*H type 4*	*H type 6*	*A type 1*	*A type 2*	*B type 2*
H type 1	0.08	0.15	0.15	0.06	0.05	0.08	0.05
H type 2		0.18	0.16	0.07	0.10	0.05	0.09
H type 3			0.08	0.15	0.13	0.15	0.15
H type 4				0.14	0.14	0.14	0.16
H type 6					0.10	0.08	0.05
A type 1						0.07	0.08
A type 2							0.10

first NMR experiments reported by Lemieux and coworkers [9], nOe data have become available for H type 1 [38, 39], H type 2 [38], H type 4 [17] and H type 6 [40] trisaccharides. Experimental data are also available for Lea [16, 41], Leb [41] and LeX [42, 43] oligosaccharides, as well as for the tetrasaccharide determinants of blood group A type 1 and A type 2 [13, 44]. In many of the above mentioned cases, the NMR studies, often in association with molecular mechanics or a molecular dynamics study, proposes only one conformation for the blood group oligosaccharides. This unique conformation always belongs to the most populated conformational family in our calculations. It has to be kept in mind that these studies often use NMR data as constraints applied to the energy minimizations or to the dynamic trajectories and therefore the existence of only one family of conformation is an a priori hypothesis which is reflected in the results.

In recent studies on H type 4 [17] and H type 6 [40], NMR data were not used as constraints, and the occurrence of two conformations was predicted by the molecular modelling study. In the case of H type 6 trisaccharide [40], the two proposed conformations mainly differ by the orientation of the αFuc1-2βGal linkage. The lowest energy one corresponds to the most populated conformational family (Fam. I) of the present work, whereas the second one ($\Phi = -140°$, $\Psi = -140°$ for the αFuc1-2βGal linkage) does not have a reasonable energy. As stated by Ejchart and coworkers [40], this can probably be accounted for by a partial inadequacy of the CVFF force-field [45] which was not specifically parameterized for oligosaccharides.

In the case of H type 4 trisaccharide [17], the two conformations predicted by a force-field derived from MM2 [46] correspond closely to conformers of Fam. I and Fam. II of the present study. The authors considered the observed nOes to be in agreement with only one of their predicted conformations, the one corresponding to our most populated conformational family, Fam. I. In fact, when we calculated averaged distances over all the conformations of the CICADA potential energy surface, the agreement with NOESY derived interresidue distances was excellent. The occurrence of a second family of conformations (30% of the population) is therefore confirmed rather than disproved by the comparisons between theory and experiments.

The A group tetrasaccharides also contain a secondary family of conformations with a non-negligible population. The reported nOe data [13, 47] are mostly in agreement with the occurrence of the Fam. I type of conformers. Conformers of Fam. II, which represent almost 40% of the population in our calculations, do not seem to be detected in aqueous solution. However, in an early study, Bush and coworkers [44] noticed

