ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2012, *57*(12), 1013-1023

PAPERS

THE INFLUENCE OF PARTIAL CHARGE CALCULATION PROTOCOL ON THE SULFATED MONOSACCHARIDE CONFORMATION AND DYNAMICS**

Andrei NEAMTU,[a,b,*] Xenia PATRAS,[b] Monica NEAMTU,[b]
Anca OPREA,[a] Tudor PETREUS[b] and Nicolae HURDUC[c]

[a] "Petru Poni" Institute of Macromolecular Chemistry, 41 A, Gr. Ghica Voda Alley, 700487, Iaşi, Roumania
[b] Center for the Study and Therapy of Pain (CSTD), "Gr. T. Popa" University of Medicine and Pharmacy, 16, University Street, 700115, Iaşi, Roumania
[c] "Gheorghe Asachi" Techical University of Iasi, 73 Prof. Dimitrie Mangeron Str., 700050, Iaşi, Roumania

Received February 18, 2011

Modeling the biologically active sulfated saccharides is an important tool for understanding the molecular mechanism of their biological activity. The present study aims to provide insights on partial charge assignment methods for sulfated saccharides. *Ab initio* calculations at HF/6-31G(d), HF/6-31+G(d,p), HF/6-31++G(d,p) levels and molecular dynamics simulations have been performed in order to evaluate the impact of the theory level used for charge derivation on the geometry and solvent interactions of a model monosaccharide. The results show that higher quantum levels of theory does not necessary give better results in the quality of the obtained charges. The calculated geometry and the hydrogen bonding network properties are in agreement with the experimental crystallographic and NMR data.

INTRODUCTION

One class of important biologically active saccharides is represented by the sulfated glycosaminoglycans, the most well-known being chondroitin sulfate and heparin. Glycosaminoglycans (GAGs) are natural polymers made of specific repeating disaccharide units in which one sugar is uronic acid and the other is either N-acetylglucosamine or N-acetylgalactosamine. Due to their variable distribution of negatively charged sulfate groups, they display unique properties which are indispensable for their biological functions. They may act as structural components of connective tissue and the extracellular matrix, as co-factors in cell surface receptor activation,[1,2] as agent in blood clotting homeostasis, angiogenesis[3] or immune defense.[4] Sulfated GAGs were among the first carbohydrates studied by NMR spectroscopy (^{1}H and ^{13}C NMR),[5-7] the most important structural information being extracted through three bond proton-proton coupling constants[8,9] and chemical shifts.[10,11] Crystallographic studies have also been performed on sulfated monosaccharides[12-16] and oligosaccharide/protein complexes.[17-20] Despite the large pool of experimental data regarding this class of compounds there is still much to be understood especially regarding the specificity of interaction with biologically relevant macromolecules like proteins. This has to be done by establishing the relationship between their chemistry, microscopic three dimensional structure and biological functions.[21] The lack of high resolution structures for the large variety of sulfated GAGs limits the progress in this field.

* Corresponding author: neamtuandrei@gmail.com
** Supplementary material on http: //web.icf.ro/rrch/ or http: revroum.getion.ro

Molecular modeling methods represent a major tool of theoretical investigations and are routinely applied for simulation of biological macromolecules,[22-24] giving useful insights at atomic-molecular level with time resolution on picosecond scales that are not readily accessible by experimental methods. Empirical force field methods use parameters obtained by comparison with experimental data or from higher level of theory like quantum mechanics.[25] For the description of electrostatic interactions these models use partial point charges situated at the nuclei positions. The design of accurate charges is thus an important requirement for a good force field especially when charged species are present. To derive partial charges, most of the nowadays methods employ quantum mechanics using Mulliken population analysis,[26] Lödwin population analysis[27] or fitting to quantum mechanical molecular electrostatic potential (MEP).[28,29] The methods of MEP fitting are preferable as they offer several advantages when applied to polar molecules.[30] The literature describes several approaches for the derivation of partial charges of sulfated monosaccharides using different level of quantum mechanical theory for the computation of the electrostatic potential around the studied molecules. These include *ab initio* Hartree-Fock and post Hartree-Fock methods with basis sets of more or less basis functions:[21,31,32] HF/3-21+G, HF/6-31G(d), HF/6-31+G(d,p), MP2/6-31+G(d,p), B3LYP/6-31G(d,p). It is however unclear to what extent the different levels of theory used for charge assignment affects the dynamics and the conformational preference of sulfated saccharides probed by molecular dynamics simulations. This is important not only for designing improved force fields for GAGs simulation but also if we take into account that higher level quantum mechanical methods require by far more computing power which may not be necessary for a satisfactory description of the properties of these molecules.

In the absence of a systematic study on this aspect, the present paper aims to evaluate the influence of different levels of theory for charge assignment on the simulated conformational preference and dynamics of the sulfated monosaccharide using 1-O-methyl-4-sulfated-*β*-*N*-acetyl-galactosamine (1-O-Me-*β*-GalNAc(4S)) (Fig. 1) as a model molecule.

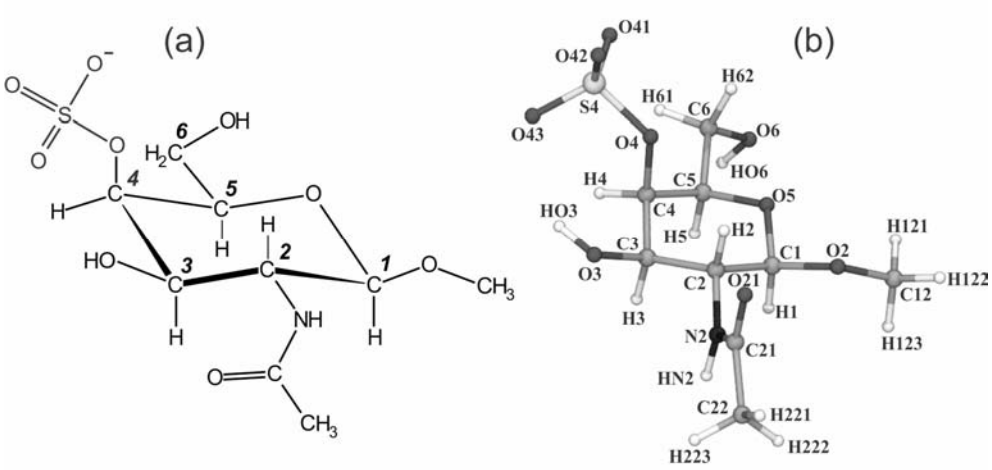

Fig. 1 – The structure (a) and atom nomenclature (b) of 1-O-methyl-4-sulfated *β* -*N*-acetylgalactosamine.

