Potential Energy Hypersurfaces of Nucleotide Sugars: Ab Initio Calculations, Force-Field Parametrization, and Exploration of the Flexibility

Pavla Petrová,[†,‡] Jaroslav Koča,[†] and Anne Imberty*,[‡]

Contribution from the Laboratory of Biomolecular Structure and Dynamics and Department of Organic Chemistry, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic, and Centre de Recherches sur les Macromolécules Végétales-CNRS (affiliated with Université Joseph Fourier), BP 53, F-38041 Grenoble Cedex 9, France

Received November 6, 1998. Revised Manuscript Received February 19, 1999

Abstract: Glycosyl esters of nucleoside di- or monophosphates, generally referred to as "sugar nucleotides", serve as a sugar donor during the biosynthesis of oligo- and polysaccharides. Therefore, they are of primary importance in carbohydrate metabolism in the living world. Not only the molecules themselves but especially their complexes with proteins are of interest in structural glycobiology. For computational studies on these molecules, it is necessary to have access to empirical methods with appropriate force field parametrization. In this work, we propose a set of parameters, developed using ab initio calculations with the 6-31G* basis set at the SCF level on model compounds, for the commonly used AMBER force field. By implementation of the new parameter set together with the CICADA conformational search program, we have obtained a semiquantitative description of conformational space, showing that nucleotide sugars can adopt several conformational families. The majority of them exhibit a "folded" rather than an "extended" geometry due to frequent intramolecular hydrogen bonds and "stacking" interactions between the base and the six-membered sugar ring. For the sake of comparison, two molecular dynamics simulations were run in an explicit water environment. The first simulation (3 ns) started with the semi-extended X-ray geometry and displayed major variations of all torsional angles, allowing for the visit of three conformational families. The second simulation (5 ns) started with the folded global minimum from the CICADA search. After about 3ns, a transition for the ribose pucker yielded to the visit of a more extended conformational family. Experimental results show that in crystalline state, or in protein/carbohydrate complexes, extended conformations which are stabilized by the interaction with surrounding molecules or with the protein surface are more frequent.

Introduction

Glycosyl esters of nucleoside di- or monophosphates, generally referred to as "sugar nucleotides", are of primary importance in carbohydrate metabolism.[1−4] These activated sugars serve as the donor during the biosynthesis of oligosaccharides, polysaccharides, and glycoconjugates in the living world. When considering the biosynthetic pathway of plant and bacterial polysaccharides and glycoconjugates from all origins, more than 35 nucleotide sugars were identified.[5] These include glycosyl esters of uridine 5′-diphosphate (UDP), thymidine 5′-diphosphate (dTDP), guanosine 5′-diphosphate (GDP), cytidine 5′-diphosphate (CDP), cytidine 5′-monophosphate (CMP), and adenosine 5′-diphosphate (ADP). The most important nucleotide sugars involved in the biosynthetic pathway of mammalian *N*-glycans are UDP-glucose (UDP-Glc), UDP-*N*-acetylglu-cosamine (UDP-GlcNAc), GDP-mannose (GDP-Man), GDP-fucose (GDP-Fuc), and CPM-sialic acid (CMP-NeuAc).[6,7]

Since the development of chemo-enzymatic methods in the field of carbohydrate synthesis,[8] nucleotide sugars have become substrates of great pharmaceutical interest. It is of paramount importance to understand the interaction of these molecules with the enzymes involved in carbohydrate synthesis: epimerases,[9] transporters,[10] and glysosyltransferases.[11] Very little is known about the three-dimensional structure and conformational behavior of nucleotide sugars.[3] Only one crystal structure, the sodium salt of UDP-Glc, is available.[12] It appears therefore necessary to develop new tools for investigating the possible conformations of sugar nucleotides.

Conformational behavior of biomolecules such as proteins, DNA, and carbohydrates has been the focus of attention in biophysics and biochemistry for a long time. If some structural

* Author for correspondence. Postal address: CERMAV-CNRS, BP53, F-38041 Grenoble Cedex 9, France. Street address: 601 Rue de la Chimie, Domaine Universitaire, F-38041 Grenoble Cedex 9, France. Telephone: +33 476 03 76 36. Fax: +33 476 54 72 03. Email: imberty@cermav.cnrs.fr.

(1) Neufeld, E. F.; Hassid, W. Z. *Adv. Carbohydr. Chem. Biochem.* **1963**, *18*, 309.
(2) Ginzburg, V. *Adv. Enzymol.* **1964**, *26*, 35.
(3) Kochetkov, N. K.; Shibaev, V. N. *Adv. Carbohydr. Chem. Biochem.* **1973**, *28*, 307−399.
(4) Shibaev, V. N. *Adv. Carbohydr. Chem. Biochem.* **1986**, *44*, 277.
(5) Gabriel O. *Methods Enzymol.* **1982**, *83*, 332−353.

(6) Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631−634.
(7) Paulson, J. C.; Colley, K. J. *J. Biol. Chem.* **1989**, *264*, 17615−17618.
(8) Ichikawa, Y.; Look, G. C.; Wong C.-H. *Anal. Biochem.* **1992**, *202*, 215−238.
(9) Thoden, J. B.; Frey, P. A.; Holden, H. M. *Biochemistry* **1996**, *35*, 5137−5144.
(10) Kawakita, M.; Ishida, N.; Miura, N.; Sun-Wada, G.-H.; Yoshioka, S. *J. Biochem.* **1998**, *123*, 777−785.
(11) Kleene, R.; Berger, E. G. *Biochim. Biophys. Acta* **1993**, *1154*, 283−325.
(12) Sugawara, Y.; Iwasaki, H. *Acta Crystallogr.* **1984**, *C40*, 389−393.

properties can be studied by experimental means, theoretical calculation is the tool of choice for exploring conformational behavior. Molecular mechanics has developed into a standard powerful method for studying molecular structure and related properties. The empirical force fields must be accurately parametrized for each species of interest. The classical protein force fields such as AMBER,[13] CHARM22,[14] GROMOS,[15] and OPLS[16] are often parametrized for nucleic acids. For all of these programs, additional parameters have been developed for carbohydrates. However, these force fields do not generally include parameters for a glycosidic linkage between a carbohydrate and another species such as glycopeptide, nucleotide sugar, or sulfated and phosphated carbohydrates. Before performing any conformational analysis of a sugar nucleotide or other phosphated carbohydrate it is necessary to develop the force field parameters.

The AMBER force field[13] was selected as a framework for the present parametrization since (1) recent carbohydrate parameters are available in the GLYCAM-93 development,[17] (2) it has been interfaced with the CICADA program [18,19] for exploring conformational space of flexible biomolecules, (3) it allows for molecular dynamics runs in explicit water and for comparison with NMR data, and (4) it has been extensively employed for proteins and could be then used for docking studies of nucleotide sugars with protein receptors. The first step of the parametrization procedure has been the ab initio study of model compounds **3** and **4** (Scheme 1). Compound **3** is a model for monosaccharides with $^4C_1$ ring shape and axial phosphate substituent, such as UDP-Glc, UDP-Gal and GDP-Man, whereas compound **4** is a model for a $^1C_4$ ring shape with an equatorial phosphate such as GDP-Fuc. The newly parametrized AMBER is then presented along with the exploration of the potential energy hypersurface of two nucleotide sugars UDP-Glc and GDP-Fuc. The CICADA program, a single coordinate driving method, that has already been shown to be very powerful in determining all the possible conformations of several dozen small and middle-sized organic molecules including pheromones,[20] peptides,[21–23] oligonucleotides,[19,24] and oligosaccharides[25–27] has been used in the present work. The resulting data represent the first conformational analysis of nucleotide sugars.

Ab Initio Study of Phosphorylated Carbohydrate Analogues. Molecules **3** and **4** are model compounds of α-D-glucose hydrogenophosphate and β-L-fucose hydrogenophosphate, respectively. The ab initio calculations of both compounds have

been conducted at the RHF/6-31G* level as implemented in the Gaussian 94 program package.[28] Partial atomic charges have been calculated separately for the ring and hydrogenophosphate group. To be consistent with the GLYCAM-93 parametrization, the ring charges were calculated using the CHelpG approach.[29] For charged groups, the RESP method[30] that is implemented in AMBER-4.1 has been used. All charge derivations were performed after optimizing the geometries at the RHF/6-31G* level.

The energy profile investigation was performed by varying the Φ torsion angles by steps of 30° and relaxing all other internal coordinates. The option of additional redundant internal coordinates was used to keep the Φ torsion angle at the given value during the energy minimization. For each step, the three staggered orientations of the C1−O1−P−OH dihedral angle were tested, and only the lowest energy one has been considered. In addition the hydroxylic hydrogen, i.e., the O1−P−O−H dihedral angle, has been given a trans orientation.

The resulting energy profile curves are given in Figure 1. For the axial phosphate group on the $^4C_1$ ring, two minima occur at Φ = 60° and Φ = 150°. The energy profile between these minima is a plateau with a very low energy barrier (ΔE = 0.53 kcal/mol). The other gauche conformer (Φ = −60°) exhibits a much higher energy and a very shallow energy well. The Φ energy curve of axial 2-methoxytetrahydropyran, a model compound of O-methyl α-D-glucose, presents a different profile,[31] with one main energy minimum at Φ = 60°. The occurrence of the second minimum (Φ = 150°) in the present study can be related to a stabilizing effect between the OH of the phosphate group and the ring. In fact, the two minima do not have the same preferred orientation for the second torsion angle Ψ: for this angle, the lowest energy is obtained for a trans orientation when Φ is in the range 0−90° and a gauche+ orientation otherwise that brings the OH group close enough for stabilization.

