Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 339 (2004) 949–959
Carbohydrate RESEARCH

Conformational analysis of complex oligosaccharides: the CICADA approach to the uromodulin O-glycans

Gianluca Cioci,[a,*] Alain Rivet,[a] Jaroslav Koča[a,b] and Serge Pérez[a,*]

[a]*Centre de Recherches sur les Macromolécules Végétales, CNRS and Joseph Fourier University, IFR 2607, BP 53, F-38041 Grenoble, France*
[b]*National Centre for Biomolecular Research, Masaryk University, CZ-61137 Brno, Czech Republic*

Received 29 September 2003; accepted 12 December 2003

Abstract—Uromodulin is the pregnancy-associated Tamm–Horsfall glycoprotein, with the enhanced ability to inhibit T-cell proliferation. Pregnancy-associated structural changes mainly occur in the O-glycosylation of this glycoprotein. These include up to 12 glycan structures, made up of an unusual core type 2 sequence terminated with one, two, or three sialyl Lewis^x sequences; this type of O-glycans could serve as E- and P-selectin ligands. The present work focuses on the most complex one; a tetradecamer made up of a type 2 core carrying three sialyl Lewis^x branches. Five different monosaccharides are assembled by 14 glycosidic linkages. The conformational behavior of the constituting disaccharide segments was evaluated using the flexible residue procedure of the MM3 molecular mechanics procedure. For each disaccharide, the adiabatic energy surface, along with the local energy minima were established. All these results were used for the generation, prior to complete optimization of the tetradecamer. This was followed by a complete exploration of conformational hyperspace throughout the use of the single coordinate method as implemented in the CICADA program. Despite the potential flexibility of the tetradecasaccharide, only four conformational families occur, accounting for more than 95% of the total low energy conformations. For each family, the molecular properties (electrostatic, lipophilicity, and hydrogen potential) were studied. The shape of the tetradecasaccharide is best described as a flat ribbon, flanked by three branches having terminal sialyl residues. Two of the branches interact through nonbonded interactions, bringing further energy stabilization, and limiting the conformational flexibility of the sialyl residues. Only one branch maintains the original conformational features of sialyl Lewis^x. This O-glycan can be seen as a fascinating example of 'dendrimeric' structure, where the spatial arrangement of three S-Le^x epitopes may favor its complementary 'presentations' for the interactions with E- and P-selectins.

Keywords: Core 2 branched O-glycans; Sialyl Lewis^x

1. Introduction

Tamm–Horsfall (THp) is the major glycoprotein produced by the kidney,[1,2] where it is expressed via a glycosylphosphatidylinositol (GPI) anchor on the endothelium of the thick ascending limb of the loop of the Henle. This 94 kDa glycoprotein can be released from its membrane anchor through proteases or phospholipase and is excreted in urine. The physiological roles of THp are still under dispute, and many hypotheses have been formulated. THp could act as an inhibitor of microbial infection of the urinary tract and of antigen-specific T-cell proliferation. Also, the aggregation and gel formation capabilities of THP could play important roles in several physiological and pathological states of the kidney. Pregnancy-associated THp is called uromodulin.[3] This is an immunosuppressive molecule whose ability to inhibit T-cell proliferation is increased 13-fold during pregnancy.[4] THp exhibits extremely large glycosylation microheterogeneities, as over 150 glycan structures that are mainly sialylated and/or sulfated to varying extents have been found on seven glycosylation sites.[5] The carbohydrate part can account for 28% of the total weight of the THp. Mass

* Corresponding authors. Tel.: +33-4-76037630; fax: +33-4-76037629; e-mail: gianluca.cioci@cermav.cnrs.fr

0008-6215/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2003.12.022

spectrometric strategies have recently uncovered pregnancy-associated changes in the O-glycosylation of this glycoprotein. THp from nonpregnant females and males expresses primarily core 1 type O-glycans terminated with either sialic acid or fucose, but not with the sialyl Lewis^x epitope. By contrast, the O-glycans linked to uromodulin include unusual core type 2 glycans terminated with one, two, or three sialyl Lewis^x sequences (Fig. 1). These changes in the glycosylation pattern could account for the enhanced immunomodulatory effects of uromodulin.[6] Meanwhile, different studies have established that sialyl Lewis^x (S-Le^x) in core 2 branched O-glycans serves as an E- and P-selectin ligand while S-Le^x containing type 1 core O-glycans are potential ligands of L-selectin.[7] The Le^x blood-group antigens have been the subjects of many studies in the past and several NMR and conformational studies converged to the assumption that these molecules exist in just a few, if not only one, conformations in solution.[8] This does not mean that these oligosaccharides are completely rigid but that their glycosidic torsion angles exhibit small oscillations around the average values.[9] In the case of S-Le^x, its structure has been idealized as a rigid Le^x core linked to sialic acid via a flexible bond, which allows the co-existence of not more than two or three conformations accessible to the molecule.[10] The S-Le^x has already been extensively studied throughout molecular mechanics calculations.[8,10] Therefore, from both the computational and biological point of view, it is of interest to see how the interactions with neighboring groups in complex oligosaccharides can modulate or alter its conformational behavior.

There are 12 proposed structures for the uromodulin O-glycans and in this work we focused our attention on the most complex one. This molecule, which we named Core2(S-Le^x)$_3$, is a tetradecamer made up of a type 2 core, which is linked to three S-Le^x groups (Fig. 1). Five different monosaccharides are assembled by 14 glycosidic linkages in quite a unique fashion to give a very complex oligosaccharide of 366 atoms.