METHODS

Molecular Dynamics simulations has been performed on 1-O-Me-*β*-GalNAc(4S) using partial atomic charges computed by fitting the quantum mechanical derived Molecular Electrostatic Potential (MEP) at gradually increasing levels of theory. 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,[33] we have chosen instead of a single molecular conformation approach, an Ensemble Average (EA) method.[30] 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 can be obtained from a conformational search or may be extracted from a molecular dynamics trajectory, the latter method having the advantage that the conformations obtained in this way reflect the conformational preference of the studied

molecule at the simulation temperature and in the presence of the solvent (denoted in the following as SEA – Solvated Ensemble Average). The conformations used to compute the charges of the 1-O-Me-*β*-GalNAc(4S) monosaccharide were generated from an initial molecular dynamics simulation as explained below. One drawback of this approach is that the partial charges needed for the initial molecular dynamics simulation are not known. To overcome this problem methyl sulfate was selected as a model compound to determine an initial guess of the charges for the sulfate group while using the GLYCAM_06c partial charges set for the rest of the 1-O-Me-*β*-GalNAc molecule.

Charge derivation procedure and quantum mechanical (QM) calculations

The partial charges of the methyl sulfate and 1-O-Me-*β*-GalNAc(4S) were computed applying the RESP method of Bayly and co-workers[34] and CHELPG algorithm[28] with 1 fitting stage and an weighting factor of 0.01. Prior to all MEP quantum evaluations the geometries of the modeled molecules were optimized using the Quadratic Approximation (QA) algorithm, taking as the convergence criteria the largest component of the energy gradient to be less than 10^{-8} Hartrees/Bohr (Ha/b) for the small molecule of methyl sulfate and respectively 10^{-3} Ha/b for GalNAc(4S) monosaccharide. For the two-electron integrals the Direct-SCF method was used with a cutoff of 10^{-11}. The Huckel method was selected for the initial orbital guess. The SCF was considered to be converged when the density change between two consecutive SCF cycles was less than 10^{-8}. All the geometry optimizations have been performed at the Hartree-Fock *ab initio* level using the Pople split-valence double zeta basis set approach with 6 Gaussian functions for core orbitals and 3 plus 1 Gaussians for the valence orbitals; d-type polarization functions for the heavy (non hydrogen) atoms were added to the basis set (HF/6-31G(d)). The Quadratic Aproximation (QA) algorithm was used to locate the optimized geometry with a gradient convergence tolerance of 10^{-3} Hartree/Bohr. In order to preserve the molecular conformation during the optimization step all the rotatable exocyclic torsional angles were constraint to their particular values as resulted from the molecular dynamics simulation. To evaluate if the optimization procedure reached a true minima and not a saddle point on the potential energy surface vibrational analysis was performed on the methyl sulfate optimized structure. The positive value for the frequency of the normal vibration of minimum energy (ν_0 = 79.7 cm^{-1}) shows that a true minima for methyl sulfate molecule was reached in the optimization process. Due to the fact that the optimization process in the case of GalNAc(4S) monosaccharide is performed in the presence of applied constraints, in this case a more direct method was applied. Taking into account that the geometry optimization is done in a subspace of the full coordinate space, at the end of optimization process the force on the constraint is very likely not zero and the final structure is not a stationary point in the full cartesian space. Vibrational analysis in the GAMES-US package is always performed in the full cartesian space and so it is meaningless in this case. A less arbitrary way of checking whether a true minima or a saddle point was reached is to add random displacement to the atom coordinates of the optimized structure and repeat the constrained optimization. As long as the displacements are truly random, the structure will optimize away from a saddle point, but will return to a (constrained) minumum. This procedure was applied for all of the 25 conformers and the data obtained shows the reach of true minima in each case in the course of the initial optimization process (see Table 1 and Fig. 1 in "Supplementary Material"). *Ab initio* calculations using different basis sets had been performed on the molecules included in the study as described below. For the initial charge assignment for the sulfate moiety the methyl sulfate was considered as the reference molecule and the quantum HF/cc-pVTZ,[35] calculated molecular electrostatic potential was fitted based on electrostatic point charges located at the nuclei positions using the RESP procedure. The correlation-consistent polarized cc-pVTZ basis set incorporates a triple zeta valence split model. It is designed to systematically converge to the complete-basis-set (CBS) limit using extrapolation techniques. The selection of the cc-pVTZ basis set in the case of methyl sulfate was suggested by the GLYCAM_06c developing methodology,[36] maintaining in this way the consistency with this force field. For the 1-O-methyl-*β*-*N*–GalNac(4S) conformers three *ab initio* levels of theory were used: (1) the Hartree-Fock method with the 6-31G(d) basis set (HF/6-31G(d)); (2) the Hartree-Fock method with the 6-31+G(d,p) basis set (HF/6-31+G(d,p)) which adds, supplementary to the

previous basis set, one polarization function on each hydrogen atom and one diffuse function ("+" sign in the usual nomenclature) on each heavy atom; (3) the Hartree-Fock with the 6-31++G(d,p) basis set (HF/6-31++G(d,p)) which consequently adds, compared with the previous one, diffuse functions on hydrogen atoms too. Quantum calculations were done in parallel on a Dell Precision 690 Workstation with 8GB of RAM and 8 compute cores with the GAMESS-US software[37] and R.E.D. v3 suite (http://q4md-forcefieldtools.org/RED/).

Molecular dynamics simulations

The parameters that describe the bonded intramolecular interactions (bonds, angles and dihedrals) and the Lenard-Jones non-bonded interactions of the 1-O-Me-*β*-GalNAc(4S) molecule were taken directly from the GLYCAM_06c parameter set, except for the dihedral terms of the sulfate moiety. The following GLYCAM atom types were associated with the sulfate group atoms: (S) for the sulphur (S4 – Fig. 1a); (O2) for the terminal oxygen atoms (O41, O42, O43) and respectively (OS) for the oxygen atom bonded to the sugar cycle (O4). The rest of the atoms in the GalNac(4S) monosaccharide had usual GLYCAM atom types. The parameters for the bonds and angles were taken from the GLYCAM parameters set and the dihedral parameters for the O2-S-OS-CG, S-OS-CG-H1 and S-OS-CG-CG torsional angles were reparameterized (data not shown) by reproducing the quantum computed energy profiles along the two mentioned dihedral coordinates following the procedure described in the GLYCAM paper.[36] For the sulfate group the methyl sulfate computed charges were used.