The equatorial phosphate at the anomeric position of the $^1C_4$ ring exhibits a rather different energy profile with one main energy minimum at Φ = 120° and a weak shoulder at Φ = −60°. The preferred orientation of the OH group is trans for all values of Φ. This conformational behavior is very different from equatorial 2-methoxytetrahydropyran[31] that presents an absolute minimum at Φ = 60° (when in the same $^1C_4$ ring shape), a shoulder at Φ = 180° and a secondary minimum (ΔE about 5 kcal/mol) at Φ = −60°.

The bond lengths and valence angles at the anomeric center are internal parameters that show a dependence on the value of the Φ torsion angle. Variations of O5−C1 and C1−O1 distances, along with O5−C1−O1 and C1−O1−P valence angles as a function of Φ are depicted in Figure 2. As observed for 2-methoxytetrahydropyran,[31] the C1−O1 distance is maximum for trans and cis orientation of Φ and is shorter for the equatorial compound than for the axial one. The O5−C1 distance also varies with the Φ torsion angles but in an opposite way. For both axial and equatorial hydrogenphosphate, the O5−C1−O1 valence angle has a minimum value for a trans

(13) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Chio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−764.

(14) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(15) Koehler, J. E. H.; Saenger, W.; van Gunsteren W. F. *Eur. Biophys. J.* **1987**, *15*, 97.

(16) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1995**, *118*, 11225.

(17) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832−3846.

(18) Koča, J. *J. Mol. Struct.* **1994**, *308*, 13−24.

(19) Fadrna, E.; Koča, J. *J. Biomol. Struct. Dynam.* **1996**, *14*, 137.

(20) Koča, J.; Carlsen, P. H. J. *J. Mol. Struct.* **1992**, *268*, 263−281.

(21) Koča, J.; Carlsen, P. H. J. *J. Mol. Struct.* **1993**, *291*, 271−286.

(22) Koča, J.; Carlsen, P. H. J. *J. Mol. Struct.* **1995**, *337*, 17−24.

(23) Koča, J.; Kriz, Z.; Carlsen, P. H. J. *J. Mol. Struct.* **1994**, *306*, 157−164.

(24) Fadrna, E.; Koča, J. *J. Phys. Chem.* **1997**, *101*, 7863−7868.

(25) Koča, J.; Pérez, S.; Imberty, A. *J. Comput. Chem.* **1995**, *16*, 796−310.

(26) Engelsen, S. B.; Pérez, S.; Braccini, I.; Hervé du Penhoat, C.; Koča, J. *Carbohydr. Res.* **1995**, *276*, 1−29.

(27) Imberty, A.; Mikros, E.; Koča, J.; Mollicone, R.; Oriol, R.; Pérez, S. *Glycoconjugate J.* **1995**, *12*, 331−349.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian-94*, Gaussian Inc., Pittsburgh, PA, 1995.

(29) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(30) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *11*, 1357−1377.

(31) Tvaroska, I.; Carver, J. P. *J. Phys. Chem.* **1994**, *98*, 9477−9485.

Scheme 1

1

2

3

4

5
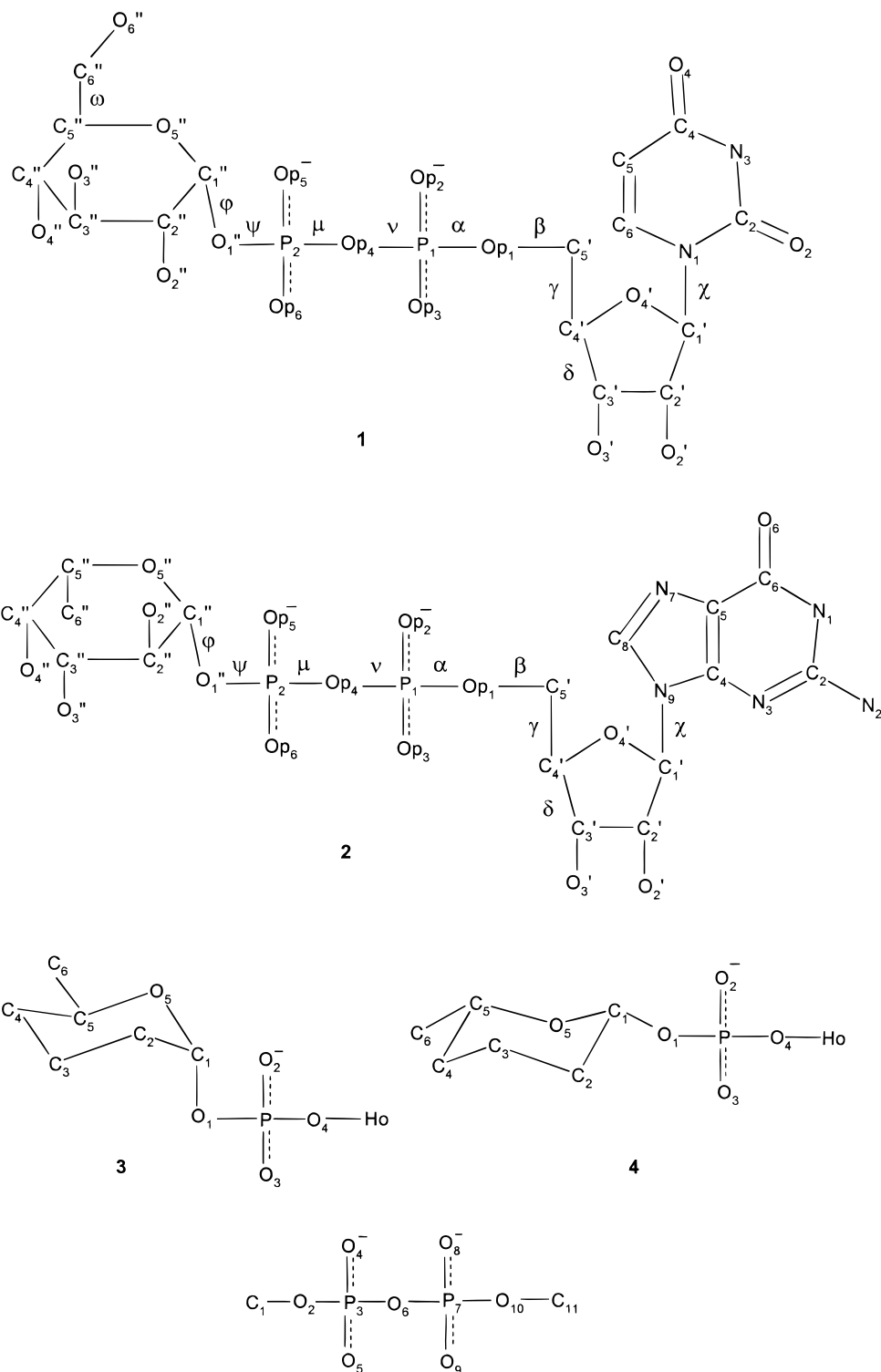
orientation of Φ, whereas the C1−O1−P valence angle varies with the energy curve: the minimum values occur at Φ = 60° for molecule **3** and Φ = 120−150° for molecule **4**. This dependence of the valence angles upon the Φ torsion is identical to what has been calculated for 2-methoxytetrahydropyran.[31]

The conformers hyporresponding to energy minima on the Φ energy curve (Φ = 60° and 150° for the axial compound and Φ = 120° for the equatorial one) have been fully optimized at the RHF/6-31G* level. The resulting energies and geometries are given in Table 1. Upon relaxation, the two minima of the axial compounds adopt Φ angles of 68° and 161°, that correspond to gauche−trans and trans−gauche orientation when referring to O5 then C2. Both conformers have equivalent ab initio energy (ΔE = 0.11 kcal/mol in favor of the TG conformer). In the crystal state, values of 75.7° and 63.8° are observed for the Φ angle of Glc-1HPO3 and UDP-Glc, respectively. The higher energy of the equatorial compound (ΔE = 2.29 kcal/mol) is due to the anomeric effect.[32−34]
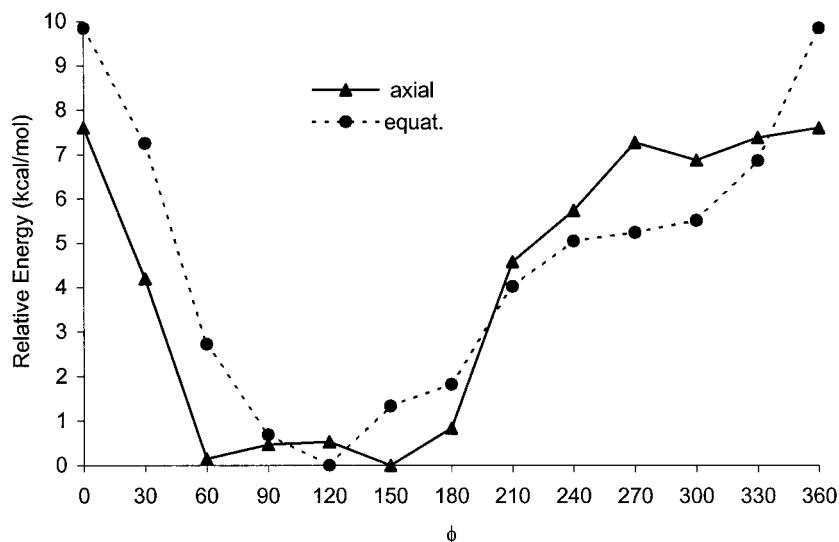
Figure 1. Ab initio 6-31G* rotational potential energy for the Φ torsion angle of compounds **3** and **4**.
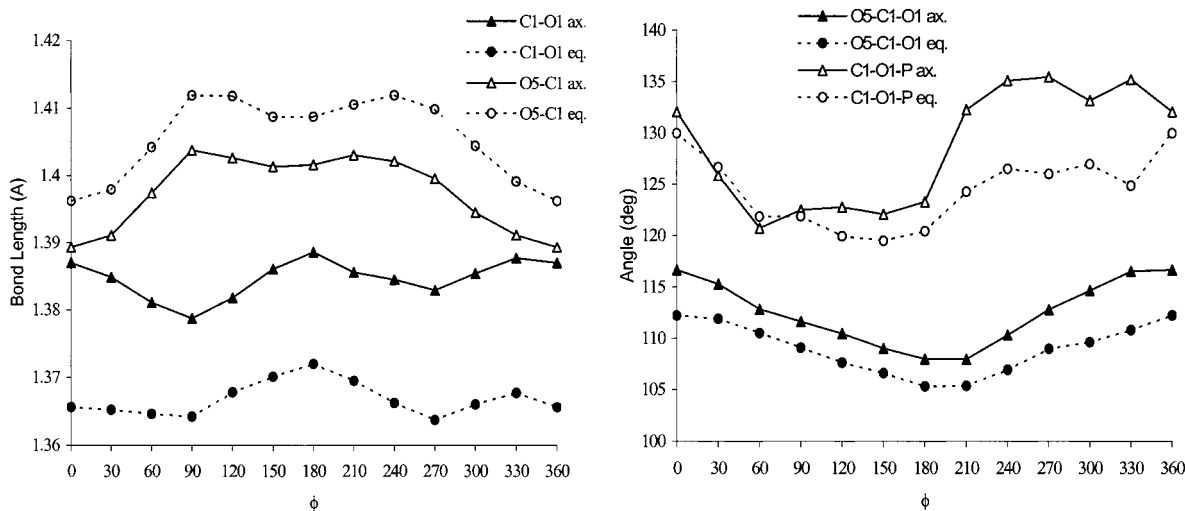