The treatment of a glycan at this level of complexity can be undertaken using the following route: (a) identification of the monosaccharide constituents; (b) identification of disaccharide segments of Core2(S-Le^x)$_3$; (c) molecular modeling of each segment, and description of the potential energy surfaces; (d) collection of the 3-D structures of low energy conformers; (e) building realistic possible starting conformations from assembly of the low energy conformers; (f) exploration of potential energy hyperspace; (g) clustering of the low energy conformers into families; (h) analysis and depiction of representative within each low energy family. The present investigation makes use of well-established computational protocols based on the molecular mechanics MM3[11] and single coordinate driving method (SCD) as implemented in the CICADA program.[12,13] There are three main advantages of the method over other conformational search algorithms: (i) the computer time required for a calculation has a polynomial dependence on the degrees of freedom, in contrast to the grid search, which has an exponential one; (ii) all the interconversion pathways between families of conformations are explored but; (iii) the algorithm spends almost its time in the essential highly populated areas, giving a realistic and time-inexpensive description of conformational space. It has been shown that CICADA is an efficient tool to search conformational space, and in the case of small and middle sized oligosaccharides it produces reliable results, which are comparable to a carefully performed grid search, but

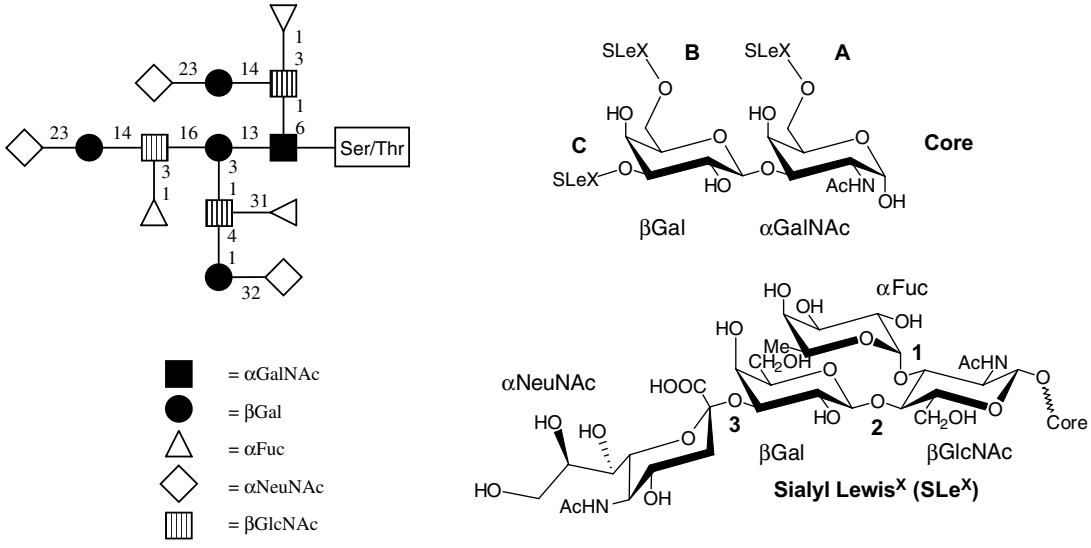

Figure 1. On the left, the structure of the whole molecule. On the right, the representations of the core disaccharide and of the S-Le^x structure. The S-Le^x groups are linked to the core structure via the linkages labeled A, B, and C.

much more time efficient.[14–16] Despite that, an SCD search with a very complex molecule has never been carried out before. In the present paper, we present the methodology of the study together with information about the three-dimensional structure of the molecule and some biomolecular implications that can be extrapolated from the results of the calculation.

2. Computational procedure

2.1. Nomenclature

The recommendation and symbols proposed by the Commission on Biochemical Nomenclature are used throughout this paper. The torsion angles describing the glycosidic linkages are defined as $\Phi = (O5–C1–O1–C'X)$, $\Psi = (C1–O1–C'X–C'X+1)$, and $\omega = (O1–C6–C5–O1)$.

The labeling of the different glycosidic linkages was made according to the following rules. The core disaccharide is β-Gal(1 → 3)-α-GalNAc. With reference to Figure 1, the three branched S-Le^x and the linkages from the core to them are labeled as A, B, and C, respectively. Within each S-Le^x moiety, the α-Fuc(1 → 3)-β-GlcNAc linkage is referred to as 1, the β-Gal(1 → 4)-β-GlcNAc as 2 whereas the α-NeuNAc(2 → 3)-β-Gal is referred to as 3. For example, the β-Gal(1 → 4)-β-GlcNAc linkage on the S-Le^x group attached (1 → 3) to the core disaccharide will be referred to as C2. The nomenclature is summarized in Table 1.

2.2. Building blocks

The starting geometry and suggestion of each of the five different constituting monosaccharides was taken from the MONOBANK data base[17] or constructed. Each residue has been submitted to a complete energy optimization through MM3.

2.3. Relaxed map of disaccharides

The conformational space of each disaccharide segment was explored by stepping the glycosidic linkage Φ, Ψ, and ω torsion angles in 15° increments over the whole angular range. Adiabatic contour maps were drawn for each disaccharide by using different orientations of the pendent groups. In particular, three staggered positions of the hydroxymethyl groups (*gt*, *gg*, and *gt*) and the two possible networks of hydrogen bonds around the sugar rings (clock- and anticlockwise). As for the sialic acid moiety, two orientations were considered for the acidic group. At each conformational microstate, a geometry optimization was performed by allowing the coordinates of each atom to vary except those defining the Φ and Ψ (and ω) torsion angles. Only the lowest energy conformer at each (Φ, Ψ) point is used for the Ramachandran-like contour plots. This procedure has the advantage of overcoming the well-known multiple-minima problem of the potential energy hypersurface due to the specific orientations of the pendant groups, and also of fully describing the conformational flexibility around the glycosidic linkages. The exact position of the different minima was determined after additional minimizations by removing the constraints on the Φ and Ψ torsion angles.