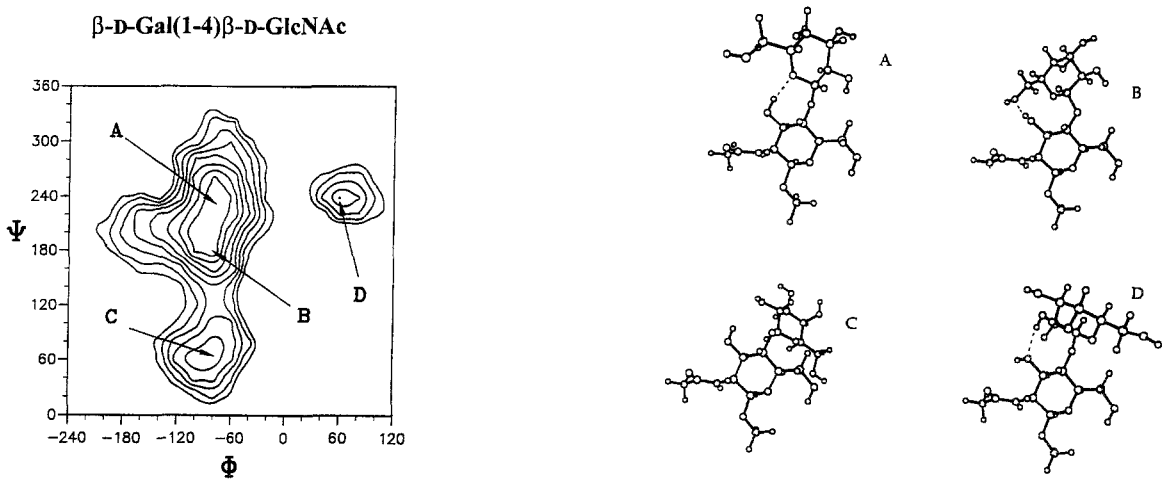


Figure S1. Adiabatic relaxed map of the βGal1-4βGlcNAc-*O*-Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. Isoenergy contours have been drawn at 1 kcal mol^{-1} increments above the absolute minimum. The outer line represents the 8 kcal mol^{-1} isoenergy contour. The lowest energy conformation of each domain has been represented by a ball and stick model and its location is indicated on the map. Inter residue hydrogen bonds are drawn as dotted lines.

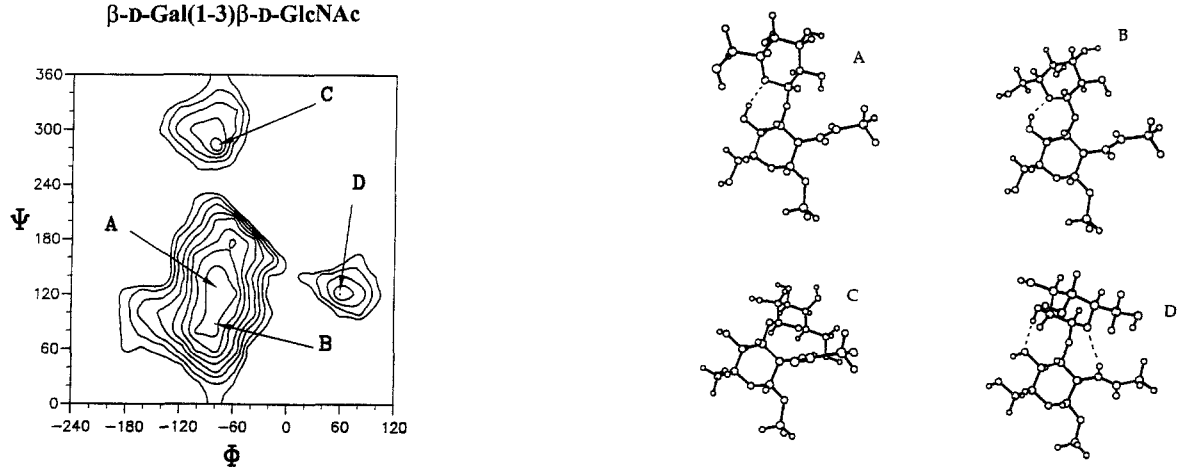


Figure S2. Adiabatic relaxed map of the βGal1-3βGlcNAc-*O*-Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. See Fig. S1 for details.

