MOLECULAR DYNAMICS SIMULATIONS OF CHONDROITIN SULFATE IN EXPLICIT SOLVENT: POINT CHARGE WATER MODELS COMPARED

ANDREI NEAMTU,[*,**] BOGDAN TAMBA[**] and XENIA PATRAS[*]

[*] *"Petru Poni" Institute of Macromolecular Chemistry, 41 A, Gr. Ghica Voda Alley, 700487, Iasi, Romania*
[**] *Center for the Study and Therapy of Pain (CSTD), "Gr. T. Popa" University of Medicine and Pharmacy, 16, University Str., 700115, Iasi, Romania*

Received October 1, 2012

Sulfated glycosaminoglycans (GAGs) are an important class of biomacromolecules involved in many biological functions. Due to the inherent limitations of experimental techniques in assessing the aspects related to their dynamics and interactions in solution, theoretical approaches like molecular dynamics simulations found extended applicability in complementing the experimental field. Recognizing the importance of the accurate solvent representation in carbohydrate simulations in explicit solvent, in this work we have tested four popular water models, namely TIP3P, TIP4P, TIP4P-EW and TIP5P, with respect to their influence on the conformation of a model chondroitin sulfate octasaccharide. The conformation adopted by the simulated saccharide showed no dependency on the water model used on the β(1-4) glycosidc linkage type, while in the case of β(1-3) linkage detectable differences were noticed. In general, it was found that TIP3P favors the direct intra-molecular hydrogen bonding of chondroitin sulfate, while the 4-site and 5-site models disfavor it, facilitating in turn water mediated bridges. Taking into consideration not only the accuracy of the models, but also the computational cost, the data presented here suggest that the TIP4P and TIP4P-EW models give the optimum quality/performance balance.

Keywords: sulfated glycosaminoglycans (GAGs), molecular dynamics, water model, free energy landscape (FEL)

INTRODUCTION

Glycosaminoglycans (GAGs) are a biologically important class of polysaccharides impinging a large variety of fundamental physiological processes. They bind and organize surrounding water molecules, being characterized by high viscosity and low compressibility, which makes them ideal for lubricating fluid in the joints.[1] Also their intrinsic rigidity assures structural integrity to the living cells and form scaffolds for cell migration.[2,3] In contrast with the early view that assess GAGs as inert materials having only structural roles in living tissues, recent advances in experimental techniques render this class of polymers as participating in tissue remodeling, normal tissue homeostasis and disease, including osteo-articular pathology, immune and inflammatory disorders, pulmonary and vascular diseases and cancer.[4,5] These roles are achieved to a large extent by the interaction with different protein ligands, such as polypeptide growth factors,[6,7] extracellular membrane proteins[8,9] and cell-cell adhesion molecules.[10]

In general, these interaction processes are highly specific and the recognition of GAG oligosaccharides by proteins depends on their structural arrangements and the particular conformation they adopt in physiological media.[11] Therefore, the study of GAGs three-dimensional structure and dynamics is of great importance not only from the perspective of biochemistry, but also for a large variety of medical fields. The experimental methodologies used to elucidate the three-dimensional features of glycosaminoglycans alone and/or in interaction with proteins include high resolution NMR spectroscopy,[12,13] X-ray crystallography,[14,15] neutron and electron diffraction[16,17] and hydrodynamic measurements.[18] Despite the fact that these methods have been successfully used for a long time to decipher the carbohydrate structure, they

generally fail to describe in a satisfactory manner the detailed dynamic properties of this class of biomacromolecules and, more importantly, their interaction with water at a molecular level.[19,20] Theoretical computational methods, such as molecular dynamics simulations, can provide a powerful and complementary alternative to experiment for the study of conformational properties of carbohydrates in solution.[21] Molecular dynamics simulations reproduce the spatial-temporal evolution of a system by integrating the classical equations of motion for its component atoms using models described by a suitable force field.[22] In this way, the behavior of the analyzed system can be followed at atomic resolution and on time scales ranging from femtoseconds to microseconds (on today's computers), depending on the complexity of the simulated system.[23] Due to the limited computing power, early simulations on saccharides were usually performed *in vacuo* or in implicit solvent represented as a continuous medium with a certain dielectric constant, but the limitations imposed by the absence of an explicit representation of the solvent are considered to drastically affect the results.[24] It is well known that oligosaccharides and polysaccharides are molecules of high conformational flexibility, which is greatly influenced by the surrounding solvent.[25,26] Therefore, the solvent description in molecular dynamics simulations of this class of molecules is of great importance.

In recent decades, partly due to the importance of water for biology, extensive efforts to build accurate models for water have been made, resulting in more than a few tenths of proposed models.[27] The complexity of these models varies in the number of interaction sites/water monomer, geometry, charges and the presence/absence of polarizability.[28] The parametrization of these models is based on fitting the physical properties of liquid water, such as density anomaly, radial distribution functions and critical parameters.[29] The relatively large number of developed water models expresses their lack of success in quantitatively reproducing all the properties of real water within a single model. However, even simple models can successfully predict at least some of the water properties with unexpected accuracy.[30]

Routinely, in molecular dynamics simulations of biomacromolecules with explicit solvent representation, fixed point charge models (i.e. no polarizability) are used for water, due to the large spatial extension of these systems, which requires many water molecules for a proper solvation. The non-bonded energy evaluations associated with the solvent are the most time-consuming step of a simulation and limit the time scale reached in a certain pre-allocated computing time. In the fixed point charge approach, the electrostatic potential around a water molecule is derived from point charges located at predefined interaction sites, which may or may not coincide with the positions of the oxygen and hydrogen nuclei. The number of these sites may vary from two[31] to six[32], depending on the complexity of the model. One of the mostly used classes of water models for simulation of biomacromolecules is the TIP (Transferable Intermolecular Potential) family of models. These models were developed for the Amber force-field suite[33,34,35,36,37] and they include models with 3, 4 and 5 interaction sites/water molecule. TIP3P[38] is a three-site model with the charges located on the oxygen and hydrogen atoms plus a Lennard-Jones interaction site for oxygen only. TIP4P[39] and TIP4P-EW[40] are four-site models, which place the negative charge on a dummy atom (labeled M in Figure 1a) located near the oxygen along the bisector of the HOH angle. The displacement of the negative charge away from the oxygen position improves the electrostatic distribution around the water molecule. The TIP4P-EW model is a re-parametrization of the standard TIP4P model, improved for use with Ewald summation techniques for electrostatics. There are also other variants of TIP4P model, namely TIP4P/2005,[29] TIP4P/Ice[41] suitable for the study of ices, and TIP4P/FQ[42], the latter including polarizabillity. TIP5P[43] is a 5-site model with 4 sites for electrostatic and 1 site for Lennard–Jones interactions. In this model, the negative charge corresponding to the oxygen atom is divided between two dummy atoms (labeled L in Figure 1a), representing the lone pairs of the oxygen atom, with a tetrahedral-like geometry. TIP5P shows significant improvement in many areas over prior simpler fixed-charge models for liquid water. However, the increased computational cost associated with 5- or 6-site models limits their general use in biomolecule simulations.