2.4. Oligomeric structure

Information from the maps computed for the parent disaccharides was used to construct the oligomeric structure. The starting geometry-(ies) of Core2(S-Le^x)$_3$ was constructed using the POLYS program.[18]

2.5. Exploring the hyperdimensional potential energy surface using CICADA

The potential energy hypersurface of Core2(S-Le^x)$_3$ was explored by the CICADA program using the MM3 force field. Several starting structures were used as input for the CICADA analysis with different orientations around the glycosidic torsion angles, based on the low-energy conformations found in the relaxed maps of the corresponding disaccharides. During the CICADA calculations, all glycosidic torsion angles were driven by 15° step as well as those related to the orientation of the COOH and of the primary hydroxyl groups. The remaining torsion angles were only monitored. When CICADA detects a minimum, the conformation is fully optimized. The resulting structure is compared with the previously stored ones and stored if not yet detected. Structures corresponding to energy maxima, that is, the transition states, are also stored. For Core2(S-Le^x)$_3$,

Table 1. Nomenclature of all the disaccharides

	Sialyl Lewis^x labeling		
	A	B	C
Core to:	β-GlcNac(1 → 6)-α-GalNAc	β-GlcNAc(1 → 6)-β-Gal	β-GlcNAc(1 → 3)-β-Gal
Inside the S-Le^x:			
1	α-Fuc(1 → 3)-β-GlcNAc	α-Fuc(1 → 3)-β-GlcNAc	α-Fuc(1 → 3)-β-GlcNAc
2	β-Gal(1 → 4)-β-GlcNAc	β-Gal(1 → 4)-β-GlcNAc	β-Gal(1 → 4)-β-GlcNAc
3	α-NeuNAc(2 → 3)-β-Gal	α-NeuNAc(2 → 3)-β-Gal	α-NeuNAc(2 → 3)-β-Gal

the dimensionality of the conformational hyperspace explored in the CICADA run was 28. In addition, all the hydroxyl groups were monitored, which means that the orientations were taken into account for detecting new energy minima. The run was considered as complete when no new minima were detected in an energy window of 4 kcal mol^{-1} above the absolute minimum.

2.6. Force field

Geometry optimization was performed using the molecular mechanics program MM3. This force field contains a correction for the anomeric effect and has been shown to be especially adapted for the study of carbohydrates.[19] The block diagonal minimization method with the default energy convergence criterion of 0.00008**n* kcal mol^{-1} per five iterations, *n* being the number of atoms, was used for grid point optimization. To mimic the hydrated environment of the molecules, a dielectric constant, $\varepsilon = 78.5$, was used in all the calculations. All calculations were performed on an SGI Origin 300 (R14000, 600 MHz processor).

2.7. Determination of families of low energy conformers

The raw data from the CICADA calculation were analyzed with the program PANIC,[21] which extracts the low energy conformations along with the transition states and analyzes the network of interconversions between them. Then the population of the low energy conformations has been analyzed by an 'in house' program in order to cluster the conformations into families.[22] In this approach, inside a given energy window, a conformer is considered to belong to a family if all its torsion angles differ by no more than a given value from the torsion angles of at least one of the members of the family. In the present study, the value considered for determining the difference has been set to 30°. This means that two conformations are classified into different families if there is at least one glycosidic torsion angle, which differs by more than 30°. The coordinates of the energy minima will be deposited or made available for users on the internet through the CERMAV website.

3. Results and discussion

The exploration of the 28 dimensions of the conformational space throughout CICADA took slightly less than three months of calculation on an R14000 600 MHz processor. From this massive amount of data, 11,471 points in the potential energy hyperspace were stored, of which 2637 were low energy conformations and 5691 were transition states. One should not be surprised at the time employed by the program to perform this calculation, as a classical grid search method on such a complex molecule would simply not have been feasible. For instance, a 20° step scan on all the 28 glycosidic bonds would require something like 18^{28} energy minimizations! Even taking into account only the combination of the lowest energy conformations shown by each disaccharide unit, say three each one, we can estimate the number of conformations as 3^{14} that equals to 4,782,969.

3.1. Analysis of linkage conformations

Because of the large amount of data produced by the calculations, together with the intrinsic very high multidimensionality of this conformational space, the analysis of the results is not an easy task to accomplish. A general way to proceed is to first take into account the structural information related to the conformations taken at each glycosidic linkage. It is possible to superimpose the conformations found by CICADA on the energy maps of the constituting disaccharides.

The energy map calculated for the α-Fuc(1 → 3)-β-GlcNAc disaccharide (Fig. 2) displays a main low energy region centered at $\Phi = 270$, $\Psi = 70$ and a second one, slightly higher in energy (+1 kcal mol^{-1}) at $\Phi = 280$, $\Psi = 150$. When looking at conformations found by CICADA, the major part falls over the zone centered at $\Phi = 280$, $\Psi = 150$, while the other minimum, which is the lowest one for the disaccharide, is slightly populated only in the case of the linkage C1.

The same reduction in flexibility takes place at the β-Gal(1 → 4)-β-GlcNAc linkages. While the map for the disaccharide exhibits three low energy zones (Fig. 3), most of the conformations found for this glycosidic linkage in the oligosaccharide falls around the upper edge of the main energy well, indicating a complete reduction in flexibility for the A2 and B2 linkages. Only the linkage C2 appears to be more flexible, with some conformations occurring within the third low energy zone centered at $\Phi = 280$, $\Psi = 60$.

In the case of an 'isolated' Le^x group, these reductions in flexibility are quite a general feature of the molecule. From many NMR and theoretical investigations,[15,23–25] it has been proposed that important nonbonded interactions occur between the galactose and fucose residues, thereby reducing the conformational flexibility of both linkages. These features have been confirmed by the crystal structure elucidation of this oligosaccharide.[26]

The orientations of the α-NeuNAc(2 → 3)-β-Gal linkages in the conformers found by CICADA correspond closely to the low energy zones of the disaccharide map. It can be clearly seen that the conformational flexibility arises mainly from the rotation around the Φ angle. The α-NeuNAc(2 → 3)-β-Gal is a flexible glycosidic linkage that is not perturbed from the effects shown before. Nevertheless, also in this case, some differences can be observed between the three linkages, as indicated

