# Free Energy Simulations of D-Xylose in Water and Methyl D-Xylopyranoside in Methanol

**Christer Höög and Göran Widmalm***

*Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University,S-106 91 Stockholm, Sweden*

*Received: November 9, 2000; In Final Form: March 6, 2001*

Free energy simulations have been performed to analyze the anomeric equilibrium of D-xylose in water and methyl D-xyloside in methanol. The anomeric $\alpha/\beta$ ratios were determined experimentally by $^1$H NMR spectroscopy. The molecular dynamics simulations employed the canonical ensemble to calculate the Helmholtz free energy change ($\Delta A$). The free energy perturbation and the thermodynamic integration methods were used for the evaluation. The duration of simulations was 500 ps with 18 and 24 $\lambda$-points in water and methanol, respectively. The relative free energy change $\Delta\Delta A$(solv−gas) calculated as $\beta-\alpha$ was for both solvents small, $\leq$0.3 kcal, but favored different anomeric configurations. Analysis of solvent structure indicated that the $\beta$-anomeric form possessed greater hydrogen bond ability at the anomeric center in xylose, whereas for the methyl xyloside there was no significant difference.

## Introduction

Free energy simulations have during the past decade found increasing use in biological applications because they aim at determining configurational changes or ligand binding processes in a quantitative manner.[1−7] Methodological work often calculates self-transformations to validate or evaluate different approaches to the calculation of free energy changes ($\Delta G$).[8,9] In the realm of biology, the area most commonly addressed by the free energy simulations is that of binding of different ligands to proteins, including the effect of mutation of the protein for the binding process.[10−15] A configurational change or a mutation of a molecule from state $\alpha$ to state $\beta$ can be represented by the thermodynamic cycle shown in Figure 1. In this way solvation effects can be investigated by simulation via calculation of the relative free energy difference ($\Delta\Delta G$). In studies of ligand binding, the calculation of $\Delta\Delta G$ is particularly useful because errors due to inadequate potential energy functions as well as those associated with finite-length simulations will approximately cancel for the two perturbations. Thus, the resulting relative free energy change should describe the difference in binding of the two ligands to the protein.

In molecular simulations one may use different ensembles. When the canonical (N,V,T) ensemble is applied, the free energy is represented by $A$, or Helmholtz free energy.[16] Subsequently, the free energy difference is denoted by $\Delta A$ (as in the present simulations). Assuming that the equipartition principle of energy applies, it should be possible to describe the Hamiltonian by a potential energy function,[17,18] such as CHARMM. In free energy simulations a coupling parameter $\lambda$ is used to bridge different states. For the initial state, $\alpha$, $\lambda = 0$ and for the final state, $\beta$, $\lambda = 1$. In CHARMM it is possible to use a "dual topology" approach in which two sets of atoms representing the initial and final states in a perturbation are included explicitly.[8,19] Any intermediate points are then effectively represented by the coupling parameter $\lambda$ defining the mixture of the two topologies.

There are several methods for calculation of free energy changes. Two in common use are: (i) the free energy perturbation (FEP) method[20] and (ii) the thermodynamic integration (TI) method.[21] The former is described by

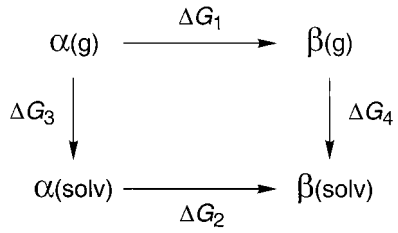



**Figure 1.** Thermodynamic cycle representing a molecule of states $\alpha$ and $\beta$ in the gas phase and in solution. The relative free energy change ($\Delta\Delta G$) is given by $\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$. The (nonphysical) changes $\Delta G_1$ and $\Delta G_2$ can be calculated in a straightforward manner.

$$\Delta G = -kT \ln\langle e^{-(H_\beta - H_\alpha)/kT}\rangle_\alpha \qquad (1)$$

where $H_\alpha$ and $H_\beta$ are the Hamiltonians of the states $\alpha$ and $\beta$, and the other symbols have their usual meaning. The free energy between two adjacent (intermediate) states should differ by no more than $\sim 2kT$, and, subsequently after a number of simulations, $\Delta G$ between the end states can be obtained. The latter method is described by

$$\Delta G = \int_{\lambda_\alpha}^{\lambda_\beta} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad (2)$$

Numerical integration performed over the discrete $\lambda$-points subsequently yields $\Delta G$.