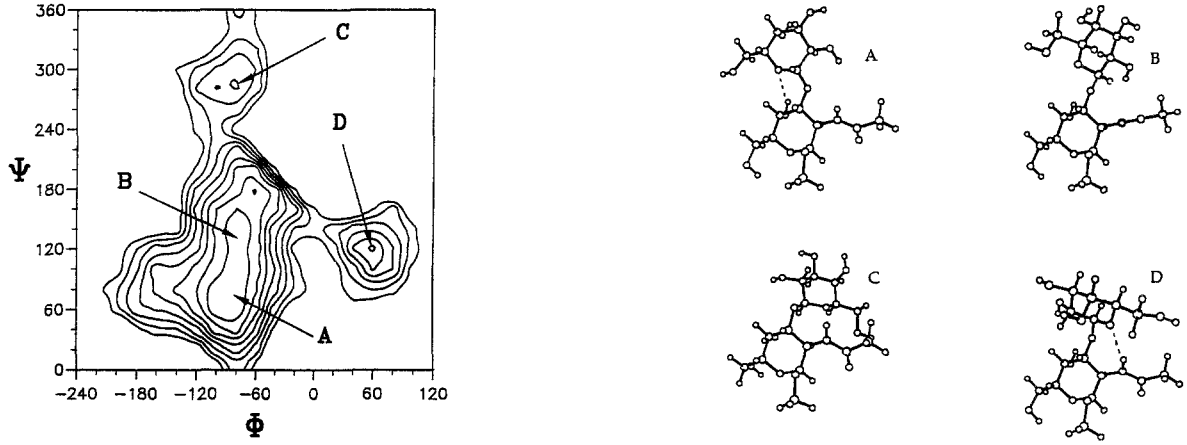


Figure S3. Adiabatic relaxed map of the βGal1-3αGalNAc-*O*-Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. See Fig. S1 for details.

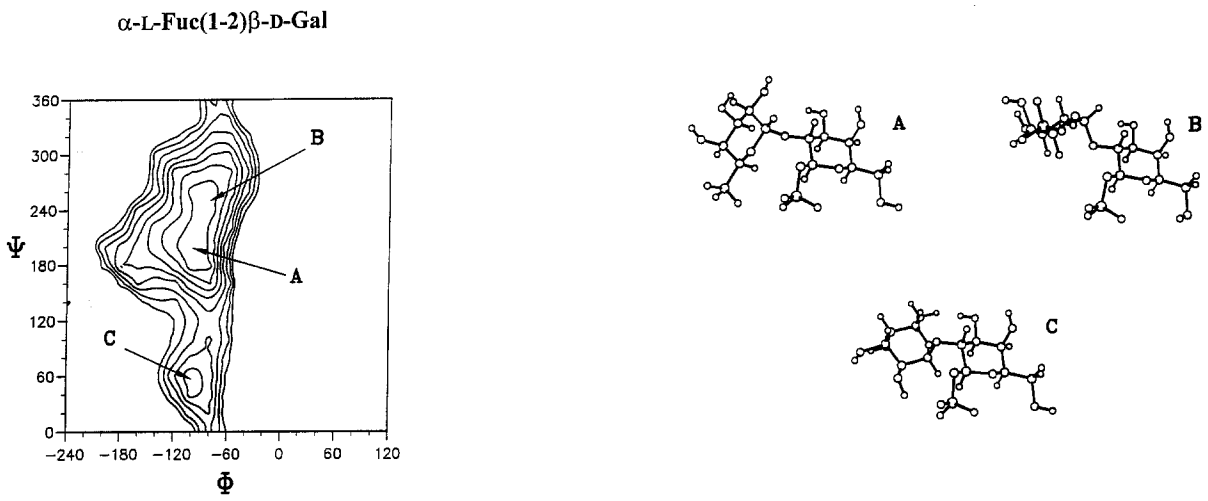


Figure S4. Adiabatic relaxed map of the αFuc1-2βGal-*O*-Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. See Fig. S1 for details.