The choice of a water model in simulations is dictated by the balance between the necessary accuracy of the solvent representation for the particular system subjected to simulation and the computational cost. For example, three-site water models are the most commonly used for protein–

water simulations because they provide the optimum ratio between accuracy and computational cost, even though the protein force fields in common use were not especially developed for the three-site models. On the other hand, it is currently believed that water models with improvements in bulk water properties will also benefit the solution properties.[40]

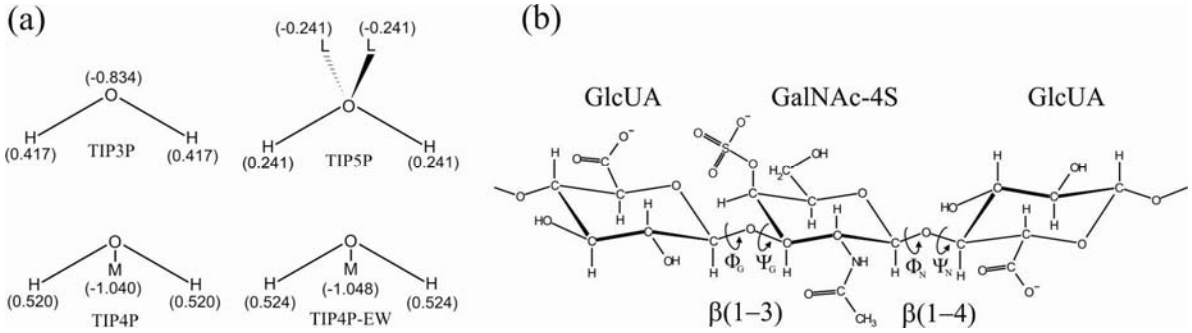

Figure 1: Structural characteristics of (a) water models with their specific charges and interaction sites and (b) chondroitin 4-sulfate (CS4) with the nomenclature of the glycosidic dihedral angles; GlcUA – glucuronic acid, GalNAc-4S – N-Acetyl-Galactosamine sulfated in position 4

Taking into consideration these aspects and the above-outlined importance of the solvent on the conformation that GAGs adopt in solution, the present study aims to evaluate, through a series of molecular dynamics simulations, the impact of the water model on the shape and dynamics of a model GAG, namely an octasaccharide of chondroitin 4-sulfate (Figure 1b), and compare the simulation results with known X-ray crystallography and NMR structural data. The water models evaluated in this respect were the TIP3P, TIP4P, TIP4P-EW and TIP5P and the properties followed during the simulations include the construction of the free energy landscape in the plane of the dihedral angles that define the β(1-3) and β(1-4) glycosidic linkages, radial distribution functions (RDF) of water molecules around selected atoms of the saccharide oligomer and the hydrogen bonds statistics.

EXPERIMENTAL

Molecular dynamics simulations have been performed on octasaccharide molecule models characteristic of chondroitin sulfate (4S) in explicit water. For water, four popular point charge models were used: TIP3P, TIP4P, TIP4P-EW and TIP5P. The obtained trajectories were comparatively analyzed with respect to the conformational preference of the simulated octasaccharide using free energy landscape calculations (FEL) in order to locate the characteristic conformational basins and to estimate how they are influenced by the solvent model used. The solvent organization around the solute was analyzed by means of radial distribution functions (RDFs) and hydrogen bond statistics.

The force field used for the GAG octasaccharide description was the GLYCAM force field[44] with modifications for inclusion of new torsional parameters for the dihedrals involving the $-SO_3^-$ moiety. Due to the absence in the GLYCAM force field of the partial charges for the sulfated monosaccharides, the partial charge for the 1-O-Me-β-GalNAc(4S) monomer was computed from scratch using a similar method with the one described in the original GLYCAM paper.[44] In short, this was done by applying the RESP method of Bayly and co-workers[45] and CHELPG algorithm with 1 fitting stage and a weighting factor of 0.01. Due to the well-known problem of strong dependency of the computed charges on the particular molecular conformation used in the calculations of the target MEP, an ensemble average (EA) method was used. The EA methodology employs a certain number of different conformations of the same molecule in order to determine the charges, followed by an averaging procedure. The ensemble of conformations has been obtained by extracting 50 conformations (equally spaced in time) from a molecular dynamics trajectory (50 ns) on a simple 1-O-Me-β-GalNAc(4S) monosaccharide. This method has the advantage that the ensemble of conformations reflects the conformational preference of the studied molecule at the simulation temperature and in the presence of the solvent (Solvated Ensemble Average).[46] Prior to all quantum MEP evaluations, the geometries of the modeled molecule were optimized taking as the convergence criteria the largest component of the energy gradient to be less than 10^{-5} Ha/b.