For simple sugars, the anomeric equilibrium is especially interesting. The monosaccharide glucose shows a small preference for the $\alpha$-anomeric configuration as determined by ab initio calculations,[22] whereas in water solution mutarotation establishes the $\beta$-anomer in $\sim$2:1 excess. For the methyl glucosides, the anomeric effect is more dominant, favoring the $\alpha$-anomer in methanol as solvent. Free energy simulations have been performed by several groups on different monosaccharides in water, e.g., glucose,[23−25] xylose,[26] and mannose.[27] The small free energy differences for these systems make these simulations extremely challenging. In the present study we investigate by

* E-mail address: gw@organ.su.se.

10.1021/jp0041252 CCC: $20.00 
Published on Web 06/16/2001

**TABLE 1: Energy Difference ($\beta-\alpha$) from Experiment and Simulations in kcal/mol**

| | $\Delta G$(solv) | $\Delta E$(g) | | $\Delta A$(g) | | $\Delta A$(solv) | | $\Delta\Delta A$(solv-g) | |
|---|---|---|---|---|---|---|---|---|---|
| compound | EXP | EM | simulation | FEP | TI | FEP | TI | FEP | TI |
| xylose | −0.4 | −0.8 | W1 | −0.8 | −0.9 | −0.7 | −0.6 | 0.1 | 0.3 |
| | | | W2 | −0.8 | −0.8 | −0.5 | −0.6 | 0.3 | 0.2 |
| methyl xyloside | 0.3 | 1.9 | M1 | 2.2 | 1.9 | 1.9 | 1.8 | −0.3 | −0.1 |
| | | | M2 | 2.2 | 2.1 | 1.9 | 2.1 | −0.3 | 0 |

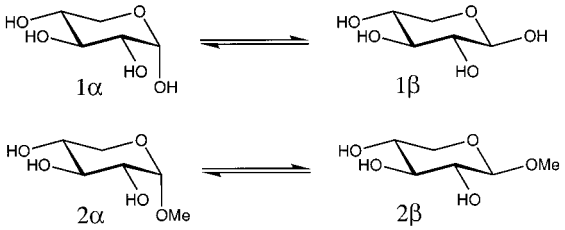



**Figure 2.** Schematic of the anomeric equilibrium of D-xylopyranose (**1**) and methyl D-xylopyranoside (**2**).

free energy simulations the anomeric equilibrium of D-xylose (**1**) in water and methyl D-xyloside (**2**) in methanol (Figure 2) and the effect of solvent thereupon.

## Materials and Methods

**General.** Methyl D-xylopyranosides were produced by reflux of D-xylose in acidified methanol overnight followed by equilibration for several weeks at room temperature. D-Xylose was equilibrated in deuterated water for several weeks at ambient temperature. $^1$H NMR measurements were performed at 25 °C using a Varian 400 NMR spectrometer.