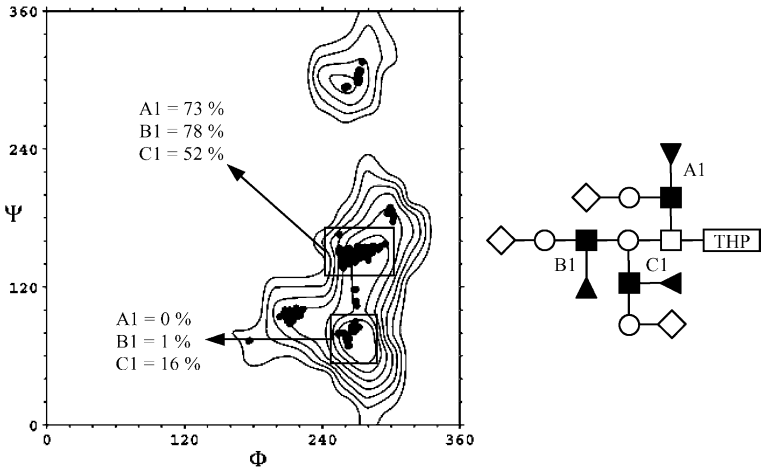

Figure 2. The map for the α-Fuc(1 → 3)-β-GlcNAc disaccharide, with the superimposed energy minima found by CICADA for the A1, B1, and C1 linkages (the black ones on the molecule representation on the right). The isocontours, with 1 kcal mol^{-1}, represent the energy levels of the disaccharide calculated using the grid search method. The map is plotted with the program Xfarbe.[20] Only the conformations within an energy window of 10 kcal mol^{-1} are shown. The areas with consistent population have been framed and labeled. The zone centered on $\Phi = 280$, $\Psi = 150$ is the most populated region. The lowest minimum of the disaccharide is populated only in the case of the linkage C1.

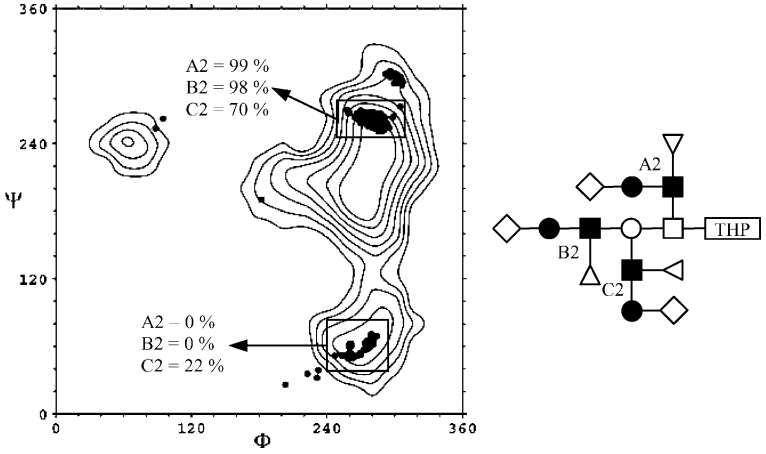

Figure 3. The map for the β-Gal(1 → 4)-β-GlcNAc disaccharide with the superimposed energy minima found by CICADA for the A2, B2, and C2 linkages. The zone centered on $\Phi = 270$, $\Psi = 200$ shows a complete lack of conformer population. Only the linkage C2 shows conformers in the other low energy regions.

by the relative populations of the lowest energy region centered on $\Phi = 60$, $\Psi = 100$ (Fig. 4).

The effects of the reduction in conformational flexibility are well illustrated below in the case of the linkages A and B, corresponding to the β-GlcNAc(1 → 6)-β-GalNAc and β-GlcNAc(1 → 6)-β-Gal disaccharides. The maps for the disaccharides A and B display the same features, both in positions and energy levels. However, the conformations found for these linkages in the oligosaccharide are totally different. In the case of β-GlcNAc(1 → 6)-β-GalNAc (A linkage) the lowest energy region is well populated (Fig. 5a) whereas for β-GlcNAc(1 → 6)-β-Gal (B linkage) the great majority of the low energy conformers are located around the minimum at $\Phi = 270$, $\Psi = 60$.

The ω torsion angles of the (1 → 6) linkages are known to have three possible orientations, *gauche–trans* (*gt*), *trans–gauche* (*tg*), and *gauche–gauche* (*gg*). As for the β-GlcNAc(1 → 6)-β-GalNAc (A linkage), the following percentages of occurrence were calculated: $gt = 72\%$, $tg = 12\%$, and $gg = 16\%$. The *gt* conformation is clearly the preferred one, while the other orientations start to have a consistent percentage only at higher energies (+6 kcal mol^{-1} over the global minima). The β-GlcNAc(1 → 6)-β-Gal (B linkage) is more rigid having shown a *gt* orientation that occurs for more than 98% of the total conformations and neglectable percentages of *tg* and *gg* orientations.

The disaccharide maps for the C and core linkages, which are buried inside the structure of the

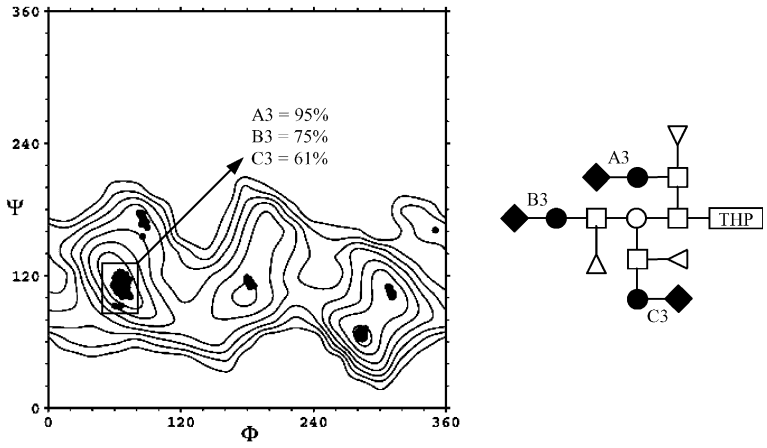

Figure 4. The map for the α-NeuNAc(2 → 3)-β-Gal disaccharide with the superimposed conformations found by CICADA for the A3, B3, and C3 linkages.