Molecular dynamics simulations were performed as described further. In each of the four models used for water, the octasaccharide molecule was placed in the center of a cubic simulation box and then hydrated with a corresponding number of water molecules from a pre-equilibrated box of water at 300K to entirely fill the box. Eight Na^+ ions were added to neutralize the system. Taking into account that two molecules in solution can indirectly influence each other through their hydration layers, the simulation box dimensions were carefully chosen to eliminate the artifactual interaction between the simulated octasaccharide and its periodic images. It was shown that a macromolecule can affect the hydrogen bonding network in the first four shells of water, which extend up to 1 nm.[47,48] The box dimensions were chosen to have an initial value of 6.5 nm, which gives a distance between the neighboring images of about 3 nm, thus minimizing the interaction between periodic images. All the simulations were done in the NPT ensemble (i.e. number of particles, pressure and temperature constant). The Berendsen pressure scheme was used for the pressure coupling with a compressibility of 4.5×10^{-5} (bar^{-1}) and a reference pressure of 1 atm. For temperature coupling, V-rescale thermostat[49] was used with a relaxation constant of 0.1 ps and an equilibrium temperature of 300K. The starting configurations constructed as described above were energy minimized using the steepest-descent method prior to all subsequent simulations. Following the energy minimization, the production run molecular dynamics simulations were performed for 200 ns in each case, which assured a sufficient sampling of the conformational space of the oligosaccharide molecule studied, the time frame of most internal motions in glycans being of the order of hundreds of nanoseconds[50]. All the simulations were done on a 64 core cluster (Dell PowerEdge 1950 servers) with Infiniband computing network infrastructure (Molecular Modeling Laboratory, CSTD – "Gr. T. Popa" University of Medicine and Pharmacy, Iasi). The resulting trajectories were analyzed on a Dell Precision 690 workstation using tools in Gromacs 4 suite for RDFs and hydrogen bonds statistics and specific software for FEL calculations.

The free energy plots were computed from the distribution of the system states in the bidimensional space of glycosidic dihedral angles by taking the histogram (density of states) and transforming it in accordance with the Boltzmann formula:

$$\Delta G = -k_B T \ln \frac{P_{i,j}}{P_{max}}$$

where ΔG is the Gibbs free energy difference, k_B is the Boltzmann constant, T is the absolute temperature, $P_{i,j}$ is the frequency corresponding to the i,j interval in $\Phi\times\Psi$ plane and P_{max} is the maximum frequency over the entire range of Φ and Ψ dihedral angles.

RESULTS AND DISCUSSION

The Φ and Ψ glycosidicdihedral angles define the relative orientation of two adjacent sugar rings and when plotted against each other in a plane the coupling between these two angles can identify the conformational basins visited by the saccharide chain with respect to backbone conformation. These types of plots are similar to Ramachandran plots[51] for polypeptide chains. In the original approach for proteins, the graphs were plotted by using the density of points, each point being associated with a "conformational state" in the plane defined by the two dihedral angles. This can readily be transformed into a free energy surface by constructing the two-dimensional histogram of states and applying the Boltzmann formula as described in the "Experimental" section.

The obtained FELs in the (Φ, Ψ) plane for both β(1-3) and β(1-4) linkage types are presented in Figure 2 comparatively for the four water models evaluated in this study. It can be seen that there are no observable differences between the simulations with different solvent models for β(1-4) linkage type. A single conformational basin can be distinguished for all solvent models considered here. The mean values are Φ_N = 289° and Ψ_N = 242° with a standard deviation between the water models under 2.5° for Φ_N and under 5.5° for Ψ_N. The values obtained here from the simulations agree very well with the NMR experimental values in solution for Φ_N and Ψ_N : Φ_N = 290°, Ψ_N = 240° (GalNAc-6Sβ(1-4)GlcUA); Φ_N = 296°, Ψ_N = 234° (GalNAc-4Sβ(1-4)GlcUA)[52]; Φ_N = 287°, Ψ_N = 243° (unsulfated form of CS4)[53], with those from combined NMR (solution)/ computational studies: Φ_N = 280°, Ψ_N = *250*° (GalNAc-4Sβ(1-4)GlcUA)[54] and with other computational study results: Φ_N = 290°, Ψ_N = 240° [55]; Φ_N = 282°, Ψ_N = 249° .[56]

The NMR technique was successfully used in the past for extracting structural data on sulfated GAGs,[57,58] but the proton NMR spectra are very complex for this class of saccharides,[59,52] so the first structures for chondroitin sulfate oligomers were resolved by X-ray fiber diffraction, PDB code 1C4S and 2C4S.[60,61] The dihedral angles obtained here by simulation for β(1-4) linkage type differ slightly from the ones in the 1C4S structure (Φ_N = 280°, Ψ_N = 231°), but deviate significantly by 27° and respectively 56° from the ones in the 2C4S structure (Φ_N = 262°, Ψ_N =

186°). This is expected due to the fact that the simulations were performed in solution, while the packing of the polymeric chains may influence the conformation in the crystal state. Moreover, the crystal for the 2C4S structure was obtained in the presence of Ca^{2+} ions, which due to their high positive charge make a link between the sulfate moiety of GalNAc-4S and the carboxyl of the GlcUA residues, visible in the resolved crystallographic structure.

The hydrogen bond analysis (Table 1) shows that the conformation of the β(1-4) linkage is stabilized by two intramolecular hydrogen bonds, in accordance with the earlier predicted models[62,60].