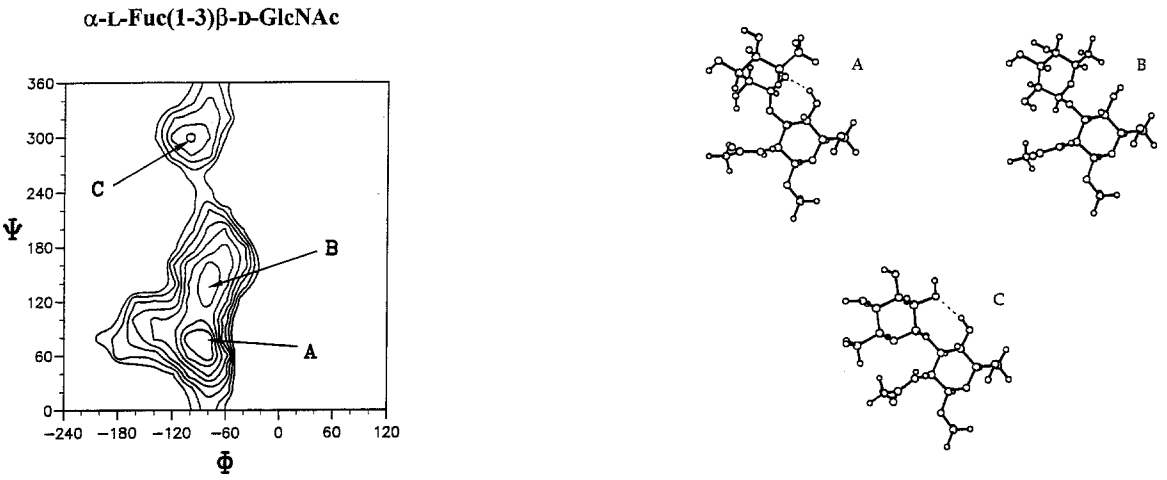


Figure S5. Adiabatic relaxed map of the αFuc1–3βGlcNAc–O–Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. See Fig. S1 for details.

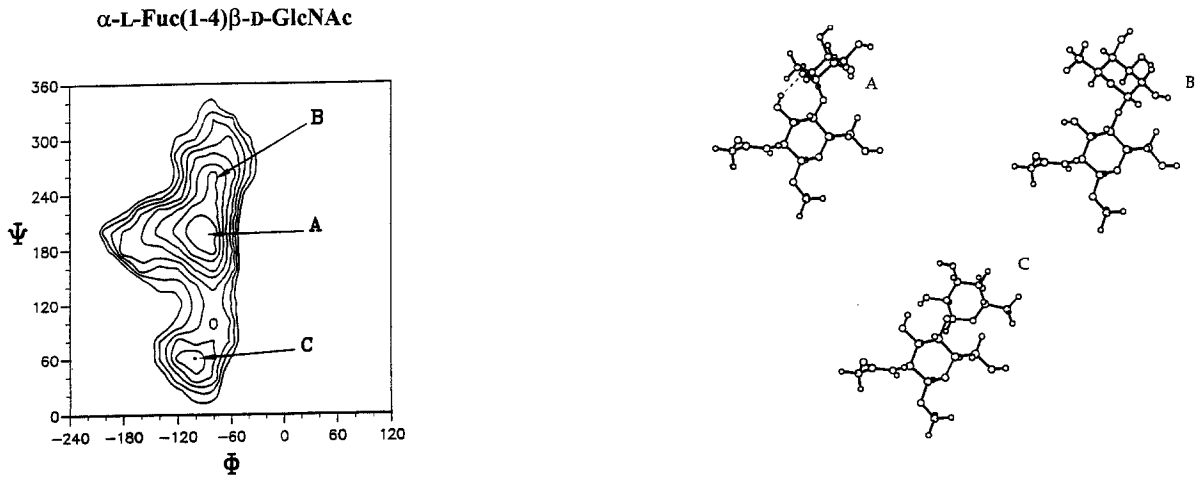


Figure S6. Adiabatic relaxed map of the αFuc1-4βGlcNAc-*O*-Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. See Fig. S1 for details.