Figure 2. Ab initio 6-31G* variations of bond lengths and valence angle at the anomeric center for compounds **3** and **4**.

Table 1. Ab Initio 6-31G* Calculated Energies and Geometries of the Energy Minima for Compounds **3** and **4**

starting conformer	ϕ (°)	6-31G* energy (hartree)	relative energy (kcal/mol)
ax-$^4C_1$ ($\phi = 60°$)	68.3	−949.3839753	0.11
ax-$^4C_1$ ($\phi = 150°$)	161.3	−949.3841462	0.00
eq-$^1C_4$ ($\phi = 120°$)	141.5	−949.3806567	2.29

The geometrical characteristics at the anomeric center are depicted in Figure 3. Marked differences between the two anomers are the shorter C1−O1 distance and smaller O5−C1−O1 angle for the equatorial compound. This has been shown to be a structural consequence of the anomeric effect[35] and is now well-documented.[34] The ab initio geometry of the axial phosphated compound can be compared with mean values from crystal structure of glucose derivatives (HPO3-Glc and UDP-Glc). With the exception of the O1−P bond, the ab initio distances are smaller than the experimental ones. This discrepancy has already been observed in theoretical studies on carbohydrates[31] and has been attributed to two particular effects of the crystalline state: the influence of packing on geometries and the difference between computed distances (r_e) that correspond to equilibrium values and measured distances (r_g) that correspond to internuclear distances averaged over molecular vibrations.[36] Nevertheless, the difference between theoretical and experimental C1−O1 distances for the axial phosphate is larger than the average.

Comparison can also be done with theoretical[31] and experimental[37] data for axial or equatorial OCH$_3$ groups. The only noticeable difference is the largest C1−O1−P valence angle, when compared to the C1−O1−C one, that may be due to the size of the phosphate group.

AMBER Parametrization for Nucleotide Sugars. Strategy. The procedure proposed by Hopfinger and Pearlstein[38] has been adopted for AMBER parametrization. Using selected model compounds, atomic partial charges were derived from ab inito calculations through electrostatic potential fitting. Equilibrium values for bond lengths and valence angles were selected to reproduce the mean values observed in crystals, when available. After all of the parameters have been incorporated in the force

(32) Edward, J. T. *Chem. Ind. (London)* **1955**, 1102−1104.

(33) Lemieux, R. U. *Molecular Rearrangements*; Interscience: New York, 1964; pp 709−769.

(34) Tvaroska, I.; Bleha, T. *Adv. Carbohydr. Chem. Biochem.* **1989**, *47*, 45−123.

(35) Berman, H. M.; Chu, S. S. C.; Jeffrey, G. A. *Science* **1967**, *157*, 1576−1577.

(36) Hargittai, M.; Hargittai, I. *Int. J. Quantum Chem.* **1992**, *44*, 1057−1067.

(37) Jeffrey, G. A.; Taylor, R. *J. Comput. Chem.* **1980**, *1*, 99−109.

(38) Hopfinger, A. J.; Pearlstein, R. A. *J. Comput. Chem.* **1984**, *5*, 486.
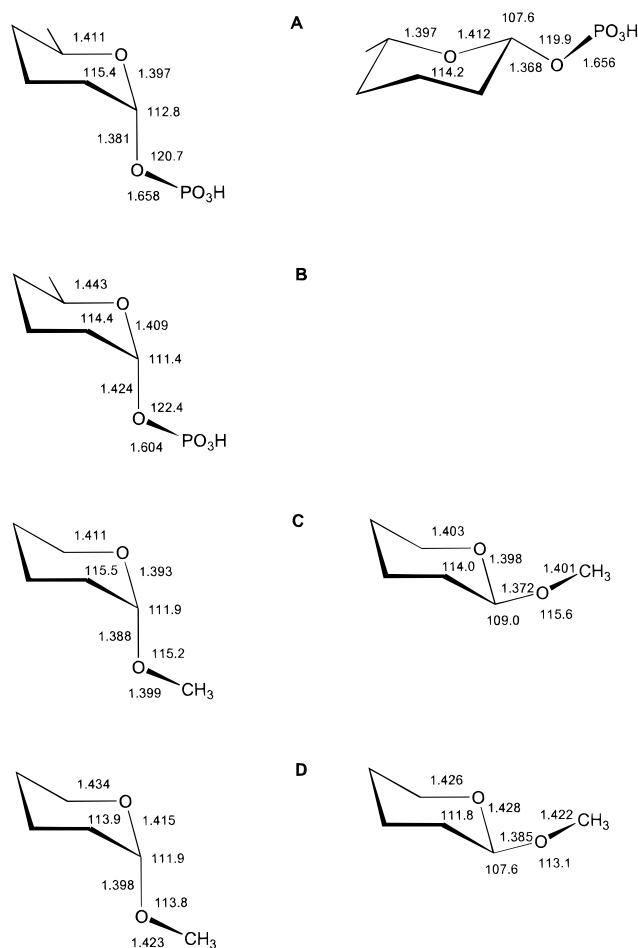
Figure 3. Comparison of geometries at the anomeric centers for (a) ab initio calculations on compound **3** and **4,** (b) mean values for crystal structures of UDP-Glc and HPO3-Glc, (c) ab initio calculations on 2-methoxytetrahydropyran,[31] and (d) mean values for crystal structures of carbohydrates.[37]

field, torsional parameters were finally derived by first calculating molecular mechanics torsional energy profiles with the parameter to be optimized set to zero and then adjusting the parameters to best reproduce the difference function between this energy profile curve and the one from ab initio calculations above. During the parametrization procedure, the dielectric constant corresponded to vacuum conditions in order to be in agreement with ab initio gas-phase calculations.

The parametrization presented here for the AMBER force field[13] is an extension of GLYCAM-93 that was proposed recently for carbohydrates.[17] No new atom types are introduced in the present parametrization. The atom types AC (axial anomeric carbon), EC (equatorial anomeric carbon), and OG (glycosidic oxygen) that were introduced in GLYCAM have been kept, together with the phosphorus atom (P) present in native AMBER.

Partial Atomic Charges. For deriving partial charges, the nucleotide sugars were considered as a combination of several fragments. For the nucleotide moiety, the partial charges were taken from native AMBER. For the carbohydrate moiety, the charges were calculated by using the CHELPG procedure[29] with 6-31G* basis set for neutral fragments, in agreement with the GLYCAM-93 parametrization.[17] For the pyrophosphate the RESP procedure was used together with the hyperbolic restraints described in the literature.[30] The partial charges for UDP-Glc and GDP-Fuc are listed in Table 2.

Parametrization of the Phosphate Group in the Axial Position. Several crystal structures of phosphated monosaccharides are available in the Cambridge Structural Database (http://www.ccdc.cam.ac.uk/index.html). For the present study, only the structures with one negative charge on the phosphate group at the anomeric position were retained: i.e., glucose hydrogenophosphate ($1K^+$)[39] and UDP-glucose ($2Na^+$, $2H_2O$).[12] Mean observed bond lengths and valence angles at the anomeric position are given in Figure 3. The equilibrium values proposed in native AMBER for equivalent atom types allowed for good agreement with the mean values and were therefore retained in the present parametrization. The stretching and bending force constants were kept identical to the ones proposed for equivalent bonds and angles in the AMBER/GLYCAM parametrization. The parametrization of the Φ torsion angle was performed as described above. To fit the difference curve to the ab initio energy profile curve of model compound **3**, the parameters of the three torsions CT−AC−OG−P, OS−AC−OG−P and H2−AC−OG−P were adjusted. The resulting nine V_n parameters are listed in Table 3. All of the phase parameters have been kept to zero. The comparison between the ab initio and newly parametrized force-field calculations for the Φ torsion energy profile curve of model compound **3** is displayed in Figure 4.