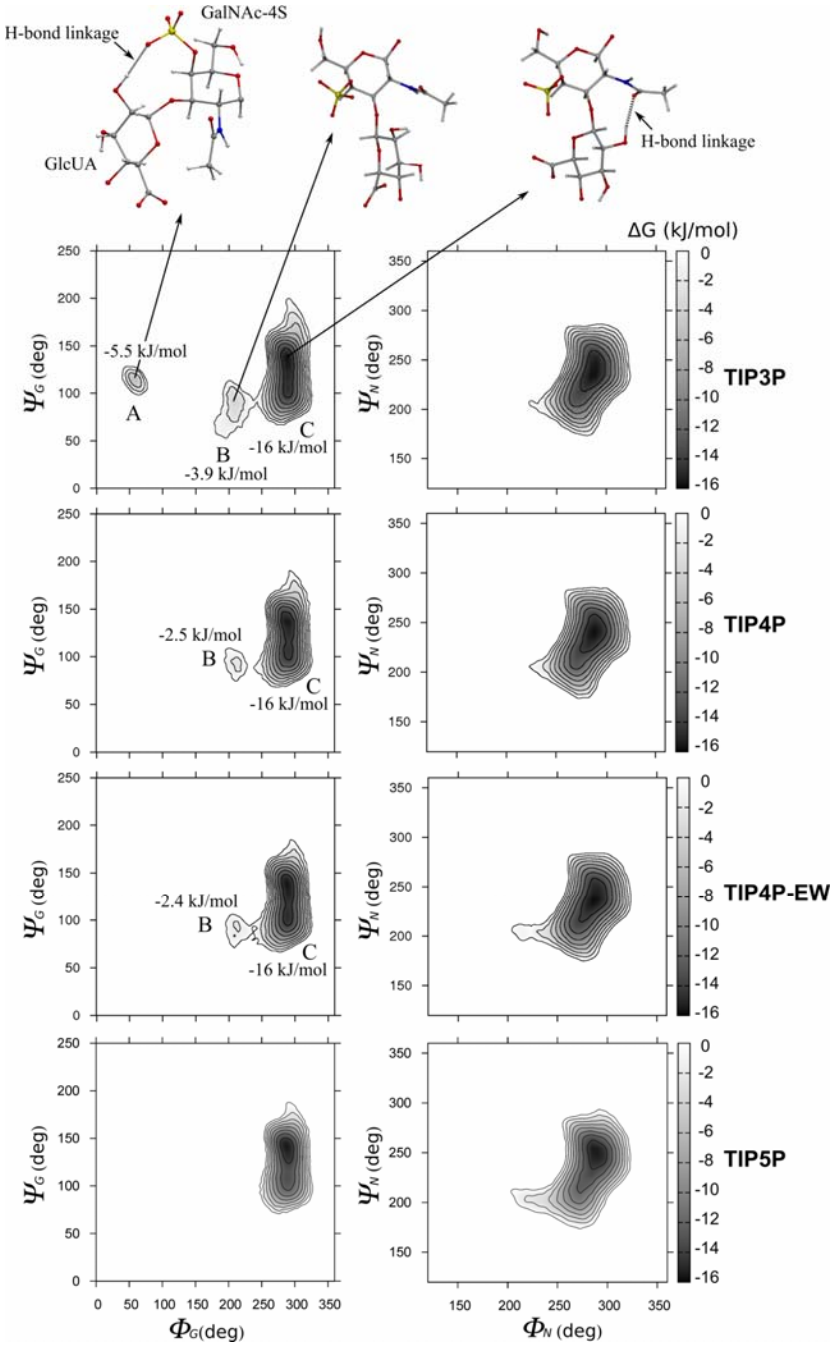

Figure 2: Free energy landscapes for chondroitin sulfate (4S) octasaccharide, using the glycosidic torsional angles as generalized coordinates for the conformational space

One of these bonds is formed between the hydroxyl groups in position 3 of GlcUA and the ring oxygen of the adjacent GalNAc-4S. This is the most prevalent hydrogen bond that affects the β(1-4) conformation, as can be seen from Table 1. The other one appears between the carboxyl group of GlcUA and the –NH group of GalNAc. This bond is less frequent and is characterized by a shorter lifetime than the first one. Contrary to the β(1-4) linkage, for the β(1-3) linkage type, the FEL shows variations dependent on the model used for the solvent description. In all simulations, one main conformational basin with a mean Φ_G = 288° and Ψ_G = 137° and a standard deviation under 1.5° for Φ_G and under 4° for Ψ_G was identified. The value predicted for Φ_G correlates well with the experimental NMR solution structures in the literature, while Ψ_G shows a deviation of nearly 30° from the experiment (Φ_G = 299°, Ψ_G = 109° (GlcUAβ(1-3)GalNAc-4S);[52] Φ_N = 288°, Ψ_N = 108° (unsulfated form of CS4).[53] Combined NMR/computational and computational studies give values closer to the ones presented here only for Φ_G (Φ_G = 280°, Ψ_G = 90° (GlcUAβ(1-3)GalNAc-4S);[54] Φ_G = 281°, Ψ_G = 90°;[55,56] Φ_G = 270°, Ψ_G = 70° (GlcUAβ(1-3)GalNAc-6S).[56] However, a closer analysis regarding the shape of the main conformational basin reveals a rather flat aspect around the minimum free energy value with variations under 1 kJ/mol, when Ψ_G spans an interval between 103° and 141°. Consequently, it can be considered that Ψ_G dihedral angle encompasses an interval of several dozens degrees during the molecular dynamics simulations at normal temperatures (1 kJ/mol is well below 1 k_BT, which is a measure for thermal energy). The minimum energy conformation that β(1-3) linkage adopts in solution is stabilized by a hydrogen bond that forms between the hydroxyl group in position 2 of GlcUA and the carbonyl oxygen of the acetamido of GalNAc-4S.

One minor basin denoted by 'B' in Figure 2 is identified in TIP3P, TIP4P and TIP4P-EW based simulations with a minimum at Φ_G = 209° and Ψ_G = 92°. For this conformation there is no hydrogen bond between the GlcUA and GalNAc-4S residues in the proximity of β(1-3) linkage. It is most visible for the TIP3P solvent model, less populated in the case of TIP4P and TIP4P-EW models and it vanishes in the TIP5P based simulation.

For TIP3P solvent model, another distinct conformation is favored, expressed by the 'A' basin defined at its minimum by Φ_G = 56° and Ψ_G = 116°, which is stabilized by a hydrogen bond formed between the OH group in position 2 of GlcUA and the SO_3^- group of GalNAc-4S. This conformation does not appear in the FEL plot, when a more complex water model is used (TIP4P, TIP4P-EW and TIP5P) and it is also not detectable in the experiment.

To further investigate the interactions between the solvent and the saccharide molecule and the influence of the ordering of the solvent around the solute and how they participate to the conformations detected in the FEL maps, the radial distribution functions and the hydrogen bonds statistics were evaluated.

Figure 3 shows the radial distribution functions (RDF) of water oxygen atoms (OW) around the selected solute oxygen and nitrogen atoms of the solute (a-e) along with the water-water oxygen atom RDFs (f). The atoms for RDF analysis were selected on the basis of their ability to participate at intra-molecular hydrogen bonding, as it is known that the pattern of this hydrogen bond network is greatly influenced by the solvent.

The RDFs between the OW and the hydroxyl oxygen atoms (O2 on the GlcUA and O6 on the GalNAc-4S) shows that the OH groups significantly contribute to the inter-molecular hydrogen bonding, which is in agreement with other simulation studies on similar sugar compounds.[63] Both RDFs for O2 and O6 show a well-defined first layer of hydration with the peak at 0.28 nm. In the case of the O6 atom, this first solvent layer is more organized for the TIP3P, TIP4P and TIP4P-EW models with a peak value of 1.7, compared to 1.5 in the case of the TIP5P model. The second layer of hydration on the other hand, with the peak at 0.55 nm is better represented for the TIP5P model (for both O2 and O6 atoms) than for the rest of the models used, which do not show this second peak. This suggests that in the case of the TIP5P model the solvent develops a better defined long range order around the hydroxyl groups of the sugar ring compared with the other models studied here. The RDFs of OW around the oxygen atoms of the sulfate moiety of GalNAc-4S residues also show a well-defined first layer of hydration at 0.27 nm for TIP3P, TIP4P and TIP4P-EW and at 0.28 nm for TIP5P.