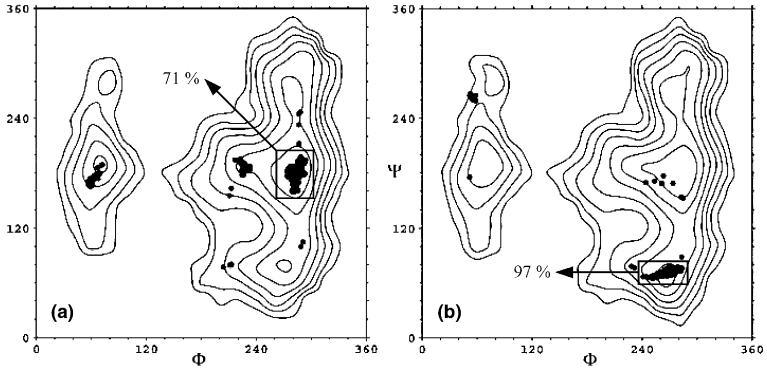

Figure 5. (a) and (b) From left to right, respectively, the maps for the disaccharides A and B, corresponding to the β-GlcNAc(1 → 6)-β-GalNAc and β-GlcNAc(1 → 6)-β-Gal linkages with the superimposed conformations found by CICADA. While the maps look rather the same, the conformations are distributed differently.

oligosaccharide, exhibit the same aspect, with a deep low energy well and two secondary regions. The conformations found for the oligosaccharide are all located in the lowest energy region whereas the secondary minima are not populated at all (Fig. 6).

Moreover, when the energy of the conformers is plotted as a function of the glycosidic torsion angle Ψ (Fig. 7), the occurrence of two distinct families with comparable percentages of occurrence for the β-Gal(1 → 3)-β-GalNac linkage is found.

The results that have been obtained for the linkages type 1, 2, and 3 groups can be compared with those obtained for an isolated S-Le^x molecule.[27] There are no major differences on the overall distribution of the conformations, which means that when linked to such a complex oligosaccharide, the S-Le^x moiety can be considered as a side group whose overall conformational behavior remains unchanged. However, we can see that the distribution of the conformations for the linkages in the C type S-Le^x matches closely the distributions calculated by CICADA for the isolated S-Le^x molecule. On the other hand, the reductions in flexibility exhibited by the linkages in A and B S-Le^x groups are more significant. This is of course the consequence of the nonequivalency (both in terms of position and neighboring groups) of the three S-Le^x moieties; in other words they 'feel' their environments in different ways. For all the linkages, we have shown how the nonbonded interactions that take place between adjacent residues can influence the conformational flexibility of these linkages. These results finally agree with the current idea that any occurrence of interactions between different residues can only result in a reduction of the available conformational space found for the disaccharides.[28,29]

3.2. Families of conformations and energy minima

To understand how nonbonded interactions can influence the overall shape of Core2(S-Le^x)$_3$ we first have to cluster all the conformations into families. In Table 2

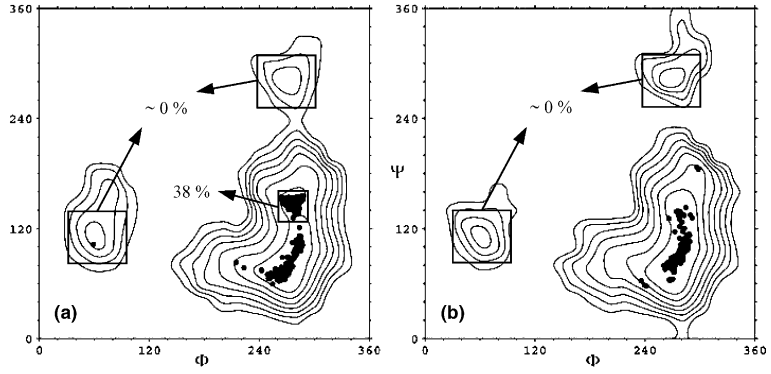

Figure 6. (a) and (b) From left to right, respectively, the maps for the disaccharides C and core, corresponding to the β-GlcNAc(1 → 3)-β-Gal and β-Gal(1 → 3)-α-GalNAc linkages, with the superimposed conformations found by CICADA.

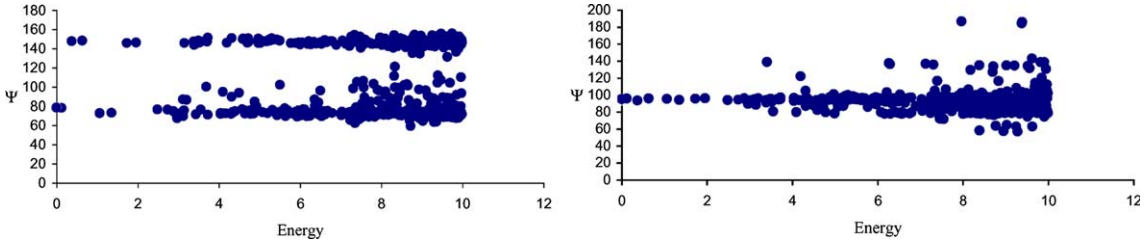

Figure 7. The energy of the conformations is plotted against the Ψ angle for the β-GlcNAc(1 → 3)-β-Gal linkage (on the left) and for the β-Gal(1 → 3)-α-GalNAc linkage (on the right). It is clear that for the β-GlcNAc(1 → 3)-β-Gal there are two main conformational families while the β-Gal(1 → 3)-α-GalNAc linkage only exhibits oscillations around its average value.

Table 2. The 15 best families found in a window of 10 kcal mol^{-1} over the global minimum

Family	Energy of the best conformer (kcal mol^{-1})	%	Glycosidic torsion angles, which differ by more than 30° from those of the global minimum (±Δ°)		
1	0.00	56.7		Global minimum	
2	+0.37	27.6	C Ψ (+69.3)		
3	+1.06	8.4		C3 Ψ (+57.4)	
4	+1.72	2.8	C Ψ (+67.4)	C3 Ψ (+57.7)	
5	+2.48	0.5		B3 Ψ (+63.9)	
6	+2.73	<0.5	C1 Φ (−69.5)	C1 Ψ (−55.9)	
7	+2.87	<0.5		C3 Φ (+244.3)	
8	+2.96	<0.5	C1 Ψ (−64.0)	C2 Ψ (−200.2)	
9	+2.97	<0.5	C1 Ψ (+142.7)	C2 Ψ (−199.0)	
10	+3.08	<0.5	C3 Φ (+215.9)	C3 Ψ (−54.1)	
11	+3.11	<0.5	B1 Φ (−58.0)	B1 Ψ (−49.8)	
12	+3.14	<0.5	C Ψ (+67.4)	B3 Ψ (+63.6)	
13	+3.19	<0.5	A1 Φ (−68.9)	A1 Ψ (−56.4)	
14	+3.37	<0.5	A1 Φ (−69.2)	A1 Ψ (−56.3)	C Ψ (+69.1)
15	+3.40	<0.5	C Ψ (+65.9)	B3 Φ (+220.3)	B3 Ψ (−46.1)

one can see that despite the great number of glycosidic torsion angles in the molecule, there are only a few families with a significant percentage of occurrence. Indeed, the global distribution of the conformations indicates the occurrence of only a few low energy conformers situated in deep and narrow energy wells. On the other hand, there exist families of conformations, with higher energy, that fill up all the conformational space. Such a feature is an indication of the occurrence of transition states that the oligosaccharide may follow to go from one low energy conformation to another one.