Parametrization of the Phosphate Group in the Equatorial Position. The same procedure as that above was used for deriving the charges of GDP-fucose. Since there is no crystal structure available for the equatorial carbohydrate−phosphate linkage, the native AMBER equilibrium values for bond stretching and angle bending were retained. The torsional parameters for the C1−O1 bond were derived using the ab initio energy profile curve calculated for model compound **4**. The resulting Φ torsion energy profile curve is displayed in Figure 4.

Parametrization of Pyrophosphate. Since pyrophosphate is a fragment of small molecules of biological importance such as coenzymes (NAD^+ and NADH) and nucleotides (ADP and ATP), it has recently been the subject of a parametrization study for the CHARMM22 force field.[40] These authors calculated the mean bond lengths P−OS and P=O2 in inorganic diphosphate crystal to be 1.613 and 1.499 Å, respectively, and the mean valence angle P−OS−P to be 131.5°. These values are in good agreement with those measured for the crystal structure of UDP-Glc (the mean bond length for P−OS is 1.60 Å and for P=O2 it is 1.486 Å and the valence angle P−OS−P is 132.6°). The native AMBER equilibrium values for P−OS (1.61 Å) and P=O2 (1.48 Å) have been retained for the present parametrization, but the P−OS−P angle bending equilibrium value has been set to 130°. The force constants were kept unmodified. The torsion around the P−OS bond has been parametrized. The ab initio energy profile curve was calculated by using RHF/6-31+G* on the dimethyldiphosphate (model compound **5**). The new angle bending and torsional parameters are listed in Table 3.

Validation of the Parametrization. The two molecules with known crystal structures, i.e., glucose-1-hydrogenphosphate (Glc-1PO3H) and UDP-glucose have been submitted to full optimization with the newly parametrized AMBER force-fields. The overall root-mean-squares (rms) calculated for all non-hydrogen atoms are 0.197 and 0.178 Å for Glc-1HPO3 and UDP-Glc, respectively. Comparison of experimental and calculated internal parameters for the glycosidic junction and

(39) Lis, T. *Carbohydr. Res.* **1992**, *229*, 33−39.
(40) Pavelites, J. J.; Gao, J.; Bash, A. B.; Mackerell A. D., Jr. *J. Comput. Chem.* **1997**, *18*, 221−239.

Table 2. Atom Type and Atomic Charges for Nucleotide Sugars

UDP-glucose						GDP-fucose					
atom	type	charge	atom	type	charge	atom	type	charge	atom	type	charge
O1″	OG	−0.4763	O2″	OH	−0.6139	O1″	OG	−0.5458	O2′	OH	−0.6139
O2″	OH	−0.67	O3′	OH	−0.6541	O2″	OH	−0.7071	O3′	OH	−0.6541
O3″	OH	−0.7317	O4′	OS	−0.3548	O3″	OH	−0.6847	O4′	OS	−0.3548
O4″	OH	−0.7214	C1′	CT	0.0674	O4″	OH	−0.6642	C1′	CT	0.0191
O5″	OS	−0.5674	C2′	CT	0.067	O5″	OL	−0.5237	C2′	CT	0.067
O6″	OH	−0.7493	C3′	CT	0.2022				C3′	CT	0.2022
C1″	AC	0.2005	C4′	CT	0.1065	C1″	EC	0.4796	C4′	CT	0.1065
C2″	CT	0.1873	C5′	CT	0.0558	C2″	CT	0.1895	C5′	CT	0.0558
C3″	CT	0.469	H1′	H2	0.1824	C3″	CT	0.2766	H1′	H2	0.2006
C4″	CT	−0.0903	H2′	H1	0.0972	C4″	CT	0.1648	H2′	H1	0.0972
C5″	CT	0.4055	H3′	H1	0.0615	C5″	CT	0.4044	H3′	H1	0.0615
C6″	CT	0.3465	H4′	H1	0.1174	C6″	CT	−0.2959	H4′	H1	0.1174
H1″	H2	0.1274	HO′2	HO	0.4186	H1″	H2	0.0381	HO′2	HO	0.4186
H2″	H1	0.1056	HO′3	HO	0.4376	H2″	H1	0.0885	HO′3	HO	0.4376
H3″	H1	−0.0218	H5′1	H1	0.0679	H3″	H1	−0.0334	H5′1	H1	0.0679
H4″	H1	0.0767	H5′2	H1	0.0679	H4″	H1	0.0248	H5′2	H1	0.0679
H5″	H1	0.0542	O2	O	−0.5477	H5″	H1	−0.0556	O6	O	−0.5597
H2O″	HO	0.4114	O4	O	−0.5761	H2″O	HO	0.4474	N1	NA	−0.4787
H3O″	HO	0.4295	N1	N*	0.0418	H3″O	HO	0.4224	N2	N2	−0.9672
H4O″	HO	0.4735	N3	NA	−0.3549	H4″O	HO	0.4399	N3	NC	−0.6323
H6O″	HO	0.4491	C2	C	0.4687	H1M	HC	0.0524	N7	NB	−0.5709
H61	H1	−0.0478	C4	C	0.5952	H2M	HC	0.0832	N9	N*	0.0492
H62	H1	0.0435	C5	CM	−0.3635	H3M	HC	0.1025	C2	CA	0.7657
P1	P	1.1128	C6	CM	−0.1126	P1	P	1.1128	C4	CB	0.1222
P2	P	1.0903	H3	H	0.3154	P2	P	1.0903	C5	CB	0.1744
Op1	OS	−0.4974	H5	HA	0.1811	Op1	OS	−0.4974	C6	C	0.477
Op2	O2	−0.7967	H6	H4	0.2188	Op2	O2	−0.7967	C8	CK	0.1374
Op3	O2	−0.7967				Op3	O2	−0.7967	H1	H	0.3424
Op4	OS	−0.4233				Op4	OS	−0.4233	H21	H	0.4364
Op5	O2	−0.7928				Op5	O2	−0.7928	H22	H	0.4364
Op6	O2	−0.7928				Op6	O2	−0.7928	H8	H5	0.164

pyrophosphate group is given is Table 4. The new force field reproduces well the anomeric geometries, and no significant distortion from experimental data is observed.

Conformational Analysis of Nucleotide Sugars

Strategy. The conformational search has been performed using the so-called single coordinate driving (SCD) method[41,42] as implemented in the program CICADA.[18] The search begins with the lowest-energy conformer available. One of the dihedral angles available is then selected and rotated by a predefined step. The new value of the dihedral angle is constrained, and all of the remaining internal coordinates are relaxed under subsequent energy minimization. The process is repeated until the new point is lower in energy than the previous one. In such a case, an approximation of a transition state has been encountered, and the point is recorded. The downhill traveling is continued until the energy increases again. The last point before the energy has increased is then fully relaxed, and the true energy minimum is obtained and again recorded. After the process is repeated for all of the selected dihedral angles and for all of the energy minima within a predefined energy window, a topological description of the energy surface containing minimum−transition state−minimum triads is obtained. The resulting data allows for flexibility calculation.[43,44] Conformational flexibility is calculated as a quantity proportional to the number of low-energy conformers which the molecule can adopt and to the height of energy barriers separating the conformers. Also the extent of movement within the energy well (size of the well) is considered for each energy minimum. The molecule is considered to be very flexible when many low-energy conformations are adopted, energy barriers separating the conformations are low, and the freedom of conformational movement within energy wells is large. The quantification of an average conformational movement can be calculated[44] and will be referred to as relative flexibility.

For conformational analysis of both nucleotide sugars, UDP-Glc and GDP-Fuc, two starting geometries were built. For UDP-Glc, the first conformer was taken directly from the crystal structure.[12] The same structure was used to build the first conformer of GDP-Fuc by substituting the glucose residue with a fucose one taken from MONOBANK,[45] a database of monosaccharides, and substituting UDP by GDP taken from the Sybyl[46] fragment libraries. For both compounds, a second starting conformer was generated by modifying the ribose pucker from C2′-endo to C3′-endo. All of the building and editing of molecular models was done within the Sybyl software.[46]

The potential energy hypersurfaces were investigated by driving all of the torsional angles affecting the shape of the nucleotide sugar: Φ, Ψ, μ, v, α, β, and γ for the backbone between the monosaccharide and the ribose ring, δ for the pucker of ribose, and χ for the orientation of the base. In addition, the ω of hydroxymethyl group of glucose was driven. This leads to a 10-dimension hypersurface for UDP-Glc and a 9-dimension one for GDP-Fuc. A 15° step was considered during the exploration, except for the δ endocyclic torsion for which a 6° increment was selected. The torsion angles governing the orientation of the hydroxyl groups of the monosaccharide and of the ribose were monitored but not driven. Two conformations were considered different if at least one of their

(41) Fadrna, E.; Koča, J. *J. Mol. Struct.* **1997**, *398−399*, 523.
(42) Koča, J. *Progr. Biophys. Mol. Biol.* **1998**, *70*, 137−173.
(43) Koča, J.; Carlsen, P. H. J. *J. Mol. Struct.* **1992**, *257*, 105.
(44) Koča, J. *J. Mol. Struct.* **1993**, *291*, 255−269.