The peak value for this first layer is again lower for TIP5P with a value of 1.5, compared to 1.7-1.8 for TIP3P and respectively for TIP4P and TIP4P-EW.

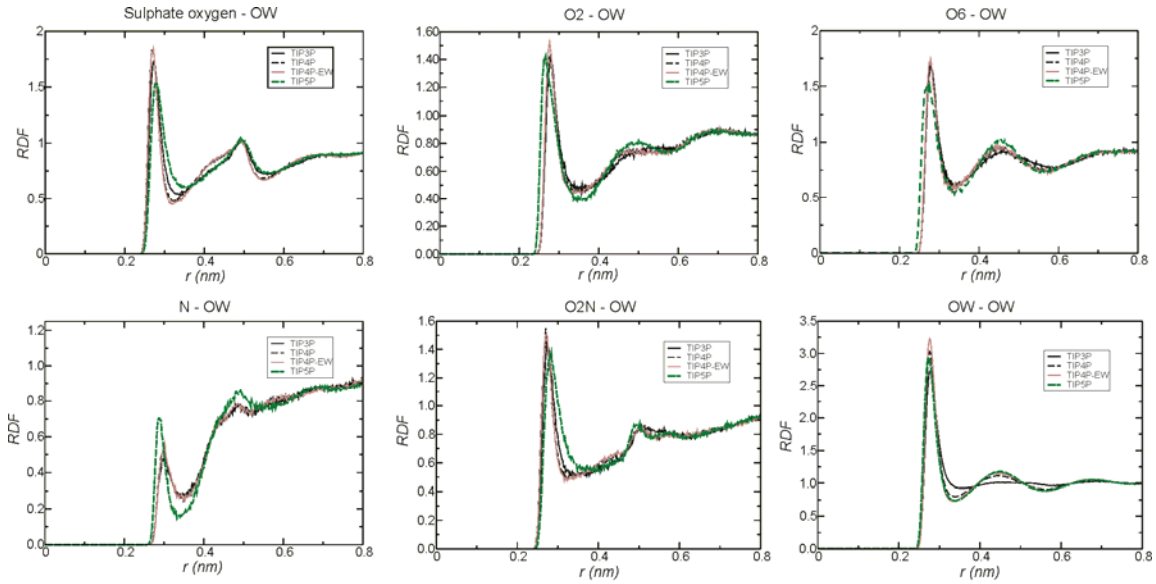

Figure 3: Radial distribution functions of water oxygen atoms (OW) around selected oxygen and nitrogen atoms of the simulated octasaccharide (a-e); f – oxygen-oxygen radial distribution function for water-water interaction

Table 1
Intra-molecular hydrogen bond statistics for chondroitin sulfate octasaccharide

	N_{hb} [a] O5-O3	τ (ps)[b] O5-O3	N_{hb} COO^--N	τ (ps) COO^--N	$N_{hb}\times10^{-3}$ $OH_{(2)}$-$O_{(N)}$	τ (ps) $OH_{(2)}$-$O_{(N)}$	$N_{hb}\times10^{-3}$ SO_3^--$OH_{(2)}$	τ (ps) SO_3^--$OH_{(2)}$
TIP3P	1.75±(0.012)	55.15	0.461±(0.005)	47.1	32.33±(4.67)	8.518	761.4±(4.6)	32.461
TIP4P	1.69±(0.012)	80.84	0.340±(0.004)	43.6	14.70±(1.47)	13.907	632.3±(5.4)	14.796
TIP4P-EW	1.71±(0.015)	106.4	0.315±(0.005)	63.8	14.00±(1.57)	13.713	588.9±(5.7)	21.388
TIP5P	0.92±(0.017)	114.0	0.331±(0.006)	75.9	5.00±(1.05)	9.884	104.9±(2.3)	22.905

[a] N_{hb} is the number of hydrogen bonds for the corresponding pairs expressed as the average of all the snapshots saved;
[b] τ is the hydrogen bond lifetime computed in accordance with Luzar and Chandler theory[65]

The lower peak amplitude for the first layer of hydration observed for the TIP5P model may be explained by the reduced partial charge values in this model compared with the other models studied here. The force-field methods rely on electrostatics and van der Waals interactions to reproduce the hydrogen bonding. The number and the lifetime of hydrogen bonds and, consequently, the water organization around the solute result from a delicate balance between the associated partial point charges, the geometry of the water model and the Lennard-Jones non-bonded parameters. The lower value for the amplitude of the first peak of hydration in the case of TIP5P is also visible in the bulk OW-OW (water to water) RDFs. Instead, the second peak corresponding to the second layer of hydration is better represented for TIP5P, expressing a more established long-range ordering of the solvent (Figure 3f). It is known that the TIP5P model gives a better geometry for the water dimer and a more tetrahedral water structure, which result in a closer agreement with the experimental RDFs from neutron diffraction data.[43,64] From the OW-OW RDF plots, it can be seen that the TIP4P and TIP4P-EW models give results for bulk water-water interactions very close to TIP5P, in contrast with the TIP3P model.

The formation and stabilization of the hydration layers around a solute molecule is

greatly influenced by the H-bond network that develops in the simulated system. Thus an H-bond statistics is needed in order to complement the data obtained by the FEL and RDF analyses. The followed parameters were the number of hydrogen bonds formed between different groups and the hydrogen bond lifetime. The hydrogen bond analysis was divided into two parts. Table 1 gives the intra-molecular hydrogen bond statistics for the CS4 octasaccharide and Table 2 presents the inter-molecular water/CS4 statistics.