A careful analysis of the features found in the first 10 families indicates that major variations of glycosidic linkages take place at the C part of the molecule, (i.e., where the S-Le^x is linked 1 → 3 to the core). This qualitative observation agrees with the results gathered

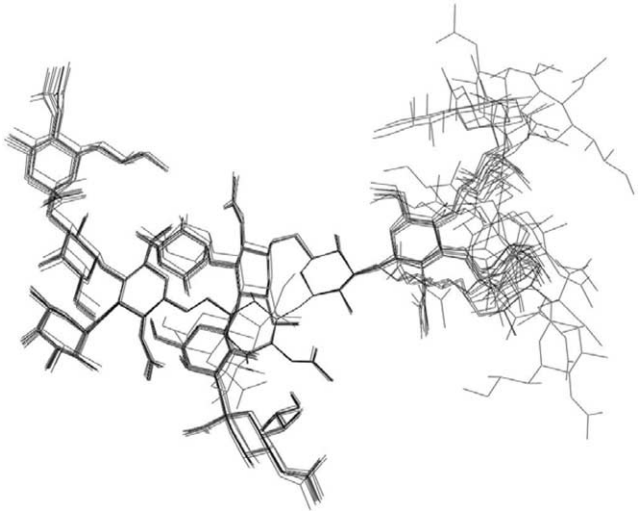

Figure 8. Superimposition of the best conformers from the first 10 families. The structures are superimposed on the core glycosidic linkage and the hydrogen atoms have been removed for clarity. The right part of the molecule is the C S-Le^x. At the extreme left, the A S-Le^x and on the center, the core disaccharide, which is partially covered by the B S-Le^x.

previously on the disaccharide potential energy surfaces. It was noticed that the linkages located at the C part of the oligosaccharide had more conformational flexibility than the linkages in the A and B S-Le^x. This is well illustrated in Figure 8, which shows the superimposition of the best conformers taken from the first 10 families. One can see that the C branch exhibits multiple conformations, particularly for the α-NeuNAc(2 → 3)-β-Gal (C3), the α-Fuc(1 → 3)-β-GlcNAc (C1) and the β-GlcNAc(1 → 3)-β-GalNAc (C) disaccharides. In the rest of the molecule, most of the glycosidic linkages show mainly oscillations around their average values.

When looking at the structure of the molecule, it is clear that the reduction in conformational flexibility of the A and B S-Le^x moieties can only be attributed to the energy stabilization that results from their spatial proximity. Indeed, several hydrogen bonds and stacking interactions take place between these two branches of Core2(S-Le^x)$_3$. These effects stabilize these two branches of the molecule in their lowest energy conformations, thereby reducing the conformational flexibility of the glycosidic linkages. In contrast, the C S-Le^x group is not involved in any long range interactions with other residues. As a consequence, it is more susceptible to exhibiting higher conformational flexibility.

3.3. Molecular shapes and properties

The whole Core2(S-Le^x)$_3$ oligosaccharide can be idealized as a backbone, formed by the core disaccharide and the GlcNAc residues of the A, B, and C S-Le^x groups, which is linked to three S-Le^x moieties. Consequently, the overall conformation of the whole molecule can be changed only by rotations about the glycosidic linkages of the backbone. As shown previously, these linkages are very rigid. Only the β-GlcNAc(1 → 3)-β-GalNAc (C) displays two distinct conformations having comparable percentage of occurrence. Therefore, the most interesting structural feature that is conserved in several families is the shape of the backbone conformation, while the terminal residues are more or less free to explore their conformational spaces. Figure 9 represents the lowest energy conformer of Core2(S-Le^x)$_3$. The backbone of the molecule is very flat, resulting from four adjacent residues that are in the same plane. The core disaccharide, together with the A and C S-Le^x branches form a curved sheet-like structure, which makes a 180° turn with respect to the plane defined by the backbone of the molecule. The maximum extension of the structure is about 27 Å. The B S-Le^x is placed above the core, forming a third arm that projects from the plane of the molecule. This architecture is highly stabilized by several hydrogen bonds between the A and B S-Le^x residues.

The lowest energy conformers of the first four families are represented in Plate 1, together with their molecular properties, electrostatic, lipophilicity, and hydrogen bonding potential. The second family differs only in the geometry of the β-GlcNAc(1 → 3)-β-GalNAc linkage, showing a twisted backbone conformation. The third family is similar to the first one except for the orientation of the α-NeuNAc(2 → 3)-β-Gal in the C S-Le^x group. The fourth family exhibits both the structural features of the second and the third family.

The Core2(S-Le^x)$_3$ oligosaccharide should be linked to uromodulin via the anomeric oxygen of the α-GalNAc residue in the core. Unfortunately, the three-dimensional structure of the protein is unknown, as is the shape of the O-glycosylation sites. From the results of the present investigation, one can observe that the anomeric oxygen of the α-GalNAc residue is located in a buried part of the oligosaccharide. As a result, the occurrence of strong interactions between Core2(S-Le^x)$_3$ and the amino-acid at the surface of the protein can be anticipated. Whether the Core2(S-Le^x)$_3$ would keep one of its low energy conformations or would induce some conformational changes of the protein in the surrounding of its O-glycosylation sites are still issues to be considered.