The charge of the C4 atom was adjusted to obtain an overall charge of -1 for the entire 1-O-Me-*β*-GalNAc(4S) molecule. Each SEA charge set was computed by RESP procedure on a collection of conformers extracted from an initial molecular dynamics trajectory. The initial (SEA) molecular dynamics simulation was started with a configuration obtained by solvating a single 1-O-Me-*β*-GalNAc(4S) molecule with 2163 water molecules inside a cubic box with the edge length of 40Å. One Na^+ ion was added to the simulation box for overall charge neutralization. Prior to the production run the system was equilibrated by slowly heating from 5K to 300K in 50ps, followed by cooling back to 5K in another 50ps. In the final step of the equilibration the system was again heated to 300K in 150ps. Due to the heating/cooling scheme the solvent can relax and adapt to the surface of the solute molecule, eliminating this way the local stresses created during the building step which can introduce artifact deformations of the solute during the subsequent production runs. After the equilibration stage the system was subjected to a production simulation for 50ns with an integration time step of 0.001ps. During the production run the system was coupled to a thermostat at 300K and to a pressure bath of 1atm. From the obtained trajectory, 25 conformations (Fig. 2) were extracted evenly at every 2ns and subjected to the RESP charge computation procedure. Prior to MEP computation an energy minimization was performed on each of the conformation at the HF/6-31G(d) level using GAMESS-US software. In order to preserve the molecular conformation during the optimization step all the rotatable exocyclic torsional angles were constraint to their particular values as resulted from the molecular dynamics simulation.

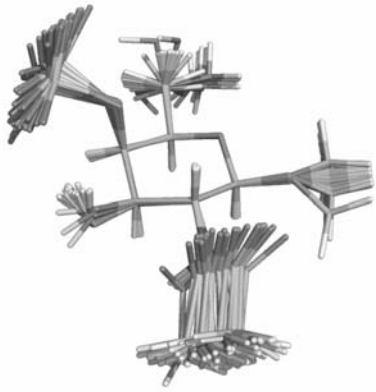

Fig. 2 – The statistical ensemble of 25 conformers on which the partial electrostatic charges were computed.

For each of the three charge sets obtained (SEA-HF/6-31G(d), SEA-HF/6-31+G(d,p) and SEA-HF/6-31++G(d,p)), molecular dynamics simulations were performed following the same protocol as above, with a final production run of 50ns. All molecular dynamics simulations were performed in parallel using GROMACS 4.0.5[38] on a HPC Dell cluster (PowerEdge 1950 nodes) with 64 computing cores. Molecular pictures were produced using VMD[39] and PyMol (Schrödinger, LLC).

RESULTS AND DISCUSSION

The atomic partial charges obtained using the Restrained Electrostatic Potential Fit (RESP)

procedure at different quantum mechanical levels for the 1-O-Me-*β*-GalNAc(4S) molecule are presented in Table 1. Along with charges, there are also given the times necessary for the calculation of the molecular electrostatic potential around one conformation of the molecule as a measure of the computational effort needed in each case. It can be seen from Table 1 that the biggest charge difference on the corresponding atoms is encountered between the HF/6-31G(d) and HF/6-31+G(d,p) sets. The charge differences are lower than 0.14e (absolute value) which is encountered for the sulphur atom. For the rest of the atoms the differences are under 0.055e. The Root Mean Squared Deviations (RMSD) between the sets of charges are 0.0323e for HF/6-31G(d) vs. HF/6-31+G(d,p) and respectively 0.001 for HF/6-31+G(d,p) vs. HF/6-31++G(d,p). This suggests that the addition of the diffuse functions set on heavy atoms influences at a certain degree the fitted point charge values, while the further addition to the basis set of diffuse functions on hydrogen atoms does not improve the quality of the results. The molecule under study is an anion carrying a net negative charge of -1. The addition of diffuse functions, which are more 'flat' Gaussian functions that more accurately represent the 'tail' portion of the wave function, to the basis set was expected to better represent the electronic charge density near the charged group.

Table 1

RESP computed SEA partial charges (in units of *e*) for the 1-O-Me-*β*-GalNAc(4S) monosaccharide atoms at different levels of quantum mechanical description[a]

Atom name	HF/6-31G(d) [b]336 bas. funct. [c](1'40s/conf)	Std. dev.	HF/6-31+G(d,p) [b]464 bas. funct. (6'26s/conf)	Std. dev.	HF/6-31++G(d,p) [b]480 bas. funct. (8'32s/conf)	Std. dev.
C12	0.2287	± 0.0090	0.2187	± 0.0108	0.2173	± 0.0107
O2	-0.3922	± 0.0264	-0.3971	± 0.0361	-0.3949	± 0.0361
C1	0.1725	± 0.0297	0.2154	± 0.0378	0.2144	± 0.0381
O5	-0.3743	± 0.0553	-0.3981	± 0.0659	-0.3971	± 0.0659
C5	0.1341	± 0.0637	0.1380	± 0.0701	0.1385	± 0.0701
C4	0.2212	± 0.0487	0.2210	± 0.0581	0.2202	± 0.0582
C3	0.1684	± 0.0557	0.1910	± 0.0682	0.1910	± 0.0683
O3	-0.6574	± 0.0331	-0.6793	± 0.0461	-0.6783	± 0.0463
HO3	0.4215	± 0.0140	0.4305	± 0.0189	0.4299	± 0.0190
O4	-0.4708	± 0.0305	-0.4897	± 0.0339	-0.4893	± 0.0340
S4	1.2820	± 0.0190	1.4229	± 0.0216	1.4230	± 0.0216
O41	-0.6516	± 0.0087	-0.6953	± 0.0097	-0.6953	± 0.0097
O42	-0.6516	± 0.0087	-0.6953	± 0.0097	-0.6953	± 0.0097
O43	-0.6516	± 0.0087	-0.6953	± 0.0097	-0.6953	± 0.0097
C6	0.3519	± 0.0462	0.3878	± 0.0494	0.3863	± 0.0496
O6	-0.7110	± 0.0252	-0.7466	± 0.0306	-0.7438	± 0.0304
HO6	0.4053	± 0.0179	0.4127	± 0.0205	0.4109	± 0.0206
C2	0.5629	± 0.0619	0.5667	± 0.0677	0.5650	± 0.0676
N2	-0.8223	± 0.0503	-0.8461	± 0.0645	-0.8434	± 0.0645
HN2	0.3297	± 0.0203	0.3259	± 0.0278	0.3244	± 0.0280
C21	0.7097	± 0.0368	0.7649	± 0.0485	0.7642	± 0.0487
O21	-0.5961	± 0.0166	-0.6372	± 0.0208	-0.6371	± 0.0209
C22	-0.0089	± 0.0088	-0.0155	± 0.0111	-0.0153	± 0.0111