Table 2
Hydrogen bond statistics between the solvent and the chondroitin sulfate octasaccharide

	N_{hb} W-W	τ (ps) W-W	N_{hb} W-CS4	τ (ps) W-CS4	Wat. bridges [(a)] W-CS4
TIP3P	7515±(1.1)	1.866	86.28±(0.046)	2.478	9.86±(0.058)
TIP4P	7903±(1.1)	3.081	86.45±(0.044)	5.275	9.66±(0.069)
TIP4P-EW	8101±(1.0)	4.563	85.58±(0.041)	11.305	9.65±(0.081)
TIP5P	7279±(1.1)	4.287	84.52±(0.049)	11.749	10.44±(0.085)

[(a)] number of water bridges per water molecule per 200 ns

As a general remark regarding the intra-molecular hydrogen bonding, it can be seen from Table 1 that the model used to describe the solvent greatly influences the frequency of formation of direct hydrogen bonds between different groups of the CS4 saccharide molecule. The TIP3P model favors these direct H-bonds, while the four-site models and especially the TIP5P five-site water model disfavor them.

The two hydrogen bonds that mainly affect the *β*(1-4) glycosidic linkage, namely between the –OH in position 3 of GlcUA and the ring oxygen of the adjacent GalNAc-4S (O5-O3 column in Table 1) and between the carboxyl group of GlcUA and the –NH group of GalNAc (COO^- - NH column in Table 1) are less frequent up to a fraction between 50% and 70% for the TIP5P water model, compared to TIP3P. The TIP4P and TIP4P-EW water models give values closer to TIP3P for the number of hydrogen bonds for the O5-O3 interaction, while the bond lifetime is closer to TIP5P for the TIP4P-EW model. For the COO^- – NH interaction, the number of hydrogen bonds per saved frame are comparable for TIP4P, TIP4P-EW and TIP5P, while again the lifetime of the formed hydrogen bonds in TIP4P-EW solvent is closer to that of the TIP5P model. The conformational preference of the *β*(1-4) linkage is probably not affected by the water model used, as it can be seen from the FEL plots in Figure 2, because the H-bond lifetime, for both hydrogen bond types, is increasingly higher for the TIP4P and TIP5P models.

The *β*(1-3) linkage conformation is, in contrast, greatly affected by the model used for the solvent description. This can be related with the H-bond statistics results in Table 1 for the two hydrogen bonds found to influence this glycosidic link. The number of direct H-bonds between the hydroxyl group in position 2 of GlcUA and the carbonyl oxygen of the N-acetyl side chain of GalNAc-4S on the one hand and between the OH group in position 2 of GlcUA and the SO_3^- group of GalNAc-4S on the other hand, is drastically reduced when the TIP5P model is used instead of the TIP3P model. The direct H-bond between SO_3^- and $OH_{(2)}$ is responsible for the unusual conformation encountered in the basin 'A' of the free energy plot (Figure 2). The lifetime of both H-bond types shows variations between the different water models, but they are small in comparison with the relative variations in H-bond numbers and do not correlate with them.

The severe decrease of intra-molecular direct H-bonding for the groups that affect the *β*(1-3) linkage conformation in the case of the TIP5P water model can be explained by the presence of solvent mediated interactions, such as water bridges involving one or more water molecules. This was observed by visually inspecting the molecular dynamics trajectories and is in accordance with the results of other studies.[63]

To quantitatively evaluate this aspect, a statistical analysis of water bridge population was performed for CS4 octasaccharide as a whole. The analysis was restrained to water bridges

containing only one solvent molecule. One water molecule is considered to form a bridge if it makes two simultaneous H-bonds with the CS4 molecule. This was evaluated for every frame in the saved trajectories for all of the four water models analyzed and expressed as number of water bridges/water molecule per 200 ns (the entire time interval of the simulations). The results are presented in Table 2. It can be seen that the TIP5P model gives a statistically significantly higher number of water bridges along the CS4 octasaccharide than the TIP3P, TIP4P and TIP4P-EW, which sustains the above assumption. Interestingly, the use of the TIP4P and TIP4P-EW models results in a smaller number of such bridges than in the case of the TIP3P model, while still showing a decreased number of direct intra-molecular hydrogen bonds along the CS4 molecule. This result is surprising and it demonstrates that one must also consider the lifetime of water bridges and the extent of higher order water bridge populations to complete the picture of how these interactions influence the conformation of polysaccharides in water. However, a quantitative description of higher order water bridges is much more difficult to assess than for simple bridges (involving only one water molecule), therefore they should make the objective of a separate study.

The water-water and water-CS4 hydrogen bond statistics shows no significant variations in the number of hydrogen bonds per saved frame but in the lifetime of these bonds. The lifetime for water-water interactions agrees well with the earlier experimental and theoretical determined values.[66,67] These are higher and closer to the experiments for the TIP4P, TIP4P-EW and TIP5P models. For the water-CS4 interactions the lifetimes are also higher and close to each other for the TIP4P-EW and TIP5P models, compared with the TIP3P and TIP4P models. A higher lifetime of hydrogen bonds formed between water and CS4 in the case of TIP4P-EW and TIP5P may also indicate a higher lifetime of water bridge-mediated H-bonding between different groups of CS4, but this assumption still needs a more in-depth and direct assessment.

The choice of the water model to be used in simulations is mainly directed by the balance between the accuracy that the model can provide for the particular class of the systems studied on the one hand and the computational cost on the other. The computing time increases as the number of interaction sites increases due to the fact that more distances (which explicitly appear in Lennard-Jones (LJ) and electrostatic energy formulas) between interacting pairs have to be evaluated for the non-bonded part of the potential energy function. In Figure 4, the computing performance is depicted for all the four water models evaluated. This is expressed as the number of nanoseconds of simulation performed/day using the hardware mentioned in the "Experimental" section.