With respect to the biological functions of the Core2(S-Le^x)$_3$ glycan on uromodulin or THP, an important feature has to be noted. If the increased activity of this protein can be related to the presence of these O-glycans at its surface, then the S-Le^x moieties are the most putative candidates for carrying out the molecular interactions that are needed for the activations of the biomolecular processes. For this purpose, the dendrimeric structure of Core2(S-Le^x)$_3$ seems to be designed 'ad hoc' as it provides the maximum capabilities with respect to intermolecular interactions. The

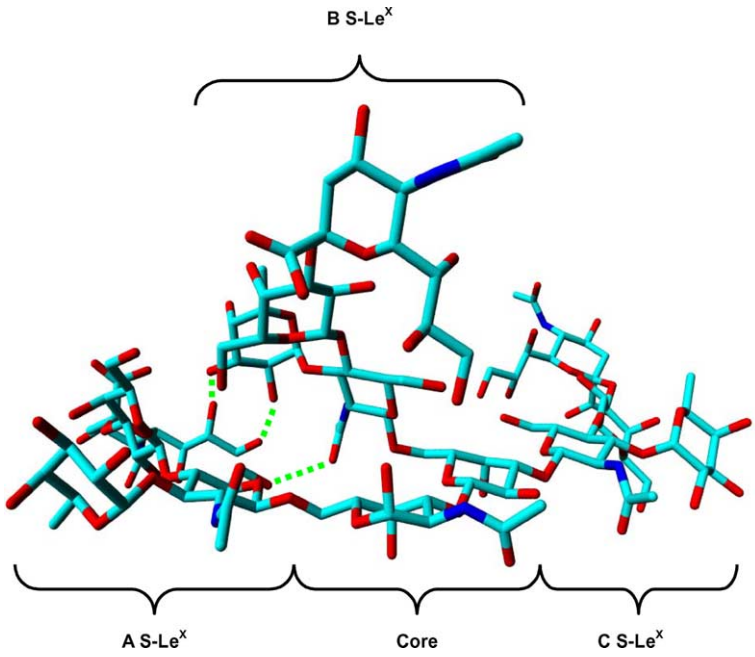

Figure 9. Pictorial three-dimensional representation of the lowest energy conformer for the first family. The hydrogen atoms have been removed and the different parts of the molecule have been framed and labeled. The possible inter-residual hydrogen bond network between A and B S-Le^x groups is suggested with dashed lines.

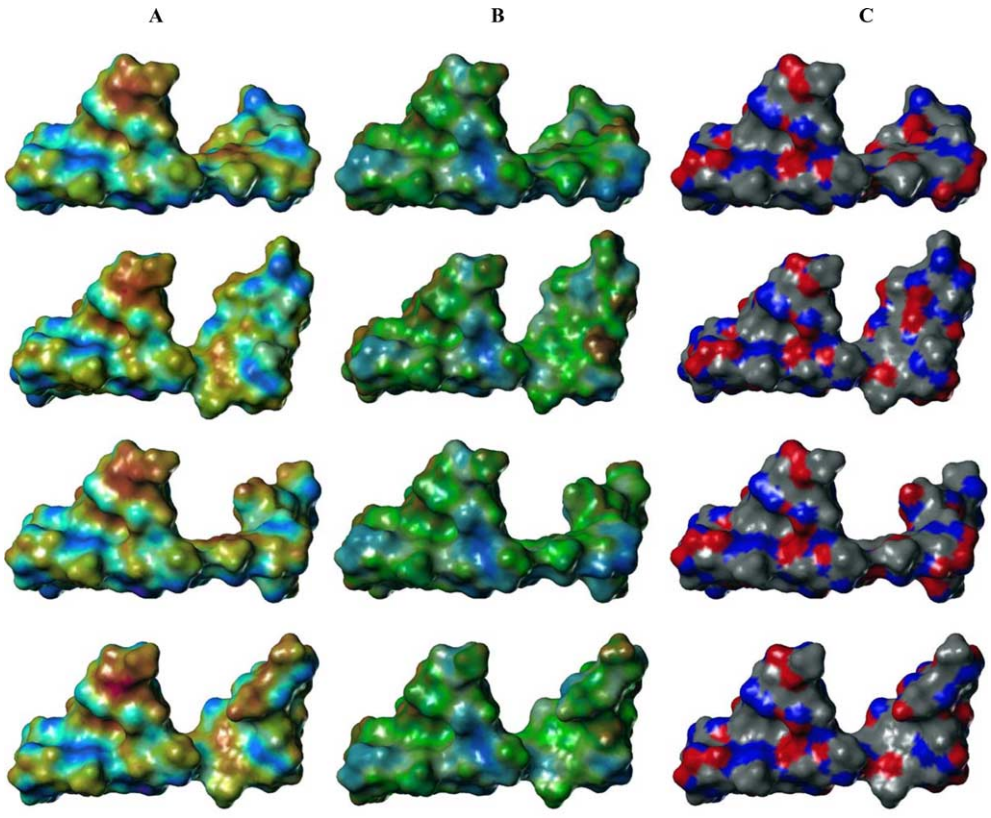

Plate 1. Projection of molecular properties on the Connolly surface for the lowest energy conformer of the four best families (from up to down). (A) Electrostatic potential: colored from blue (negative values) to red (positive values), (B) lipophilicity potential: colored from brown (hydrophobic) to blue (hydrophilic), and (C) hydrogen bonding capacity: blue (acceptor) and red (donor).

conformations of the S-Le^x groups in the lowest energy conformers does not show any differences from those found in previous studies. In particular they are compatible with the binding to E- and P-selectins as derived from crystal structure elucidations.[30]

4. Conclusions

Along with previous knowledge about the conformations of Le^x and S-Le^x oligosaccharides, the present set of information constitutes the data bank required to characterize the structural features and the dynamics of most of the unusual O-linked glycans that may be related to the enhanced immunomodulatory effects of uromodulin. For each disaccharide constituent, precise information about the low energy domains, along with a description of the features displayed by all the local minima has been provided. The construction of the largest O-linked glycan has been completed; this is a tetradecamer made up of the assembly of three S-Le^x moieties on a type 2 core.