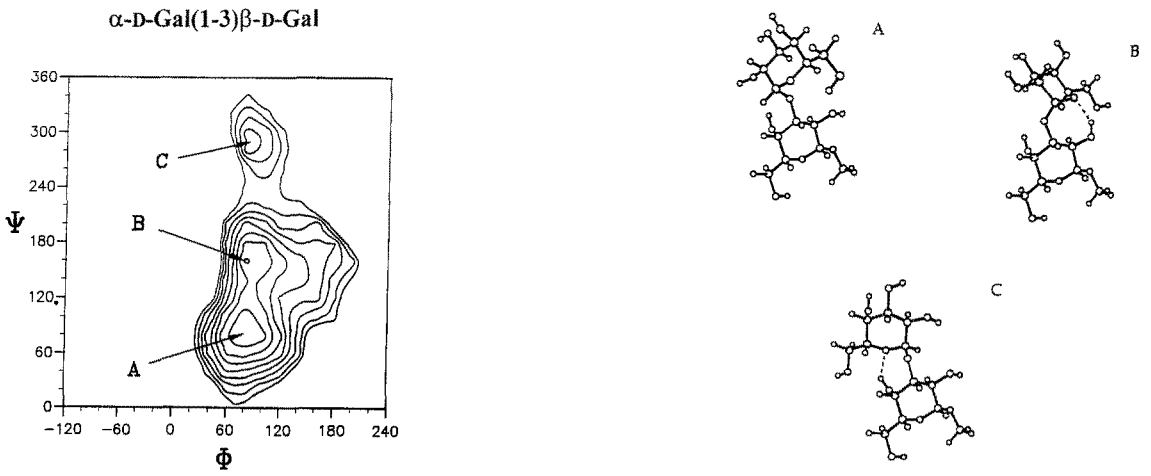


Figure S7. Adiabatic relaxed map of the αGal1-3βGal-*O*-Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. See Fig. S1 for details.

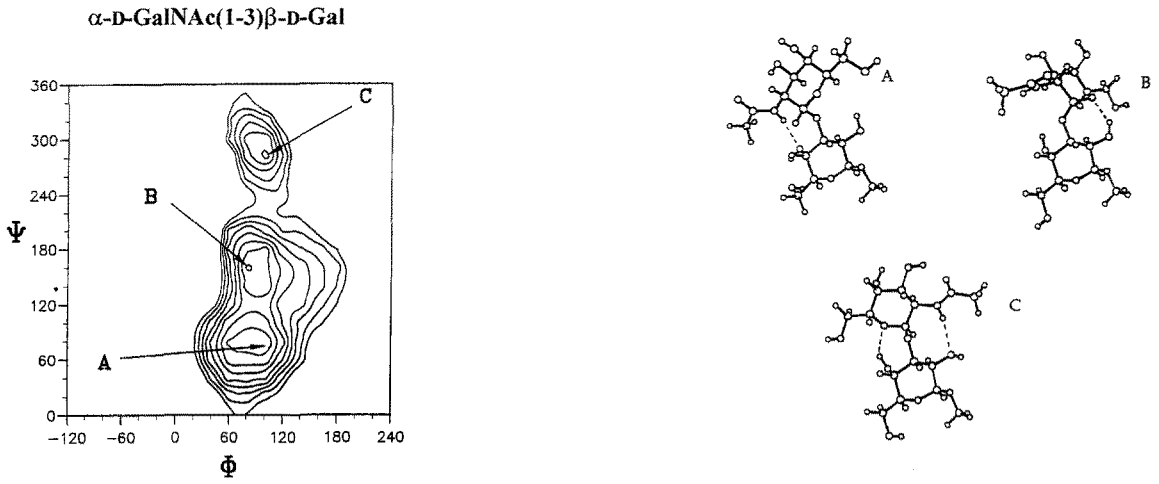


Figure S8. Adiabatic relaxed map of the αGalNAc1-3βGal-*O*-Me disaccharide as a function of the Φ and Ψ glycosidic torsion angles. See Fig. S1 for details

some important changes in the nOe on changing from D_2O to pyridine solvent. At that time, they explained this change in nOe by some small changes in conformation. In the light of the present calculations, it could be proposed that this change in nOe results from the change of populations between the two main conformational families.

Comparison with biochemical data

Validation of a computer simulation study can also be performed by testing the models against immunological or biochemical data. As for the histo-blood group series, a large amount of data are available for their molecular recognition by antibodies because of the medical importance of blood group incompatibility in blood transfusion, organ transplantation and pregnancy. To rationalize the complicated cross-reaction pattern that antibodies display towards histo-blood groups [48, 49], it is of interest to visualize the different shapes that the carbohydrate epitopes can adopt, and to quantify their similarities and differences.

From the ensemble of calculations described in this work, it is possible to propose one such general model; it consists of an alignment of the 10 histo-blood groups as either H or having an H precursor. Considering the lowest energy conformation of the most populated family (Fam. I) for each oligosaccharide, all the H trisaccharide moieties have been superimposed. The resulting alignment is displayed in Fig. 6I. The alignment of the three rings of the core trisaccharides is excellent as demonstrated by the root mean squares values listed in Table 4. This excellent alignment creates the common shape of all oligosaccharides. Differences arise from the presence or the absence of

the fourth residue (1-4 linked Fuc of the Lewis group, GalNAc of the A group, and Gal of the B group), but also from the nature of the core disaccharide. Depending on the 1-3 or 1-4 nature of the core linkage, the *N*-acetyl group is on one side or the other of the GlcNAc (or GalNAc) residue. Similarly, the *O*-methyl group can have different locations or orientations.