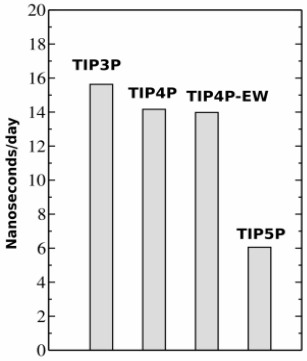

Figure 4: Computational performance (expressed as nanoseconds/day) for the simulated CS4 octasaccharide using different models for water description

The TIP3P model uses three interaction sites for the electrostatic interactions, which gives 9 distance evaluations between one pair of water molecules (distances are the same for both LJ and electrostatic interactions). The TIP4P and TIP4P-EW models also use 3 interaction sites for electrostatics, but the LJ site atom and the point charge for the oxygen do not coincide (Figure 1), which gives 9 distance evaluations for electrostatic interactions plus one for O-O LJ interactions (3×3+1=10). The TIP5P model uses 4 interaction sites for electrostatics and one for LJ

interaction, which results in 17 distance evaluations for one pair of water molecules (4×4+1=17). The number of distance evaluations for each model is reflected in the computing time required for every simulated nanosecond (Figure 4). As expected, the TIP3P model being the simplest one gives a better computational performance, while the TIP5P the poorest. The TIP4P and TIP4P-EW models show a high performance in terms of computational cost and good accuracy in examined properties, closer to TIP5P. Taking into account the free energy analysis, radial distribution functions and hydrogen bonds analysis presented here, out of the four water models analyzed the TIP4P-EW model seems to give the best balance between the quality of the solvent description and computational performance.

CONCLUSION

Molecular dynamics simulations have been performed on a chondroitin sulfate octasaccharide/water system in order to evaluate the influence of the complexity of the solvent description on the conformation of the glycosaminoglycan molecule. Four water models were used in this respect: TIP3P, TIP4P, TIP4P-EW and TIP5P. The conformation of the GAG molecule obtained by simulation is in good agreement with the experimental X-ray diffraction and NMR data for all the solvent models. However, there are detectable differences among the different TIP models, especially for the TIP3P, when compared to TIP4P, TIP4P-EW and TIP5P. While in the case of β(1-4) linkage type the differences are minimal, for β(1-3) linkage type, the free energy landscape in the (Φ, Ψ) glycosidic dihedral space shows the presence of a conformational basin not visible when 4-site or 5-site water models are used, nor in experimental data. This conformation is stabilized by a hydrogen bond favored by the TIP3P model, which forms between the OH group in position 2 of GlcUA and the SO_3^- group of GalNAc-4S. The differences among the results obtained with different TIP models for water can be explained on the one hand by the fact that the solvent around the GAG molecule influences the intra-molecular hydrogen bond network of chondroitin sulfate, as sustained by the hydrogen bond analysis. In general, it was found that TIP3P favors the direct intra-molecular hydrogen bonding of chondroitin sulfate, while the 4-site and 5-site models disfavor it. On the other hand, the radial distribution function analysis reveals that the solvent organization in hydration layers around the chondroitin sulfate is dependent on the accuracy of the water model. The TIP5P model shows long range order, while the TIP3P model gives a good description only for the first layer of hydration. Also, taking into consideration the computational performance criteria, it seems that the TIP4P and TIP4P-EW models give the optimum balance between the quality of the solvent description and computational performance in the case of glycosaminoglycan simulations. Another important aspect regarding the data presented here is not related only to the useful information about the solution conformation of chondroitin sulfate as a function of the water model used, but also to the evaluation of the time scales of conformational changes and of the intra-molecular and sugar water specific interactions, which are responsible for the conformation that the saccharide adopts. The results of this work are thus of great interest for computational studies on the conformations and dynamics of chondroitin sulfate oligomers in interaction with other biological macromolecules like proteins.

ACKNOWLEDGMENTS: This research was financially supported by the European Social Fund – "Cristofor I. Simionescu" Postdoctoral Fellowship Programme (ID POSDRU/89/1.5/S/55216), Sectoral Operational Programme on Human Resources Development 2007–2013.

REFERENCES

[1] M. David-Raoudi, B. Deschrevel, S. Leclercq, P. Galera and K. Boumediene and J. P. Pujol, *Arth. Rheum*., **60**, 760 (2009).
[2] M. M. Murray, S. D. Martin and M. Spector, *J. Orthop. Res*., **18**, 557 (2000).
[3] M. G. Haugh, C. M. Murphy, R. C. McKiernan, C. Altenbuchner and F. J. O'Brien, *Tissue Eng. A*, **17**, 1201 (2011).
[4] C. B. Knudson and W. Knudson, *FASEB J*., **7**, 1233 (1993).
[5] R. Stern, A. A. Asari and K. N. Sugahara, *Eur. J. Cell Bio*., **85**, 699 (2006).
[6] J. D. Esko and S. B. Selleck, *Annu. Rev. Biochem*., **71**, 435 (2002).
[7] J. T. Gallagher, *J. Clin. Invest*., **108**, 357 (2001).
[8] E. Hileman, J. R. Fromm, J. M. Weiler and R. J. Linhardt, *BioEssays*, **20**, 156 (1998).