[a] All non-polar hydrogens (bonded to carbon atoms) were constrained to have a zero charge to maintain the compatibility with the GLYCAM force field

[b] Number of Gaussian basis functions used to describe the 1-O-methyl-β-*N*–GalNAc(4S) molecule

[c] The time needed for MEP calculation per conformation using the parallel algorithm on 7 computing cores

The standard deviations were computed on the statistical ensemble of 25 conformations for each atom. The values shows fluctuations between 0.01e and 0.07e, the larger variability being observed for the pyranose ring atoms and the smallest for the atoms which lie farther from the ring group, like the oxygen atoms of the sulfate moiety. This can be explained by the fact that the procedure for the charge fitting of the CHELPG algorithm utilize as optimization target the potential calculated on a grid of points situated inside a certain distance interval near the molecular surface. The ring atoms, due to the larger number of chemical bonds with the neighboring heavy atoms, participate to a lesser extent to the formation of the geometrical molecular surface, in this way their charge being fitted with less accuracy. This drawback of the method is known in literature, especially for the molecules with buried atoms under the molecular surface.[37]

The time needed for the MEP computation of a single conformation rapidly increases with the number of basis functions, the most complex method used here being five times slower than the simplest one. This is important in the case of ensemble averaged charges where the charge fitting is done over multiple conformations.

To further analyze the influence of the differences found in charge values between the HF/6-31G(d) set and the HF/6-31+G(d,p), it has to be established if these differences are large enough to modify the dynamical behavior of the studied molecules during simulations. The analysis of the molecular dynamics trajectories performed with the three sets of charges can give us information about the hydrogen bonds of the solvated molecule, solvent organization near the molecular surface of the solute and conformational preferences of the analyzed molecule. All these depend at a certain degree on the charge model used to describe the electrostatic non-bonded interactions.

The hydrogen bonding network organized between the saccharide molecule and the surrounding solvent plays an important role in its conformational stabilization, hydrogen bonds (H-bonds) being one of the most important among the forces that drive this process. Because most of the force fields used in molecular mechanics do not contain explicit hydrogen bonding terms, instead relying upon electrostatic and van der Waals interactions to reproduce hydrogen bonding[25], the analysis of the H-bond statistics during the molecular dynamics simulations can be a sensitive indicator of the differences in the quality of the three sets of partial charges considered. Fig. 3 presents the distribution of the donor – acceptor distance and the donor – hydrogen – acceptor angle for the hydrogen bonds established between the oxygen atoms of the sulfate group (O41, O42, O43) and the water molecules, mediated over the entire molecular dynamics trajectories (50ns each). Most of the solvent molecules participate to the hydrogen bonding with the $-SO_3^-$ oxygen atoms at a mean distance of 2.7Å between the centers of the water oxygen and the acceptor (the maximum of the distribution curve in Figure 3a). The angle of the hydrogen bond, expressed as the angle defined by the three atoms that participate to the bond, has a maximum at 10°. The graphs shows a very good agreement between the H-bond characteristics (both distance and angle) computed successively with the three analyzed charge sets, proving that the differences in the charge values given by the HF/6-31G(d) and HF/6-31+G(d,p) sets do not perturb the hydration layer involved in this type of interaction.

This result is further sustained by the H-bond life time plots and the radial distribution functions (RDF) of the water molecules around the monosaccharide molecule (Fig. 4). The number of hydrogen bonds between the terminal oxygen atoms of $-O-SO_3^-$ group and water oxygen is rapidly decreasing with increasing uninterrupted life times (Fig. 4a). The obtained mean value for H-bond lifetime is $\tau_{1/2} = 9$ ps which is in good agreement with the experimentally measured ones (10-15ps) by time resolved infrared spectroscopy.[40,41] The life time curves show no difference between the different charges sets used. Fig. 4b depicts the radial distribution functions of the water oxygens (OW) computed from the molecular dynamics trajectories, comparatively for the three charge sets used. Two reference atom types were selected as centers for these RDFs: one of the terminal negatively charged oxygen atoms of $-O-SO_3^-$ group (O41, O42 and O43) and O3 oxygen atom of the hydroxyl in position 3. Both RDFs show an intense maximum at approximately 2.7Å which coincides with the maximum of the distribution curve in Fig. 4a. This maximum defines the first layer of hydration (first neighboring water molecules) of the terminal $-SO_3^-$ group. The second and the third layers of hydration are detectable at 4.9Å and 7Å respectively for the O43-OW pair. The second layer of hydration for the –OH hydroxyl group is disrupted by the strong interaction of the negatively charged neighboring sulfate group with the solvent. Again, all the three charge sets give similar results in very good quantitative agreement with each other.