Most of these oligosaccharides have been shown to have more than one family of possible conformations. Therefore, it is important to consider this second family for alignment purposes. Even though the energies are less favourable, these secondary minima could be selected upon binding to a receptor. Consequently, a second model is proposed in Fig. 6II, based on the alignment of the lowest energy conformations of the second main conformational family (Fam. II). Only eight oligosaccharides have been taken into account since the Le^Y and Le^b tetrasaccharides have a quite different conformational behaviour. Table 5 presents the structural variations which are very limited. As with the first alignment, main differences are due to the presence of terminal substituents and to the different location of the *N*-acetyl group on the reducing residue.

The aim of these two models is to provide a basis for further reflections about the possible mode of binding of such molecules by antibodies or by plant lectins. We are currently using these alignments for a Quantitative Structure Activity Relationships (QSAR) study in relation to the immuno-chemical characterization of histo-blood group oligosaccharides (Imberty, Mikros, Carrupt, Mollicone, Oriol and Pérez, in preparation).

Conclusions

In pioneering work [9,10], similarities in shape between H type 1 and H type 2 trisaccharides, and by extension between $Lewis^b$ and $Lewis^Y$ tetrasaccharides have been proposed. Nevertheless, it is the first time that: (i) such alignment is extended to 10 histo-blood group oligosaccharides; and (ii) that secondary minima are taken into account by the proposal of a secondary possible alignment.

The CICADA procedure used in this work allowed us to investigate in a complete way the conformational behaviour of the oligosaccharides. When compared to molecular dynamic studies, it has the advantage of exploring conformational families in a very efficient way – whereas molecular dynamic trajectories are often trapped in one single energy well [17]. Also the procedure is interfaced with MM3 which has been shown to be well suited to carbohydrate conformational analysis, whereas the force-field used in molecular dynamic studies are not always well adapted for sugars.

Acknowledgements

This work was supported by CNRS, INRA and INSERM. The stay of E.M. in France has been financed in 1993 by the French Ministy of Research and in 1994 by a grant from the EEC. Florence Casset is acknowledged for her careful reading of the manuscript.