(45) Pérez, S.; Delage, M. M. *Carbohydr. Res.* **1991**, *212*, 253−259.
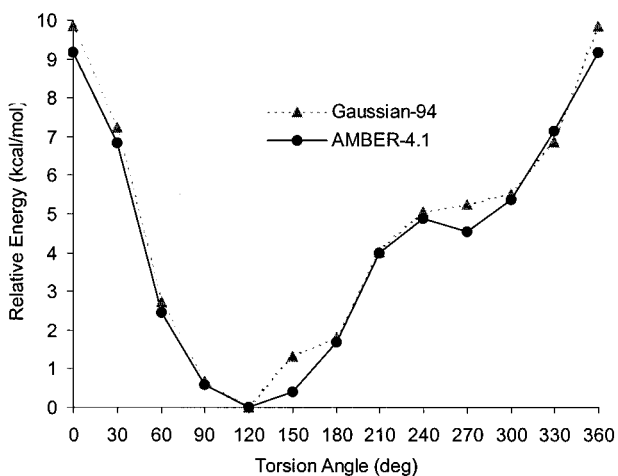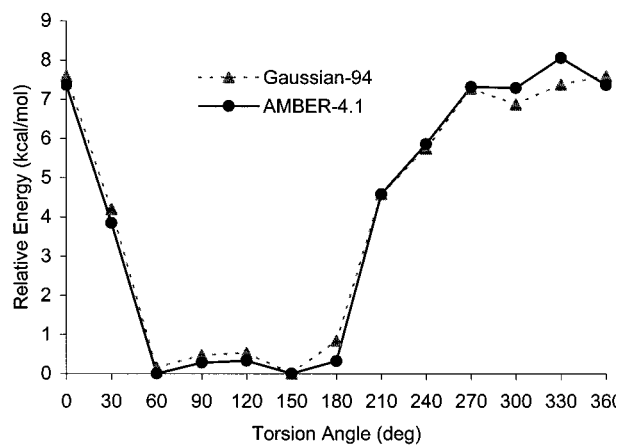(46) Sybyl 6.4, Tripos Associates, 1699 S. Hanley Road, Suite 303, St Louis, MO 63144.

Table 3. AMBER Force Field Parameters for Nucleotide Sugars

Stretching

	K_r	r_{eq}	source
AC−OG	320.0	1.420	present work
OG−P	230.0	1.610	P94 OS−P

Bending

	K_θ	θ_{eq}	source
AC−OG−P	100.0	120.50	P94 CT−OS−P
EC−OG−P	100.0	120.50	P94 CT−OS−P
OG−P−OH	45.0	102.60	P94 OS−P−OH
OG−P−O2	100.0	108.23	P94 OS−P−O2
OG−P−OS	45.0	102.60	P94 OS−P−OS
P−OS−P	100.0	130.00	present work

Torsion

	IDIVF	$PK(Vn/2)$	phase	PN	source
AC−OG−P−OH	1	0.25	0.00	−3	P94 CT-OS−P−OH
	1	1.20	0.00	2	
AC−OG−P−OS	1	0.25	0.00	−3	P94 CT−OS−P−OS
	1	1.20	0.00	2	
X−P−OG−X	3	0.75	0.00	3	P94 X−P−OS−X
CT−AC−OG−P	1	0.12	0.00	−3	
	1	−0.62	0.00	−2	
	1	−0.78	0.00	1	
OS−AC−OG−P	1	0.12	0.00	−3	present work
	1	0.09	0.00	−2	
	1	−1.00	0.00	1	
H2−AC−OG−P	1	0.12	0.00	−3	present work
	1	−0.08	0.00	−2	
	1	0.30	0.00	1	
CT−EC−OG−P	1	0.15	0.00	−3	present work
	1	−0.46	0.00	−2	
	1	−0.07	0.00	1	
OL−EC−OG−P	1	0.15	0.00	−3	present work
	1	0.73	0.00	−2	
	1	0.06	0.00	1	
H2−EC−OG−P	1	0.15	0.00	−3	
	1	−0.41	0.00	−2	
	1	−0.13	0.00	1	
P−OS−P−OG	1	−0.33	1.37	−3	present work
	1	−0.20	−10.65	−2	
	1	−0.12	137.89	1	
P−OS−P−OS	1	−0.33	1.37	−3	present work
	1	−0.20	−10.65	−2	
	1	−0.12	137.89	1	

Figure 4. Comparison of the ab initio and newly parametrized AMBER energy profile curves for axial and equatorial C1−O1 bond (compounds **3** and **4**).

Table 4. Comparison of Crystalline and AMBER Optimized Distances (Å) and Valence Angles (deg) nearby the Anomeric Center for Two Molecules

	Glc-PO3H (xray)	Glc-PO3H (AMBER)	UDP-Glc (xray)	UDP-Glc (AMBER)
C5−O5	1.444	1.42	1.446	1.42
O5−C1	1.408	1.42	1.410	1.42
C1−O1	1.423	1.43	1.425	1.42
O1−P	1.596	1.62	1.612	1.61
C5−O5−C1	113.2	115.5	114.0	114.8
O5−C1−O1	111.6	112.7	111.2	111.7
C1−O1−P	124.8	122.3	120.0	121.6

torsion angles differed by more than 30° (6° for endocyclic torsion).The AMBER dielectric parameters idiel = 0 and dielc = 4, classicaly used for simulating implicit water environment, were used during the CICADA conformational search. The calculations were performed until all the conformations within 5 kcal/mol window were fully examined, requiring a computing time of about 4 days on a Silicon Graphics R10000.

Two programs were used to analyze the results. The Panic program[47] allows for sorting of conformers and flexibility calculations. The FAMILY RMS program[48] clusters the conformers into families based on the rms values of a selection of atoms. To investigate the global shape of nucleotide sugars, the atoms selected are the seven non-hydrogen atoms of the monosaccharides (excluding O6 of glucose) together with three atoms of the base (N_1, C_4, and O_4 for UDP and N_9, N_7, and N_1 for GDP). A rms cutoff of 0.25 Å was selected for the maximum difference allowed in one family.

Potential Hypersurface of UDP-glucose. CICADA/Amber calculations resulted in a total of 6621 geometries on the potential energy hypersurface. Among them, 1712 conformers are energy minima, and 112 of them have a relative energy lower

that 4 kcal/mol. The torsion angles of interest of these conformers are represented in Figure 5. The four torsion angles of the pyrophosphate backbone, i.e., ψ, μ, ν and α, can adopt any of the staggered orientations. In fact, the pyrophosphate P−O bond displays almost unhindered rotation and even eclipsed O−P−O−P torsion angles (μ, ν) are possible, as was previously observed in a review of crystal structures.[49] Nevertheless, the most populated orientations are generally the +synclinal one (gauche+). The only exception is the ψ torsion angles: it can adopt the +synclinal (gauche+) and -synclinal (gauche-) orientations but has a slight energy preference for the +anticlinal conformation (36% of the population between 100° and 166°). Around the ribose ring, the orientation of the torsion angles correspond to what is currently observed in tRNA

(47) Koča, J.; Carlsen, P. H. J. Unpublished "in house" software, 1995.
(48) Kriz, Z. "in house" program, 1998.

(49) Saenger, W.; Reddy, B. S.; Mühlegger, K.; Weimann, G. *Nature* **1977**, *267*, 225−229.
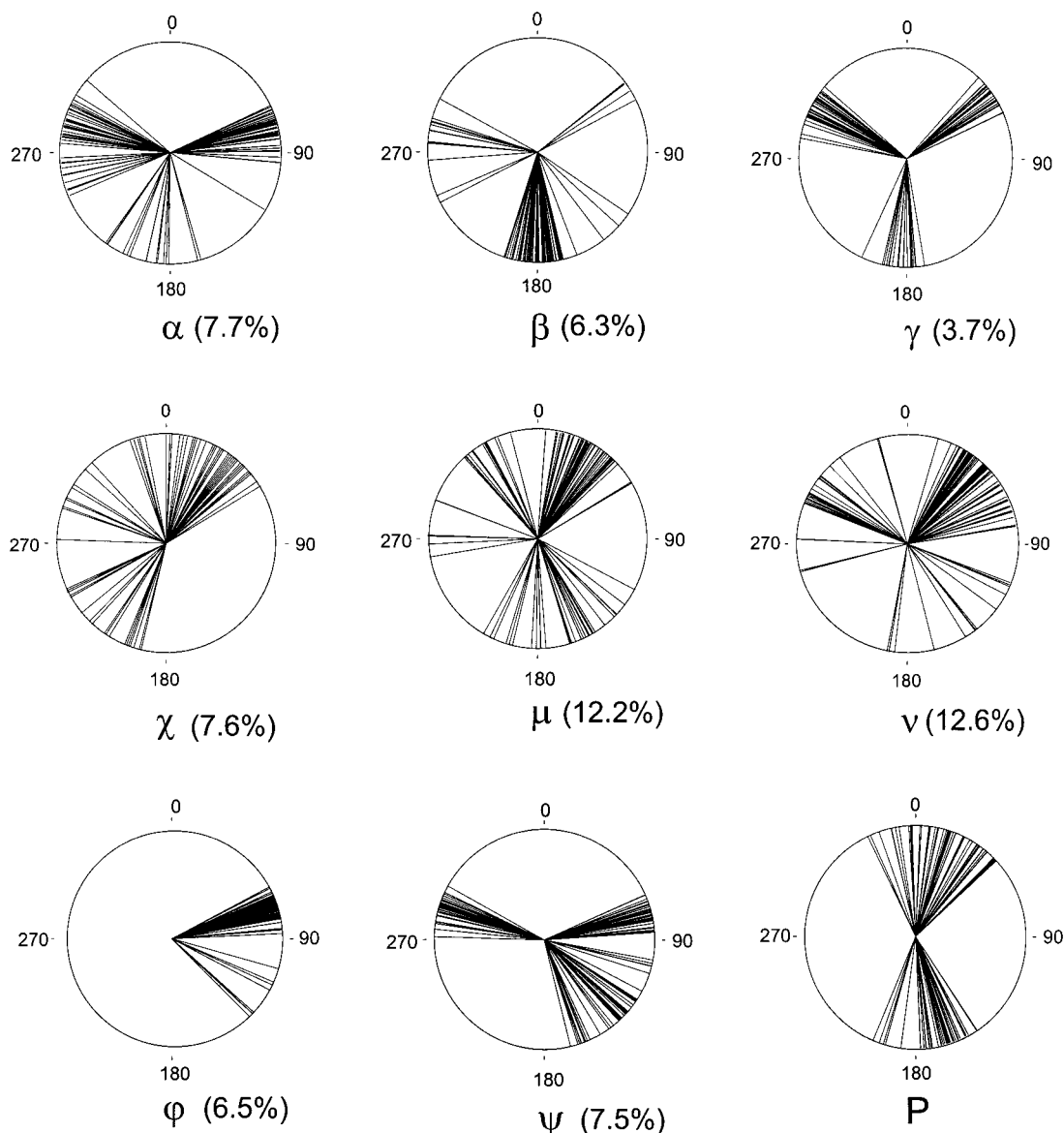
Figure 5. Distribution of UDP-Glc conformers in a 4 kcal/mol energy window for all torsion angles of interest as calculated with the CICADA conformational search method.

polynucleotide crystal structures.[50] The C4′−C5′ bond (γ) adopts the staggered orientation, whereas the C5′−O bond has a strong preference for the anti conformer, the ribose ring is almost equally distributed among C2′-endo and C3′-endo forms. The nucleotide prefers the syn orientation, but the anti one still represents 20% of the population. The ϕ glycosidic torsion angle remains in the +synclinal (centered about 70°) orientation, with some rare incursion in the +anti one (centered about +120°).