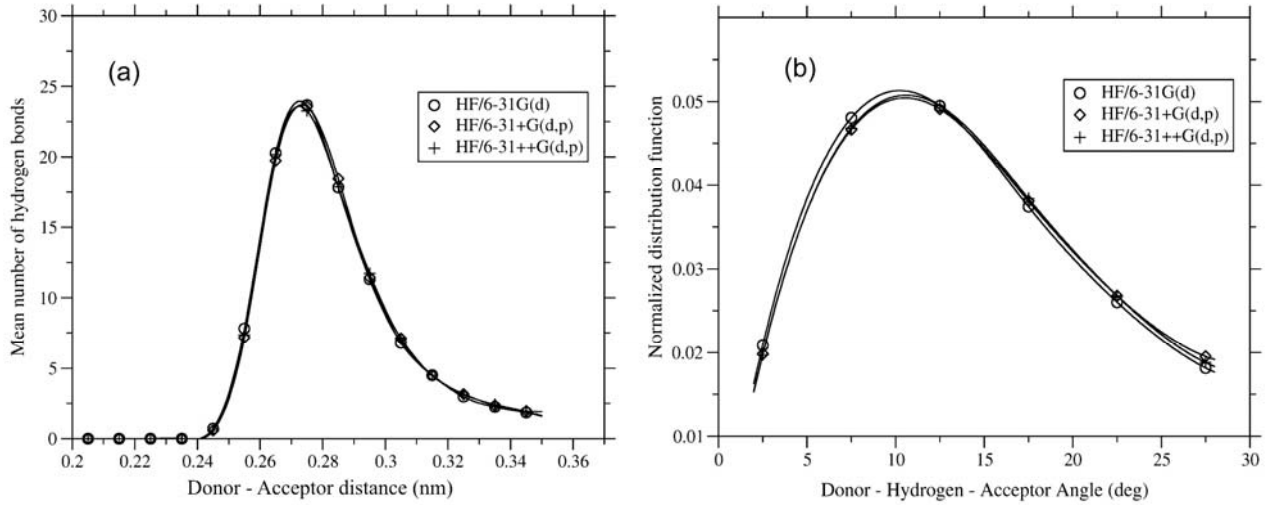

Fig. 3 – Distribution functions of (a) hydrogen – acceptor distance and (b) donor – hydrogen – acceptor angle as obtained from the molecular dynamics simulations for the $-O-SO_3^- \cdots H-O-H$ hydrogen bond interaction in 1-O-Me-β-GalNAc(4S).

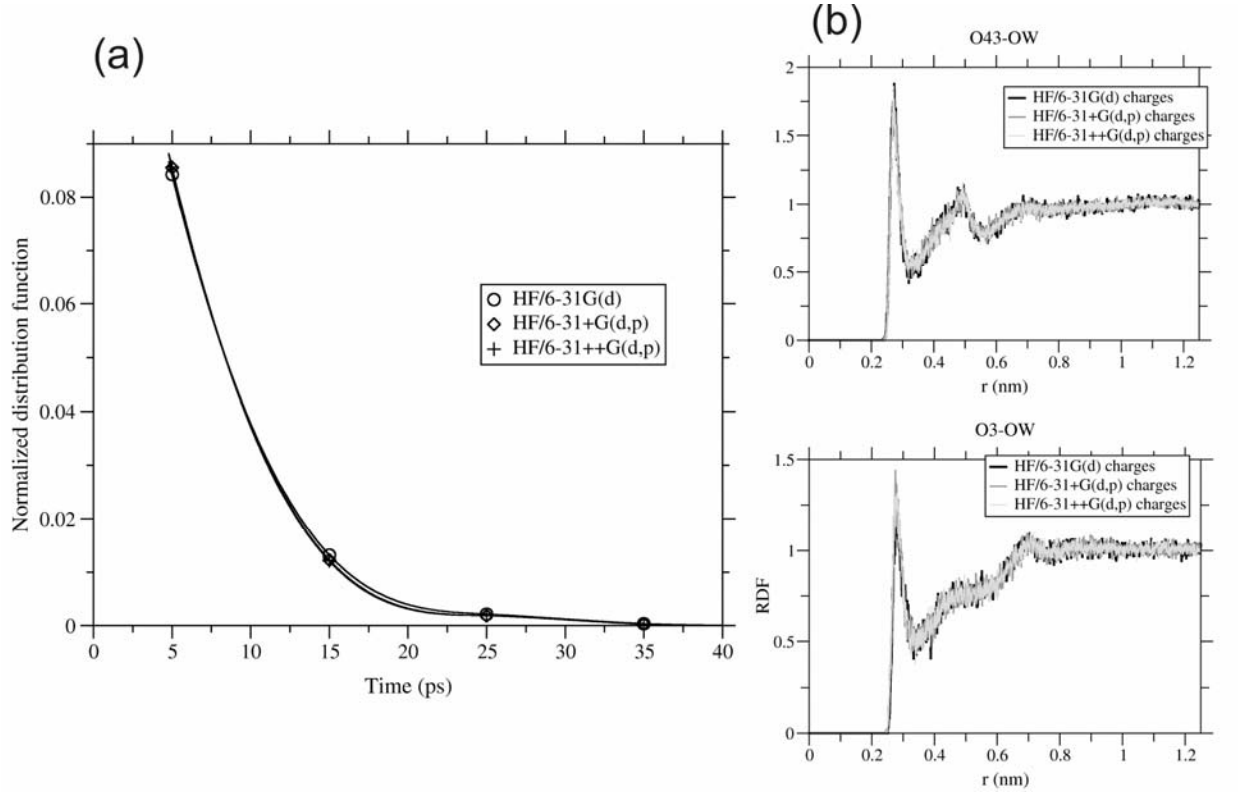

Fig. 4 – (a) The distribution functions for the hydrogen bond life time between the terminal oxygens of $-O-SO_3^-$ group and the solvent molecules; (b) radial distribution functions for the water oxygens around the O3 atom and O43 sulfate terminal oxygen.

The conformational preference of the monosaccharide molecule during the molecular dynamics simulation can also be a useful indicative of the influence that the charge set choice may have on the molecule's behavior, especially when hydrogen bonding is present intramolecularly. The conformational space of a molecule is defined by its rotatable torsion angles and is usually analyzed by representing the time evolution of some selected ones in the course of the simulation.[31] A better representation of the conformational space explored by the molecule can be obtained by using Free Energy Landscape (FEL) representations in two dimensions. Figs. 4 and 5 show, for the three charge sets analyzed, the FELs obtained from the molecular dynamics simulations, together with the corresponding frequency of microstates histograms (probability) to reveal the relative population of the detected conformational basins. This is done for two pairs of dihedral angles, one pair characterizing the

shape of the pyranozic ring and the other one the relative orientation of the sulfate terminal group to the neighboring equatorial groups –OH in position 3 and $–CH_2OH$ in position 5. Fig. 5 depicts the free energy landscapes as isocontour lines in the conformational space defined by the ν2 (C1-C2-C3-C4) and ν5 (C4-C5-O5-C1) dihedral angles. The ν2 and ν5 angles define the shape of the pyranosic ring. The FEL profile was obtained by applying the Boltzmann formula:

$$\Delta G = -k_B T \ln \frac{P_{i,j}}{P_{max}} \tag{1}$$

to the frequency matrix calculated as a histogram of the microstates (defined by particular values of the angles ν2 and ν5). In the formula (1), 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 ν2×ν5 plane and P_{max} is the maximum frequency over the entire range of ν2, ν5. The FEL landscape shows two basins. The basin 'C' characterized by mean central values of -60° for ν2 angle and respectively 60° for ν5 is by far more populated than the second one characterized by the mean central values of (-40°, -35°) for (ν2, ν5) (basin'S'). The basin 'C' corresponds to the $^4C_1$ shape of the ring which *N*-actylgalactosamine usually adopts in solution as revealed by NMR[42] and crystallographic[17-20] studies. The values of the dihedral angles for the second one (basin 'S') are specific for the skew '*S'* shape which is known to be energetically disfavored compared with the $^4C_1$ form. The form adopted by the pyranose ring is $^1S_3$ with the C1 carbon atom being above the plane defined by the C2-C4-C5-O5 atoms and C3 below it. The computed free energy differences between the $^4C_1$ and $^1S_3$ are 2.55 kcal/mol for HF/6-31G(d) charges compared with 2.23 kcal/mol for HF/6-31+G(d,p) and HF/6-31++G(d,p) charges. The free energy difference between HF/6-31G(d) and HF/6-31+G(d,p) is 0.32 kcal/mol which is the same order of magnitude as thermal energy at usual temperatures ($1RT$ = 0.59 kcal/mol at T = 300K; R is the gas constant) and cannot this way significantly influence the relative conformational behavior of the simulated monosaccharide.

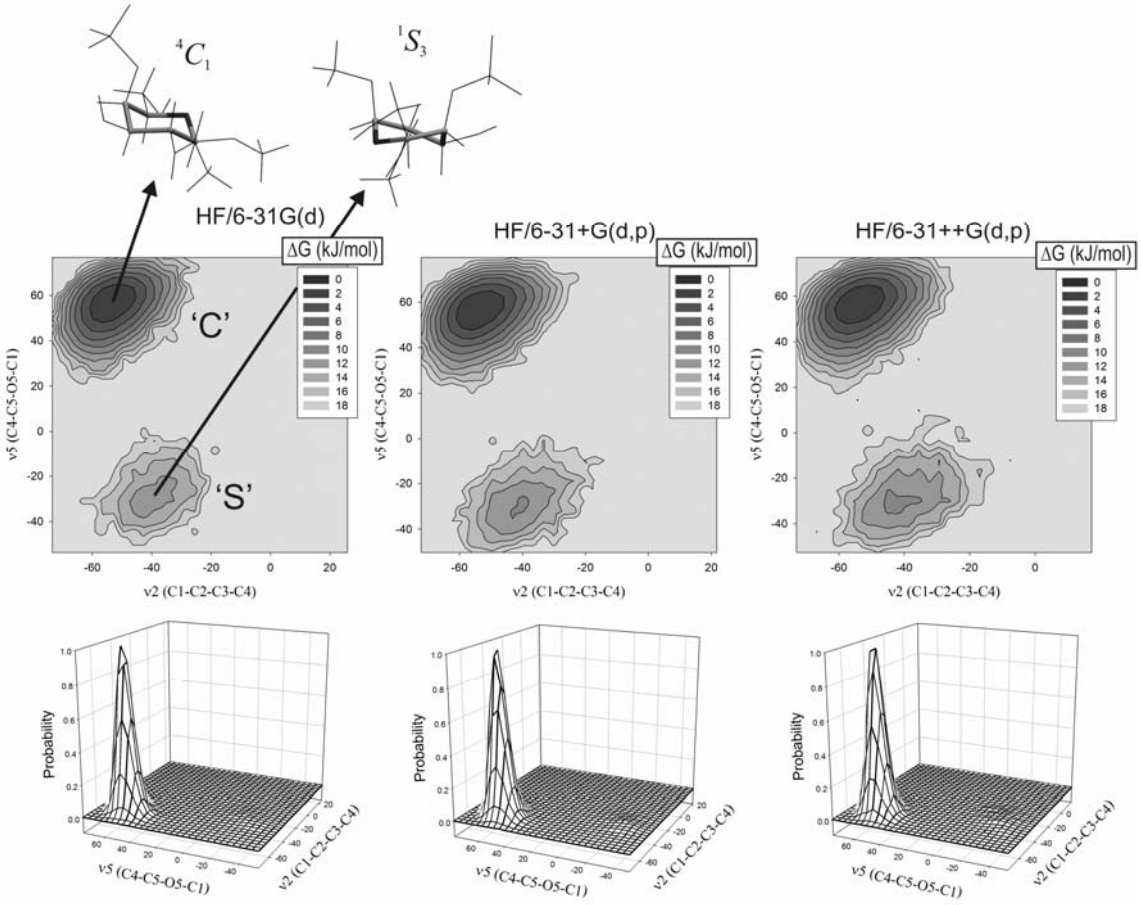

Fig. 5 – Free energy landscapes and the corresponding frequency histograms in the (ν2, ν5) conformational plane for the three analyzed charge sets ('C' – chair, 'S' – skew).

The second pair of dihedrals, which characterize the relative orientation of the $-O-SO_3^-$ moiety relative to its neighboring $-CH_2$-OH in position 5, comprises the χ4 (S4-O4-C4-C3) and χ5(O6-C6-C5-C4) dihedral angles. The (χ4, χ5) pair is important as it can give information about the intramolecular hydrogen bonding involving the sulfate group, which highly depends on the assigned charges in the classical model. Three conformational basins can be identified on the FEL plot (obtained using the same procedure as above) in (χ4, χ5) plane (Fig. 6). The 'T' ('trans' of χ5) basin with mean central values of (-110º, -180º) for the (χ4, χ5) angles; the 'ST−' ('staggered minus') basin with (-110º, -60º) mean values for the (χ4, χ5) pair and the 'ST+' ('staggered plus') basin with (-110º, 75º) mean values for the (χ4, χ5) pair. The most favorable configuration for the -O-SO_3^- group is that with the χ4 angle around -110º which favors a hydrogen bond between its negative charged terminal oxygen and either the –OH group in position 3 (basin 'T') or with the -CH_2-OH in position 5 (basins 'ST-' and 'ST+'). This is in very good agreement with experimental crystallographic determined structures.[17-20] From the correspondent probability plots it can be seen that the most favorable configuration is when the hydrogen bond is formed between the sulfate terminal oxygen and the –OH group in position 3.