References

1. King M-J (1994) *Biochim Biophys Acta* **1197**:15–44.
2. Clausen H, Hakomori S (1989) *Vox Sang* **56**:1–20.
3. Oriol R, Le Pendu J, Mollicone R (1986) *Vox Sang* **51**:161–71.
4. Oriol R (1995) In *Blood Cell Biochemistry* Vol 6 (Cartron JP, Rouger P, eds) pp. 37–73. New York; Plenum Press
5. Oriol R, Cooper D, Davies JE, Keeling PWN (1984) *Lab Invest* **50**:514–18.
6. Lasky LA (1992) *Science* **258**:964–69.
7. Yuen, C-T, Bezouska K, O'Brien J, Stoll M, Lemoine R, Lubineau A, Kiso M, Hasegawa A, Bockovich NJ, Nicolaou KC, Feizi T (1993) *J Biol Chem* **269**:1596–98.
8. Hakomori S (1989) *Adv Cancer Res* **52**:257–331.
9. Lemieux RU, Bock K, Delbaere LTJ, Koto S, Rao VS (1980) *Can J Chem* **58**:631–53.
10. Biswas M, Rao VSR (1980) *Biopolymers* **19**:1555–65
11. Rao VSR, Biswas M (1985) *Top Mol Struct Biol* **8**:185–218.
12. Yan Z-Y, Bush CA (1990) *Biopolymers* **29**:799–811.
13. Cagas P, Bush CA (1992) Biopolymers **32**:277–92.
14. Berg EL, Robinson MK, Mansson O, Butcher EC, Magnani JL (1991) *J Biol Chem* **266**:14869–72.
15. Rutherford TJ, Spackman DG, Simpson PJ, Homans SW (1994) *Glycobiology* **4**:59–68.
16. Kogelberg H, Rutherford TJ (1994) *Glycobiology* **4**:49–57.
17. Toma L, Ciuffreda P, Colombo D, Ronchetti F, Lay L, Panza L (1994) *Helv Chim Acta* **77**:668–78.
18. Imberty A, Bourne Y, Cambillau C, Rougé P, Pérez S (1993) *Adv Biophys Chem* **3**:71–117.
19. Bourne Y, Rougé P, Cambillau C (1992) *J Biol Chem* **267**:197–203.
20. Cooke RM, Hale RS, Lister SG, Shah G, Weir MP (1994) *Biochemistry* **33**:1051–96.
21. Bundle DR, Baumann H, Brisson J-R, Gagné SM, Zdanov A, Cygler M (1994) *Biochemistry* **33**:5183–92.
22. MM3(1992), QCPE, Creative Arts Building 181, Indiana University, Bloomington, IN 47405, USA.
23. Allinger NL, Yuh YH, Lii JH (1989) *J Am Chem Soc* **111**:8551–66.
24. Pérez S, Imberty A, Carver JP (1994) *Adv Comput Biol* **1**:146–202.
25. French AD, Dowd MK (1993) *J Mol Struct (Theochem)* **286**:183–201.
26. Koča J (1994) *J Mol Struct (Theochem)* **308**:13–24.
27. Koča J, Pérez S, Imberty A (1995) *J Comp Chem* **16**:296–310.
28. Pérez S, Delage M-M (1991) *Carbohydr Res* **212**:253–59.
29. IUPAC-IUB (1971) Commission on Biochemical Nomenclature. *Arch Biochem Biophys* **145**:405–21.
30. Marchessault RH, Pérez S (1979) *Biopolymers* **18**:2369–74.
31. Koča J (1993) *J Mol Struct* **291**:255–69.
32. Imberty A, Pérez S (1994) *Glycobiology* **4**:351–66.
33. SYBYL 6.04, Tripos Associates, 1699 S. Hanley Road, Suite 303, St Louis, MO 63144, USA.
34. Lemieux RU, Koto S, Voisin D (1979) In *Anomeric Effect, Origin and Consequences* ACS Symposium Series Vol 87 (Szarek A, Horton D, eds) pp. 17–29. Washington DC: American Chemical Society.
35. Delbaere LTJ, Vandonselaar M, Prasad L, Quail JW, Wilson KS, Dauter Z (1993) *J Mol Biol* **230**:950–65.
36. Scheifer L, Senderowitz H, Aped P, Tartakowsky E, Fuchs B (1990) *Carbohydr Res* **206**:21–39.

37. Brünger AT, Krukowski A, Erikson JW (1990) *Acta Crystallogr A* **47**:585–93.
38. Rao BNN, Dua VK, Bush CA (1985) *Biopolymers* **24**:2207–29.
39. Cagas P, Kaluarachchi K, Bush CA (1991) *J Am Chem Soc* **113**:6815–22.
40. Ejchart A, Dabrowski J, von der Lieth C-W (1992) *Magn Res Chem* **30**:S105–14.
41. Cagas P, Bush CA (1990) *Biopolymers* **30**:1123–38.
42. Homans SW, Forster MJ (1992) *Glycobiology* **2**:143–51.
43. Miller KE, Mukhopadhyay C, Cagas P, Bush CA (1992) *Biochemistry* **31**:6703–9.
44. Yan Z-Y, Rao BNN, Bush CA (1987) *J Am Chem Soc* **109**:7663–69.
45. Hagler AT, Lifson S, Dauber P (1979) *J Am Chem Soc* **101**:5122–30.
46. Allinger NL, (1977) *J Am Chem Soc* **99**:8127–34.
47. Bush CA, Yan Z-Y, Rao BNN (1986) *J Am Chem Soc* **108**:6168–73.
48. Oriol R, Samuelsson BE, Messeter L (1990) *J Immunogen* **17**:279–99.
49. Good AH, Yau O, Lamontagne LR, Oriol R (1992) *Vox Sang* **62**:180–89.