Calculated relative flexibility indexes are also indicated in Figure 5 as the contribution of each bond to the total molecular flexibility. These indexes confirm that the pyrophosphate torsion angles (μ, ν) are the most flexible bonds in the molecule. The other torsion angles have almost equivalent values, with the exception of the γ torsion angle. This bond cannot be referred to as "rigid" since it can adopt the three staggered orientations. Nevertheless, its low flexibility index is due to the narrow shape of the three energy wells and to the height of the energy barriers between them.

Clustering of the conformers based on the relative distance and orientation of the glucose ring and the base leads to 6 main

conformational families that have been named by the alphabetic letter that best mimics their shapes. The lowest energy conformers of each of the six most populated families are displayed in Figure 6, and their characteristics are listed in Table 5. Some of the families present long-range interactions between the glucose ring and the base and are referred as U or C. When the "stacking" between the two rings is almost perfect (in U families), the orientation of the glucose ring has been taken into account by checking if the exocyclic carbon (C6) is above or under the ring plane (U_{up} or U_{down}). Due to the stereochemistry of the sugar, the U_{down} conformation is characterized by the occurrence of hydrogen bond between the sugar and the base, whereas the U_{up} one is stabilized by hydrophobic forces involving the methine hydrogens at C3″ and C4″ and the uridine ring. This particular conformation has the lowest energy and displays the shortest distances between the glucose and the base rings (Table 5). The C shape has the same global features as the U_{up} one but is less folded. The uridine ring still interacts with the hydrophobic side of glucose but in a perpendicular way. Family J and L include conformers where the glucose and the base planes are almost perpendicular. Family J is further characterized by a small distance between the ribose and glucose

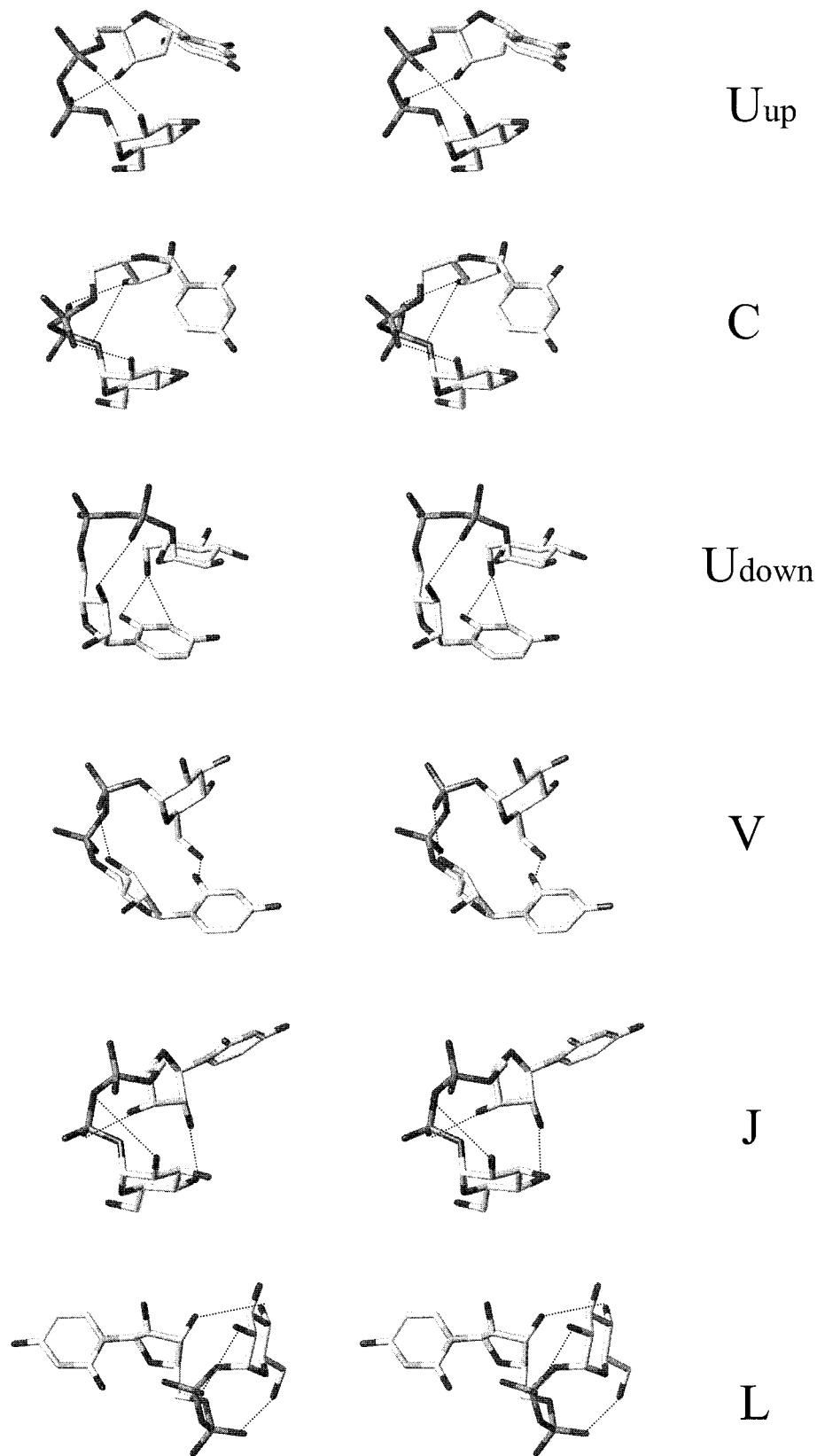(50) Holbrook, S. R.; Sussman, J. L.; Warrant, R. W.; Kim, S. H. *J. Mol. Biol.* **1978**, *123*, 631−660.

Figure 6. Stereographic representation of the lowest energy conformers of the six most probable conformational families of UDP-Glc.

rings, allowing for stabilization by hydrogen bonding. Conformers of the V family present a intermediate case.

Potential Hypersurface of GDP-fucose. The conformational search afforded 5377 points on the PES, including 1388 minima. The 121 conformers with energy lower than 8 kcal/mol were considered for torsion angle analysis (Figure 7). Most of the

conformational features of GDP-fucose are similar to those of UDP-glucose and therefore will not be reiterated again here. The differences mainly occurred at the two extremities of the molecules. The syn orientation of the nucleotide base is strongly energetically favored. This seems to be correlated to the bulkiness of the base rather than to the shape of the ribose ring

Table 5. Description of UDP-Glc Lowest-Energy Conformational Families

family	relative energy of the lowest energy conformer in the family (kcal/mol)	C4″−C4 (Å)	C1″−N1 (Å)
U_{up}	0.00	3.91−6.74	4.58−7.40
C	0.64	5.61−6.34	5.26−6.49
U_{down}	1.68	3.77−5.75	3.48−5.03
V	2.07	6.24−7.17	5.29−5.95
J	2.13	8.73−9.34	7.82−8.12
L	2.89	10.13−11.97	7.51−8.84

since this latter displays only a slight preference for the C3′endo conformation. The Φ glycosidic torsion angle can adopt a wider range of conformation than its homologue in UDP-Glc. This difference in conformational behavior is due to the equatorial versus axial nature of the glycosidic linkage.

Flexibility calculations indicated that GDP-fucose is significantly more rigid than UDP-Glc since their global flexibility indexes displayed a ratio of about 100. When looking at the relative contribution of each torsion angle, the flexibility indexes given in Figure 7 confirm the higher flexibility of the pyrophosphate fragment. Again, GDP-Fuc exhibits different conformational behavior from that of UDP-Glc, when one concentrates on the glycan and the base. The equatorial glycosidic linkage is significantly more flexible than the axial one. On the contrary, the χ torsion angle is now the most rigid bond in the molecule, due to the strong preference for the syn orientation displayed by the bulky guanine base.

The clustering analysis leads to five conformational families that present the same global characteristics as their counterpart in UDP-Glc (Figure 8). The lowest energy one, i.e., the U_{down}, displays a very favorable interaction between the guanine base and the fucose ring, with stabilization by hydrogen bonding. This creates a strong energy preference for this conformer (Table 6). However, other conformational families, displaying more extended shapes, are also described in the present work.

Molecular Dynamics Simulation on UDP-glucose

To compare the conformational behavior found by CICADA search using distance-dependent dielectric constant with modeling in explicit water, we have run two molecular dynamics (MD) simulations. The first simulation (3ns) started with the X-ray (extended) geometry, while the global CICADA minimum was used as an initial geometry for the second simulation (5ns).