[9] M. Bernfield, M. Gotte, P. W. Park, O. Reizes, M. L. Fitzgerald *et al*., *Annu*. *Rev*. *Biochem*., **68**, 729 (1999).
[10] C. R. Franco, E. S. Trindade, H. A. Rocha, R. B. da Silveira, K. S. Paludo *et al*., *Biochem*. *Cell Biol*., **87**, 677 (2009).
[11] A. Imberty, H. Lortat-Jacoband and S. Pérez, *Carbohyd*. *Res*., **342**, 430 (2007).
[12] C. Vanhaverbeke, J. P. Simorre, R. Sadir, P. Gans and H. Lortat-Jacob, *Biochem*. *J*., **384**, 93 (2004).
[13] A. Canales-Mayordomo, R. Fayos, J. Angulo, R. Ojeda, M. Martin-Pastor *et al*., *Biomol*. *NMR*, **35**, 225 (2006).
[14] I. Capila, M. J. Hernaiz, Y. D. Mo, T. R. Mealy, B. Campos *et al*., *Structure*, **9**, 57 (2001).
[15] C. Shao, F. Zhang, M. M. Kemp, R. J. Linhardt, D. M. Waisman *et al*., *J. Biol*. *Chem*., **281**, 31689 (2006).
[16] G. A. Jeffrey, *Acta Crystallogr*. *B*, **46**, 89 (1990).
[17] S. Perez, *Methods Enzymol*., **203**, 510 (1991).
[18] H. Perez-Sanchez, K. Tatarenko, M. Nigen, G. Pavlov, A. Imberty *et al*., *Biochemistry*, **45**, 13227 (2006).
[19] F. Corzana, M. S. Motawia, C. Herve du Penhoat, S. Perez, S. M. Tschampel *et al*., *J. Comput*. *Chem*., **25**, 573 (2004).
[20] A. Imberty and S. Perez, *Chem*. *Rev*., **100**, 4567 (2000).
[21] J. L. M. Jansson, A. Maliniak and G. Widmalm, in "NMR Spectroscopy and Computer Modelling of Carbohydrates: Recent Advances", edited by J. F. G. Vilegenthart, R. J. Woods, American Chemical Society, Washington, DC, 2006, pp. 20-39.
[22] A. R. Leach, "Molecular Modelling, Principles and Applications", Prentice Hall, Second edition, 2001, pp. 165-247.
[23] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror and D. E. Shaw, *Curr*. *Opin*. *Struct*. *Biol*., **19**, 120 (2009).
[24] A. Almond and K. J. Sheehan, *Glycobiology*, **13**, 255 (2003).
[25] J. W. Brady, *Adv*. *Biophys*. *Chem*., **1**, 155 (1990).
[26] K. N. Kirschner and R. J. Woods, *Procs*. *Natl*. *Acad*. *Sci*. *USA*, **98**, 10541 (2001).
[27] B. Guillot, *J. Mol*. *Liq*., **101**, 219 (2002).
[28] C. Vega, J. L. F. Abascal, M. M. Conde and J. L. Aragones, *Faraday Discuss*., **141**, 251 (2009).
[29] L. F. Abascal and C. Vega, *J. Chem*. *Phys*., **123**, 234505 (2005).
[30] H. L. Pi, J. L. Aragones, C. Vega, E. G. Noya, J. L. F. Abascal *et al*., *Mol*. *Phys*., **107**, 365 (2009).
[31] K. M. Dyer, J. S. Perkyns, G. Stell and B. M. Pettitt, *Mol*. *Phys*., **107**, 423 (2009).
[32] H. Nada and J. P. J. M. van der Eerden, *J. Chem*. *Phys*., **118**, 7401 (2003).
[33] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr. *et al*., *J. Am*. *Chem*. *Soc*., **117**, 5179 (1995).
[34] P. A. Kollman, *Acc*. *Chem*. *Res*., **29**, 461 (1996).
[35] J. Wang, P. Cieplak, P. A. Kollman, *J. Comp*. *Chem*., **21**, 1049 (2000).
[36] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong *et al*., *J*. *Comput*. *Chem*., **24**, 1999 (2003).
[37] E. J. Sorin and V. S. Pande, *Biophys*. *J*., **88**, 2472 (2005).
[38] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem*. *Phys*., **79**, 926 (1983).
[39] W. L. Jorgensen and J. D. Madura, *Mol*. *Phys*., **56**, 1381 (1985).
[40] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick *et al*., *J. Chem*. *Phys*., **120**, 9665 (2004).
[41] L. F. Abascal, E. Sanz, R. García Fernández and C. Vega, *J. Chem*. *Phys*., **122**, 234511 (2005).
[42] S. W. Rick, S. J. Stuart and B. J. Berne, *J. Chem*. *Phys*., **101**, 6141 (1994).
[43] M. W. Mahoney and W. L. Jorgensen, *J. Chem*. *Phys*., **112**, 8910 (2000).
[44] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, C. R. Daniels, B. L. Foley and R. J. J. Woods, *J. Comput*. *Chem*., **29**, 622 (2008).
[45] C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys*. *Chem*., **97**, 10269 (1993).
[46] M. Basma, S. Sundara, D. Calgan, T. Vernali and R. J. Woods, *J. Comput*. *Chem*., **11**, 1125 (2001).
[47] D. M. Leitner, M. Havenith and M. Gruebele, *J. Radiol*. *Prot*., **25**, 553 (2006).
[48] S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, X. Heugen *et al*., *Procs*. *Natl*. *Acad*. *Sci*. *USA*, **130**, 2374 (2008).
[49] G. Bussi, D. Donadio and M. Parrinello, *J. Chem*. *Phys*., **126**, 014101 (2007).
[50] J. Gonzalez-Outeriño, K. N. Kirschner, S. Thobhani and R. J. Woods, *Can*. *J. Chem*., **84**, 1 (2006).
[51] G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol*. *Biol*., **7**, 95 (1963).
[52] F. Yu, J. J. Wolff, I. J. Amster and J. H. Prestegard, *J. Am*. *Chem*. *Soc*., **129**, 13288 (2007).
[53] B. M. Sattelle, J. Shakeri, I. S. Roberts and A. Almond, *Carbohyd. Res*., **345**, 291 (2010).
[54] V. Blanchard, F. Chevalier, A. Imberty, B. R. Leeflang, Basappa, K. Sugahara and J. P. Kamerling, *Biochemistry*, **46**, 1167 (2007).
[55] A. Almond and J. K. Sheehan, *Glycobiology*, **10**, 329 (2000).
[56] M. A. Rodríguez-Carvajal, A. Imberty and S. Pérez, *Biopolymers*, **69**, 15 (2003).
[57] K. Sugahara, Y. Tanaka and S. Yamada, *Glycoconjugate J*., **13**, 609 (1996).
[58] S. Yamada, K. Yoshida, M. Sugiura and K. Sugahara, *J. Biochem*., **112**, 440 (1992).
[59] A. Kinoshita-Toyoda, S. Yamada, S. M. Haslam, K. H. Khoo, M. Sugiura *et al*., *Biochemistry*, **43**, 11063 (2004).

[60] J. J. Cael, W. T. Winter and S. Arnott, *J. Mol. Biol*., **125**, 21 (1978).
[61] W. T. Winter, S. Arnott, D. H. Isaac and E. D. Atkins, *J. Mol. Biol*., **125**, 1 (1978).
[62] E. D. Atkins, D. Meader and J. E. Scott, *Int. J. Biol. Macromol*., **2**, 318 (1980).
[63] Y. Zhong, B. A. Bauer and S. Patel, *J. Comput. Chem*., **32**, 3339 (2011).
[64] S. W. Rick, *J. Chem. Phys*., **120**, 6085 (2004).
[65] A. Luzar and D. Chandler, *Nature*, **379**, 55 (1996).
[66] F. N. Keutsch and R. J. Saykally, *Procs. Natl. Acad. Sci. USA*, **98**, 10533 (2001).
[67] V. P. Voloshin and Y. I. Naberukhin, *J. Struct. Chem*., **50**, 78 (2009).
[68]