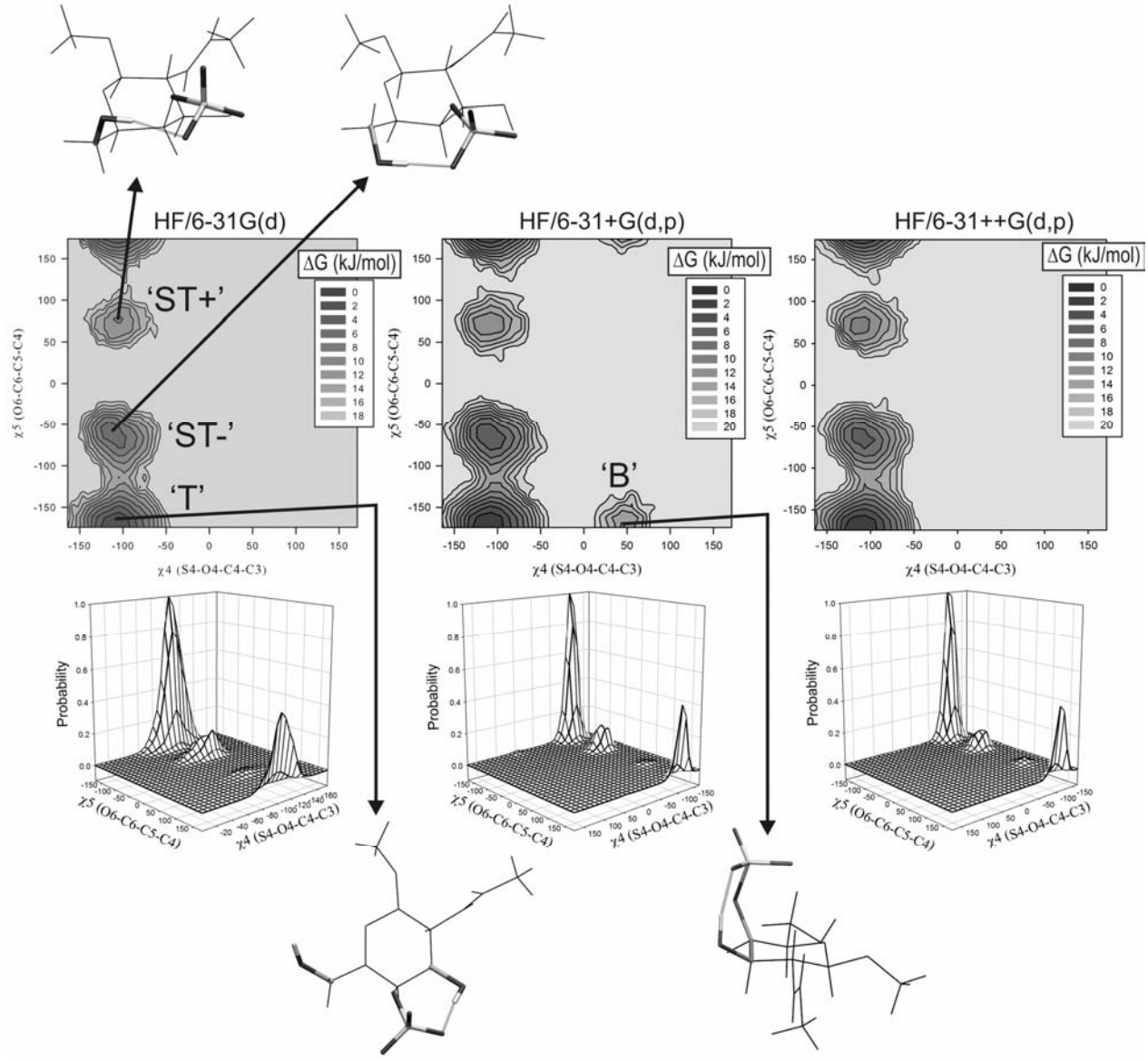

Fig. 6 – Free energy landscapes and the corresponding frequency histograms in the (χ4, χ5) conformational plane for the three analyzed charge sets ('T' – trans, 'ST' – staggered, 'B' – bent).

The hydrogen bonding with the $–CH_2$-OH group can be realized with the angle χ5 having values either around -60º or 75º, the first configuration being more probable than the latter one. For the HF/6-31+G(d,p) derived charges, a second minima for the χ4 angle around 60º can be observed. This corresponds to a conformational basin ('B') in which the $-O-SO_3^-$ group is bend over the ring. A hydrogen bond between the terminal sulfate oxygen and the –OH group in position 3 is stabilizing this conformation. It is however highly unfavorable energetically and is not present in the conformational space of the other two cases (HF/6-31G(d) and HF/6-31++G(d,p) charge sets). Due to the low probability of this state, its occurrence is presumably rare enough to be observed in the simulations time frame (50ns) other than accidentally. The data presented in Figs. 4 and 5 show that the studied molecular models, based on the three charge sets, detect the same preferred conformations of the monosaccharide molecule and their relative populations.

CONCLUSIONS

Quantum mechanical calculations and molecular dynamics simulations were performed on solvated 1-O-methyl-*β*-*N*–GalNAc(4S) monosaccharide molecules in order to evaluate the impact of partial atomic charge derivation procedure on the solvent organization around sulfated monosaccharides and on the conformational behavior of this class of substances. The results presented here show small differences in the atomic charges computed at different quantum mechanical levels, especially when diffuse functions are added for heavy atoms. However, when these different charge sets are actually used for molecular dynamics simulations there are no detectable differences in the hydrogen bonding network organization and dynamics. The predicted ring shape is in agreement with the experimental NMR and crystallographic data for these compounds. However, small differences appear for the conformational preference of the analyzed monosaccharide with the skew shape of the pyranozic ring becoming slightly more favored when the level of quantum description used for charge derivation is higher. Due to their low magnitude, these differences do not exert an important effect at usual temperatures. It can be concluded from the data presented here that for sulfated monosaccharides, like sulfated *N*-acetylgalactosamine, the increase in the complexity of the basis set beyond 6-31G(d), by adding diffuse and polarization function sets, does not give an significant improvement over the partial charge assignment. This is important because the time needed for a quantum mechanical MEP computation of the studied molecules rapidly increase with the basis set functions number. The present study gives solvated ensemble averaged charge sets for the 1-O-methyl-β-*N*–GalNAc(4S) which can be used for further molecular dynamics simulations involving GAGs.

It has to be further clarified however to what extent the small differences observed in the conformational behavior of the sulfated monosaccharides really affect the overall structural characteristics in more complex systems such as isolated polymer chains or networks. Even the simulation time scales presented here (tenths of nanoseconds) are sufficiently large for isolated monosaccharides, the influence of the partial charge derivation protocol on more complex oligo- and polysaccharide macromolecules has to be analyzed on time scales at least one order of magnitude higher.