Simulation Protocol. The molecule was hydrated in the xLEaP module of AMBER by a periodic box of TIP3P waters which was extended by approximately 10 Å in each direction from the nucleotide sugar atoms. This led to approximately 850 water molecules. All simulations were run with the SANDER module of AMBER-5.0 with SHAKE algorithm[51] (tolerance = 0.0005 Å) to constrain covalent bonds involving hydrogens, using periodic boundary conditions, a 2 fs time step, a temperature of 300 K with Berendsen temperature coupling,[52] a 9 Å cutoff applied to the Lennard-Jones interaction, and constant pressure P = 1 atm. The nonbonded list was updated every 10 steps.

Equilibration was performed by first restraining the atoms of the UDP-glucose (water molecules were allowed to move) and running 1000 steps of minimization. After this initial minimization, all subsequent simulations were run by using the

(51) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comput. Phys.* **1984**, *81*, 3684−3690.

particle mesh Ewald method (PME)[53] within AMBER-5.0 with a cubic B-spline interpolation order and a 10^{-6} tolerance for the direct space sum cutoff. To speed up the fast Fourier transform in the calculation of the reciprocal sum, the size of the PME charge grid is chosen to be a product of powers of 2, 3, and 5 and to be slightly larger than the size of the periodic box. This leads to a grid spacing of ∼1 Å or less. The first step was followed by a 25 ps dynamics with the position of the UDP-glucose fixed. Equilibration was continued with 25 kcal/(mol Å) restraints placed on all solute atoms, minimization for 1000 steps, followed by 3 ps of MD, which allowed the water to relax around the solute. This equilibration was followed by 5 rounds of 600-steps minimization where the solute restraints were reduced by 5 kcal/(mol Å) during each round. Finally, the system was heated from 100 to 300 K over 2 ps, and the production run was initiated.

The MD trajectories were analyzed with the CARNAL module of AMBER.

Analysis of the Trajectories. The rms deviation plots were analyzed along with the distances between the terminal residue centers of mass, and time averaged structures of selected intervals were calculated from the MD trajectories. These averaged geometries were compared with the conformers of UDP-glucose families obtained from the CICADA conformational search. All trajectory analyses, together with torsion angles fluctuation have been deposited.

The first simulation (3 ns) used the X-ray semi-extended structure (close to L family) as the starting point. Analysis of the trajectories, together with fluctuations of selected torsion angles, is displayed in Figure 9. The molecule displays conformational changes of large amplitude: each of the torsion angles shows flexible behavior, spanning the different regions that were predicted by CICADA (Figure 5). More particularly, the torsion angles of the pyrophosphate (ψ, μ, ν, and α) can jump between the three staggered orientations. As for the ribose ring, both puckering modes are visited. The X-ray structure moves very quickly away from the starting conformation and reaches the L conformational family that is quite similar. The V family is sampled between 600 and 1000 ps. The L−V interconversion is probably related with a large change of the γ, ψ, μ, and ν torsions (about 120°). The second region, where this family is again sampled, is between 1500 and 2000 ps. The last 500 ps of the MD trajectory (2500−3000 ps) converge to the J family. The rest of the trajectory belongs to "intermediate" geometries. In summary, the molecule passes three conformational families during the simulation, the average "lifetime" of the existence of a family is about 0.5 ns.

A completely different behavior is observed when using the global minimum from the CICADA search as starting geometry. The molecule remains in the U_{up} folded conformation during the first 3 ns. Then it converges to the J family which is stable within the last 2 ns of the run. The only observed variation in torsion angles involves the puckering of the ribose ring (see Supporting Information for analysis of the trajectories).

The results of the MD simulations confirm the existence of four conformational families that were predicted by the heuristic search. Whereas the conformational behavior of the UDP-glucose within the first MD simulation, which has been initiated by the X-ray geometry, is quite flexible, the global minimum obtained by the CICADA search seems to produce a more stable trajectory.

(53) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
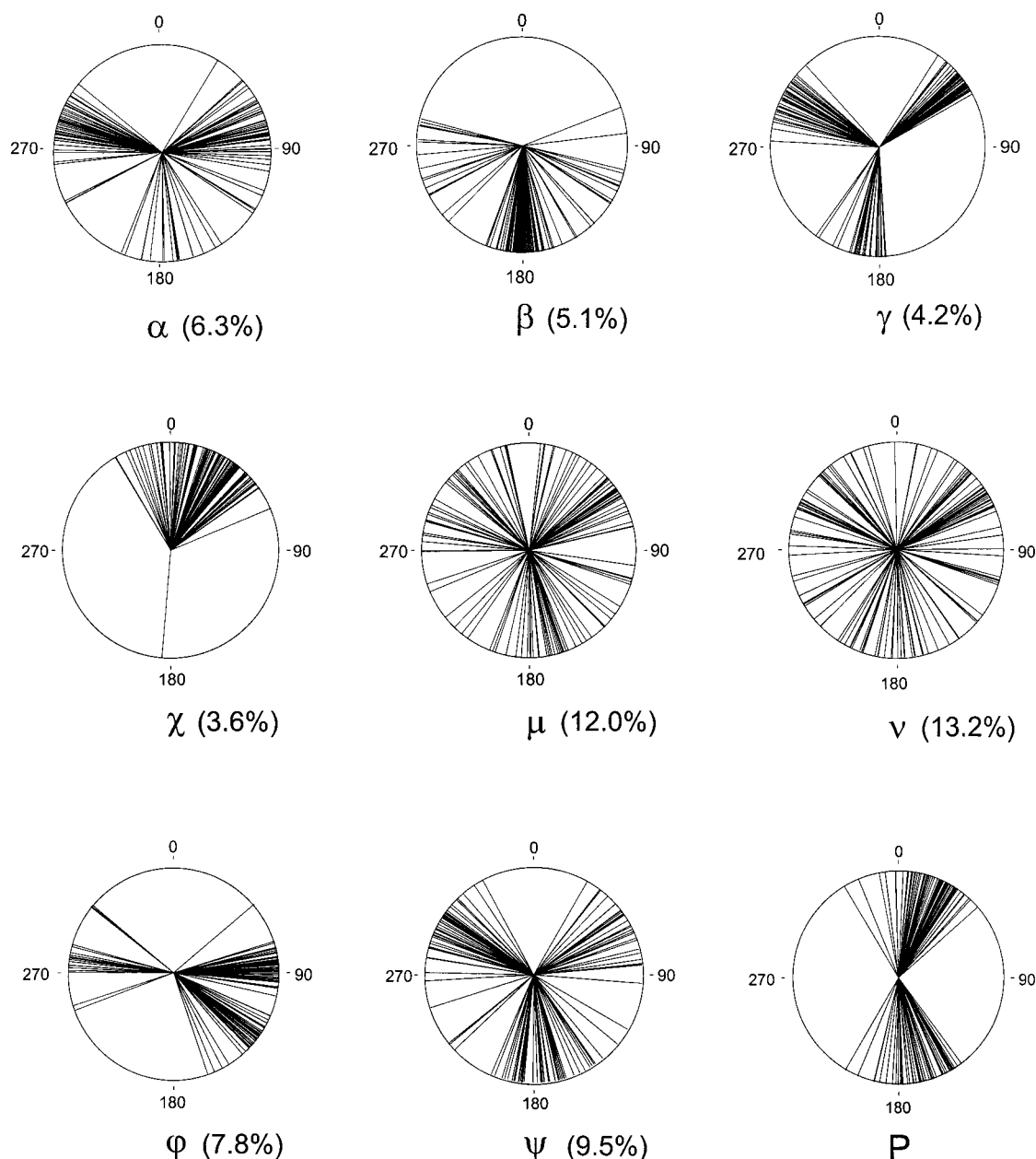
Figure 7. Distribution of GDP-Fuc conformers in a 8 kcal/mol energy window for all torsion angles of interest as calculated with the CICADA conformational search method.

Discussion

Looking for experimental validation of such families in various environments, it can be observed that UDP-Glc (Na$^+$) adopts a conformation close to family L in the crystalline state.[12] When looking at the protein/nucleotide sugar crystalline complexes, it appears that UDP-Glc adopts a rather extended shape in the binding site of UDP-Gal 4-epimerase.[9] This conformation, that maximizes the interaction of both phosphate groups with positively charged amino acids, could be described as an S shape. This conformational family was found during our conformational search procedure, but its energy was not as favorable as those of the ones described in Figure 6. In the crystal structure of glycogen phosphorylase *b*,[54] two molecules of UDP-Glc could be located: one in the glycogen storage site and one at the allosteric site. The first molecule adopts an elongated shape close to an extended S, whereas the second one belongs to the J conformational family. It is therefore confirmed that, depending on the environment, the flexible nucleotide sugar can adopt a variety of conformations. We described some of them here, but other ones could appear if enough stabilization is provided by intermolecular contacts.

All of the low-energy conformations of nucleotide sugars that are described in the present work display at least two intramolecular hydrogen bonds. This stabilization of an "ordered" conformation by hydrogen bonding was predicted to occur in solution from a chiroptical study of UDP-Glc.[55] However, a very recent NMR study of the same compound (Hervé du Penhoat et al., personal communication) evidenced a more extended averaged shape in water solution. It is likely that the very high hydrogen-bonding potential of the nucleotide sugar will yield many interactions with water molecules. It is also

(54) Oikonomakos, N.; Acharya, K. R.; Stuart, D. I.; Melpidou, A. E.; McLaughlin, P. J.; Jonhson, L. *Eur. J. Biochem.*, **1988**, *173*, 569−578.