The many possible structures have been explored and submitted to an extensive energy minimization using the molecular mechanics program MM3. This was followed by a complete exploration of conformational hyperspace throughout the use of the single coordinate driving method as implemented in the CICADA program. For the whole oligosaccharide, there exists a substantial reduction of the number of viable conformations. The interactions with neighboring groups in complex oligosaccharides can modulate or alter its conformational behavior. Some generalized pictures have emerged from the present investigation: (i) The occurrence of only four conformational families can account for more than 95% of the total conformations; (ii) The whole glycan can be idealized as a backbone formed by the core disaccharide, and the GlcNAc residue of the three S-Le^x moieties. As such, the overall conformation of the glycan can be changed only by rotations about the glycosidic linkages of this backbone. Three of these linkages are conformationally rigid. It is only the β-GlcNAc(1 → 3)-β-GalNac portion of the molecules that offers conformational flexibility as it displays two distinct conformations having a comparable percentage of occurrence; (iii) The backbone of the O-glycan is very flat; it is flanked by three branches having terminal sialyl residues. Two of the branches interact through nonbonded interactions that bring further stabilization, and limit the conformational flexibility of the sialyl residues. Only one branch maintains the original conformational feature of S-Le^x epitope; (iv) The Core2(S-Le^x)$_3$ glycan is a fascinating example of 'dendrimeric' structure, with a spatial arrangement of three S-Le^x epitopes that may favor, throughout their complementary 'presentations', the interactions with E- and P-selectins.

Acknowledgements

This work was supported in part by the 5th Framework Programme, Research Training Networks under the contract GlycoTrain (HPRN-CT-2000-00001), awarded to G. Cioci. J. Koča was supported by a research associate position from the Centre National de la Recherche Scientifique. The Centre Experimental de Calcul Intensif en Chimie (Grenoble) provided the necessary computational resources.

References

1. Tamm, I.; Horsfall, F. L. *Proc. Soc. Exp. Biol. Med.* **1950**, *74*, 108–114.
2. Kumar, S.; Muchmore, A. V. *Kidney Int.* **1990**, *37*, 1395–1401.
3. Pennica, D.; Kohr, W. J.; Kuang, W.; Glaister, D.; Aggarwal, B. B.; Chen, E. Y.; Goedder, D. V. *Science* **1987**, *236*, 83–88.
4. Muchmore, A. V., et al. *Science* **1987**, *237*, 1479–1484.
5. van Rooijen, J. J.; Kamerling, J. P.; Vliegenthart, J. F. *Glycoconj. J.* **2001**, *18*, 539–546; Williams, J.; Marshall, R. D.; van Halbeck, H.; Vliegenthart, J. F. *Carbohydr. Res.* **1984**, *134*, 141–155; Afonso, A. M.; Marshall, R. D. *Biochem. Soc. Trans.* **1979**, *7*, 170–173.
6. Easton, R. L.; Patankar, M. S.; Clark, G. F.; Morris, H. R.; Dell, A. *J. Biol. Chem.* **2000**, *275*, 21928–21938.
7. Mitoma, J.; Petryniak, B.; Hiraoka, N.; Yeh, J.; Lowe, J. B.; Fukuda, M. *J. Biol. Chem.* **2003**, *278*, 9953–9961.
8. Imberty, A.; Pérez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.
9. Woods, R. J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 591–598.
10. Imberty, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 617–623.
11. Allinger, N. L.; Zhu, Z. Q.; Chen, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 6120.
12. Koča, J. *Progr. Biophys. Mol. Biol.* **1998**, *70*, 137–173.
13. Koča, J. *J. Mol. Struct. (Teochem.)* **1994**, *308*, 13–24.
14. Koča, J.; Pérez, S.; Imberty, A. *J. Comput. Chem.* **1995**, *16*, 296–310.
15. Imberty, A.; Mikros, E.; Koča, J.; Mollicone, R.; Oriol, R.; Pérez, S. *Glycoconj. J.* **1995**, *12*, 331–349.
16. Engelsen, S. B.; Pérez, S.; Braccini, I.; Hervé du Penhoat, C.; Koča, J. *Carbohydr. Res.* **1995**, *276*, 1–29.
17. Imberty, A.; Bettler, E.; Karababa, M.; Mazeau, K.; Petrova, P.; Pérez, S. In *Perspectives in Structural Biology*; Vijayan, M., Yathindra, N., Kolashar, A. S., Eds.; Indian Academy of Science (Universities Press), 1999. http://www.cermav.cnrs.fr/.
18. Engelsen, S. B.; Cros, S.; Mackie, W.; Pérez, S. *Biopolymers* **1996**, *39*, 417–433.
19. Pérez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K., et al. *Carbohydr. Res.* **1998**, *314*, 141–155.
20. http://www.fhi-berlin.mpg.de/grz/pub/xfarbe_doc.html.
21. Koča, J. *J. Mol. Struct.* **1993**, *291*, 255–269.
22. Imberty, A.; Pérez, S. *Glycobiology* **1994**, *4*, 351–366.
23. Ichikawa, Y.; Lin, Y. C.; Dumas, D. P.; Shen, G. J., et al. *J. Am. Chem. Soc.* **1992**, *114*, 9283–9298.
24. Thogersen, H.; Lemieux, R. U.; Bock, K.; Meyer, B. *Can. J. Chem.* **1982**, *60*, 44–57.
25. Miller, K. E.; Mukhopadhyay, C.; Cagas, P.; Bush, C. A. *Biochemistry* **1992**, *31*, 6703–6709.

26. Pérez, S.; Mouhous-Riou, N.; Nifant'ev, N. E.; Tsvetkov, Y. E.; Bachet, B.; Imberty, A. *Glycobiology* **1996**, *6*, 537–542.
27. Imberty, A.; Pérez, S. In *Carbohydrates Mimics. Concept and Methods*; Chapleur, Y., Ed.; Wiley: Weinheim, 1997; pp 349–364.
28. Imberty, A.; Gerber, S.; Tran, V.; Pérez, S. *Glycoconj. J.* **1990**, *7*, 37–54.
29. Imberty, A.; Delage, M.; Bourne, Y.; Cambillau, C.; Pérez, S. *Glycoconj. J.* **1991**, *8*, 456–483.
30. Somers, W. S.; Tang, J.; Shaw, G. D.; Camphausen, R. T. *Cell* **2000**, *103*, 467–479.