The conclusions presented here for charge derivation must be considered only for *ab initio* methods with split-valence double zeta basis sets. The case of higher 'post Hartree-Fock' methods which take into consideration the electron correlation effects, like the Density Functional Theory or the Møller–Plesset perturbation theory, has to be equivalently evaluated.

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

REFERENCES

1. L. Pellegrini, F. Burke, F. von Delft., B. Mulloy and T.L. Blundell, *Nature*, **2000**, *407*, 1029-1034.
2. J. Schlessinger, A.N. Plotnikov, O.A. Ibrahimi, A.V. Eliseenkova, B.K. Yeh, A. Yayon, R.J. Linhardt and M. Mohammadi, *Moll. Cell.*, **2000**, *6*, 743-750.
3. B. Mulloy, M.J. Forster, C. Jones and D.B. Davies, *Biochem. J.*, **2000**, *293*, 849-858.
4. H.B. Nader, S.F. Chavante, E.A. dos-Santos, F.W. Oliveira, J.F. de-Paiva, S.M.B. Jerônimo, G.F. Medeiros, L.R.D. de-Abreu, E.L. Leite, J.F. de-Sousa-Filho, R.A.B. Castro, L. Toma, I.L.S. Tersariol, M.A. Porcionatto and C.P. Dietrich, *Braz. J. Med. Biol. Res.*, **1999**, *32*, 529-538.
5. A. S. Perlin, D.M. Mackie and C.P. Dietrich, *Carbohydr. Res.*, **1971**, *18*, 185-194.
6. B. Casu, P. Oreste, G. Torri, G. Zopetti, J. Choay, J.C. Lormeau and Petitou M., *Biochem J.*, **1981**, *197*, 599-609.

7. B. Mulloy and E.A. Johnson, *Carbohydr. Res.*, **1987,** *170*, 151-165.
8. H.P. Wessel and S. Bartsch, *Carbohydr. Res.*, **1995,** *274*, 1-9.
9. K. Bock, *Annu. Rep. NMR Spectr.*, **1984**, *13*, 1-57.
10. V. Bossennec, M. Petitou and B. Perly, *Biochem J.*, **1990**, *267*, 625-630.
11. P.N. Sanderson, T.N. Huckerby and I.A. Nieduszynski, *Biochem J.*, **1989**, *257*, 347-354.
12. D. Lamba, S. Glover, W. Mackie, A. Rashid, B. Sheldrick and S. Pérez, *Glycobiology*, **1994**, *4*, 151-163.
13. J.A. Kanters, B. van Dijk and J. Kroon, *Carbohydr. Res.*, **1991**, *212*, 1-11.
14. A.J. Polvorinos, R.R. Contreras, D. Martin-Ramos, J. Romero and M.A. Hidalgo, *Carbohydr. Res.*, **1994**, *257*, 1-10.
15. D. Lamba, W. Mackie, B. Sheldrick, P. Belton and S. Tanner, *Carbohydr. Res.*, **1988**, *180*, 183-193.
16. W. Mackie and E.A. Yates, D. Lamba, *Carbohydr. Res.*, **1995**, *226*, 65-74.
17. W. Huang, L. Boju, L. Tkalec, H. Su, H.O. Yang and N.S. Gunay, *Biochemistry*, **2001**, *40*, 2359-2372.
18. G. Michel, K. Pojasek, Y. Li, T. Sulea, R. Linhardt, R. Raman, V. Prabhakar, R. Sasisekharan and M. Cygler, *J. Biol. Chem.*, **2004**, *279*, 32882.
19. V.V. Lunin, Y. Li, R.J. Linhardt, H. Myazono, M. Kyogashima, T. Kaneko, A.W. Bell and M. Gygler, *J. Mol. Biol.*, **2004**, *337*, 367-386.
20. Z. Li, M. Kienetz, M.M. Cherney, M.N. James and D. Bromme, *J. Mol. Biol.*, **2008**, *383*, 78-91.
21. B.M. Sattelle and A. Almond, *J. Comput. Chem.*, **2010**, *31*, 2932-2947.
22. W.L. Jorgensen and J.J. Tirado-Rives, *J. Am. Chem. Soc.*, **1988**, 110:1657.
23. A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus and *J. Phys. Chem B.*, **1998**, *102*, 3586-3616.
24. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, **1995**, *117*, 5179-5197.
25. A.R. Leach, "Molecular Modelling, Principles and Applications", Prentice Hall, Second edition, 2001, p. 353-406, 479-483.
26. R. S. Mulliken, *J. Chem. Phys.*, **1955**, *23*, 1833-1840.
27. P. O. Lödwin, *Adv. Quantum Chem.*, **1970**, *5*, 185-199.
28. C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, **1990**, *11*, 361-373.
29. R. J. Woods, M. Khalil, W. Pell, S. H. Moffat and V. H. Jr. Smith, *J. Comput. Chem.*, **1990**, *11*, 297-310.
30. M. Basma, S. Sundara, D. Calgan, T. Vernali and R. J. Woods, *J. Comput. Chem.*, **2001**, *11*, 1125-1137.
31. G. Clipa, M. T. Hyvönen, A. Koivuniemi and M. L. Riekkola, *J. Comput. Chem.*, **2010**, *31*, 1670-1680.
32. M. Ragazzi and D. R. Ferro, *J. Mol. Struct. (Theochem)*, **1997**, 107-122.
33. D. E. Williams, *Biopolymers*, **1990**, *29*, 1367-1386.
34. C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, **1993**, *97*, 10269-10280.
35. T. H. Dunning Jr., *J. Chem. Phys.*, **1970**, *53*, 2823-2883.
36. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods*, J. Comput. Chem.*, **2008**, *29*, 622-655.
37. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery Jr, *J. Comput. Chem.*, **1993**, *14*, 1347-1363.
38. H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comp. Phys. Comm.*, **1995**, *91*, 43-56.
39. W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics.*, **1996**, *14*, 33-38.
40. F. N. Keutsch, R. S. Fellers, M. G. Brown, M. R. Viant, P. B. Petersen and R. J. Saykally, *J. Am. Chem. Soc.*, **2001**, *123*, 5938-5941.
41. S. Woutersen, Y. Mu, G. Stock and P. H. Woutersen, *Chem. Phys.*, **2001**, *266*, 137-147.
42. B. M. Sattelle, J. Shakeri, I. S. Roberts and A. Almond, *Carbohydr. Res.*, **2010**, *345*, 291-302.