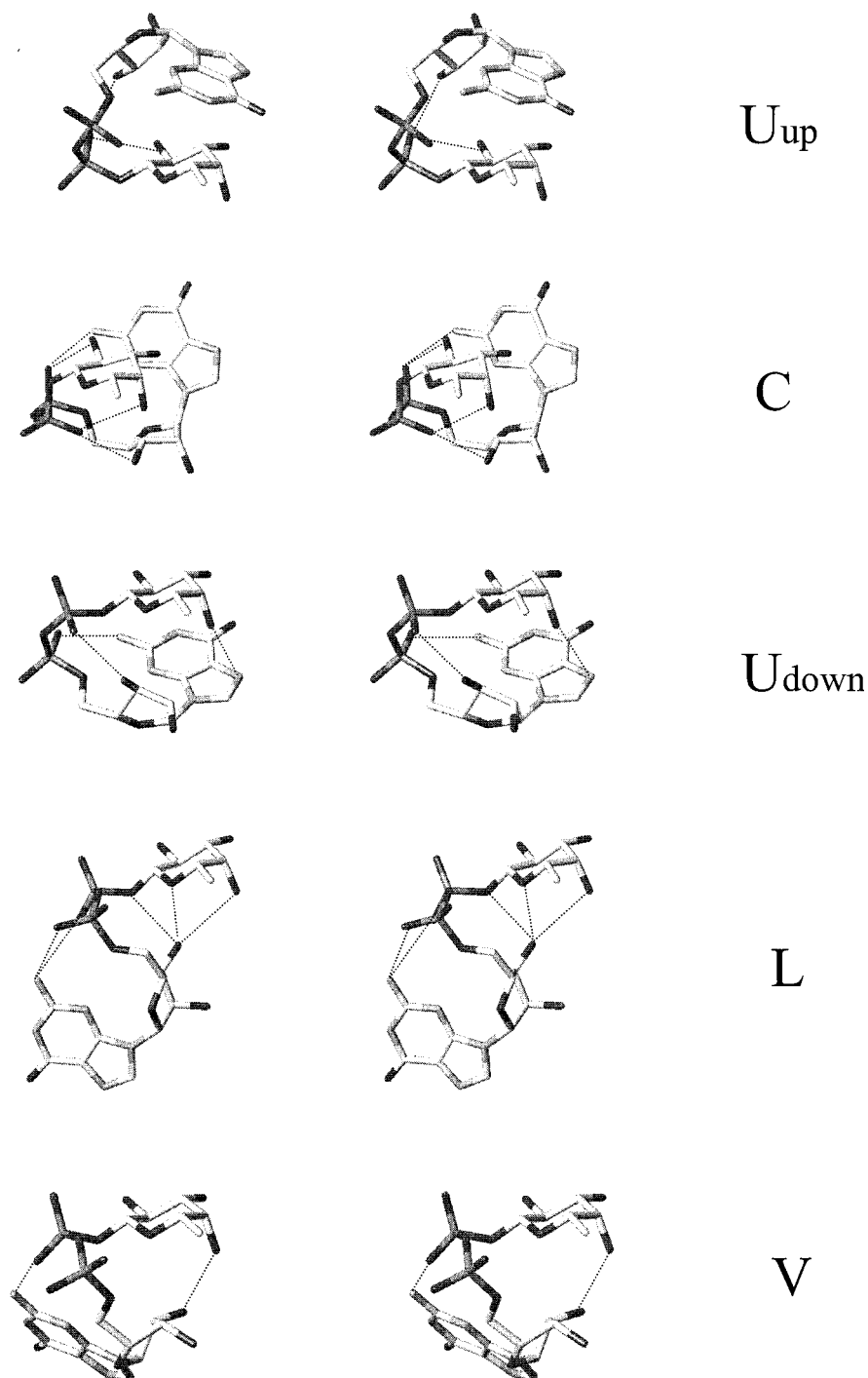(55) Hirano, S. *Biochem. Biophys. Res. Commun.* **1971**, *43*, 1219.

Figure 8. Stereographic representation of the lowest energy representatives of the five most probable conformational families of GDP-Fuc.

Table 6. Description of GDP-Fuc Lowest-Energy Conformational Families

family	relative energy of the lowest energy conformer in the family (kcal/mol)	N9$-$C1″ (Å)	C6$-$C4″ (Å)
U$_{up}$	4.35	3.58$-$7.01	3.70$-$7.53
C	4.58	4.93$-$7.10	5.42$-$7.16
U$_{down}$	0.0	4.00$-$7.47	4.32$-$7.82
V	7.63	5.41$-$6.29	7.86$-$9.56
L	7.02	6.04$-$8.56	10.20$-$12.56

observed that in the protein/nucleotide sugars complexes,[9,51] a dense network of hydrogen bonds is established between the ligand and the binding site, with no intramolecular hydrogen bonding in the UDP-sugar.

Concluding Remarks

Our effort with this work has been to develop a set of parameters for the widely used AMBER all-atoms force field, which could be implemented in conformational studies on nucleotide sugars and also on their complexes with proteins. The conformational analysis results presented in this work show that the molecules can adopt several conformational families. Comparison with experimental data clearly shows that the behavior of these molecules will be different in polar and nonpolar environment, as well as in the crystalline state or in protein binding sites. To confirm the stability of conformational families found by CICADA conformational search using distance-dependent dielectric constant, we have run two MD
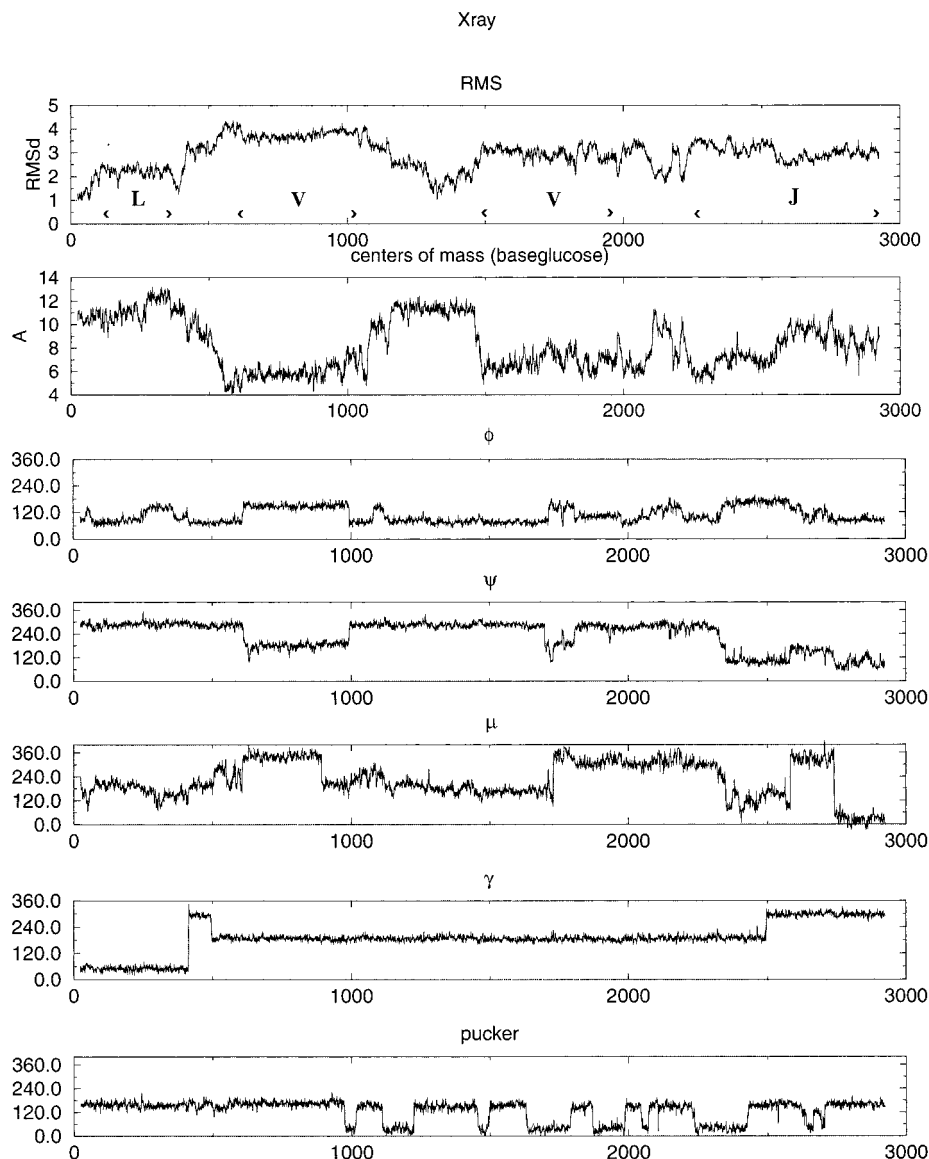
Figure 9. MD simulation of UDP-glucose. Analysis of the evolution of rmsd together with MD histories of the distance between glucose and base center of masses and selected torsion angles of interest. The starting conformation comes from the X-ray structure.

simulations (3 and 5 ns) in explicit water. The variability of the conformational sampling during the MD run is dependent on the starting structure. While the X-ray starting geometry gives quite flexible trajectory, the simulation initiated in the global minimum of the CICADA conformational search is more stable. In all cases, the simulation confirmed the existence of several conformational families in explicit water. In the case of isolated molecules, intramolecular hydrogen bonds and "stacking" interactions are dominant and result in an energy preference of "folded" conformational families. On the other hand, the same interactions, but at the intermolecular level, will favor more extended conformations in polar solution and crystalline state.

In the future, to test and improve the method as well as for obtaining more quantitative results, it will be necessary to include explicit solvent and cations, both monovalent and divalent, in the calculations while taking into account all of the conformational families found in this work.

Supporting Information Available: Figures showing MD simulations of UDP-glucose (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA983854G