**Simulation.** The Molecular Mechanics program CHARMM[19,28] with the force field PARM22 (Molecular Simulation Inc., San Diego), which is similar to the carbohydrate force field developed by Ha et al.,[29] was used for the simulations. Initial conditions for the free energy simulations (TSM module of CHARMM) of **1** were prepared by placing D-xylopyranose in a cubic water box with a side length of 25.82 Å containing 576 TIP3P water[30] molecules and removing those that were closer than 2.5 Å to any solute atom. This procedure resulted in a system with the monosaccharide and 565 waters, which was initially energy minimized using steepest descent, 300 steps, followed by adopted basis Newton−Raphson until the root-mean-square gradient was less than 0.01 kcal mol$^{-1}$ Å$^{-1}$. The system was for each $\lambda$ equilibrated for 50 ps continued by production runs for 500 ps at each window. The method proposed by Nóse[31] was used for keeping the temperature constant at 300 K, resulting in a canonical ensemble, i.e., constant number of particles, constant volume, and constant temperature (N,V,T ensemble). Thus, the free energies from simulations correspond to Helmholtz free energy. Minimum image boundary conditions were used with a heuristic update frequency of the nonbond list and a force shift cutoff[32] acting to 12 Å, using a dielectric constant of unity. SHAKE, with a tolerance gradient of $10^{-4}$, was used to restrain hydrogen-heavy atom bond stretch; the time step was accordingly set to 2.0 fs. The coupling parameter $\lambda$ used the following values: 0.004, 0.01, 0.03, 0.06, 0.12, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.88, 0.94, 0.97, 0.99, and 0.996.

The free energy simulations of **2** in methanol were performed by the same procedure with some changes. The side length of the cubic box was 31.12 Å and contained 448 methanol residues. The solvent molecules closer than 2.8 Å to any solute atom were removed. This resulted in a system with methyl D-xylopyranoside and 443 methanol residues which was energy minimized using steepest descent, 400 steps, followed by adopted basis Newton−Raphson until the root-mean-square gradient was less than 0.01 kcal mol$^{-1}$ Å$^{-1}$. The $\lambda$ used were 0.005, 0.007, 0.012, 0.02, 0.03, 0.05, 0.08, 0.12, 0.18, 0.26, 0.4, 0.6, 0.74, 0.82, 0.88, 0.92, 0.95 0.97, 0.98, 0.988, 0.993, and 0.995.

Additional simulations were performed for 2 ns each at the endpoints, i.e., $\lambda = 0$ and $\lambda = 1$, using Berendsen's weak coupling algorithm[33] for a constant temperature, and data were saved every 0.02 ps. Free energy simulations of **1** and **2** were also performed under comparable conditions in a vacuum using the same $\lambda$ as above to be able to calculate the influence of solvent. All simulations were performed twice to obtain independent data sets.

Analyses of the free energy simulations were performed with both the free energy perturbation (FEP) method, employing double-wide sampling,[34,35] and the thermodynamic integration (TI) method as they are implemented[35,36] in CHARMM. Radial distribution functions were integrated for oxygen−oxygen out to 3.5 Å and for hydrogen−oxygen to 2.5 Å to give the corresponding coordination numbers. For the generation of spatial distribution functions (SDF), the molecular reference frame was attached to the C1, C3, and C5 atoms. The SDF figures were produced with the general purpose simulation software M.Dyna*Mix*[37] and the "gOpenMol" package.[38] Computations were performed at an IBM SP2 at the Center for Parallel Computers, KTH, Stockholm. An MD production run of 500 ps required ∼115 h of CPU time on a single processor.

## Results and Discussion

The anomeric equilibrium of D-xylose in water and methyl D-xyloside in methanol was investigated by $^1$H NMR spectroscopy at 25 °C. It was found by integration of the anomeric resonances that the $\alpha$:$\beta$ ratio of **1** was 1.0:2.0 and of **2** it was 1.7:1.0. The results are similar to those reported previously under comparable conditions.[39,40] The experimentally determined $\Delta G$ are given in Table 1. As described above, the present simulations are performed at constant volume. In the case of a constant pressure simulation the following relationship exists: $\Delta G = \Delta A + p\Delta V$. For a dilute solution of the monosaccharides discussed herein, the change between the two states $\alpha$ and $\beta$ should have a negligible effect on the volume. Therefore, below we will compare directly $\Delta A$ from simulation with $\Delta G$ from experiment.

The preference of an axial electronegative group at the anomeric position of a pyranosidic sugar is known as the anomeric effect.[41] For simple unsubstituted sugars it has been argued that the effect is small[22] and that the energy difference between $\alpha$ and $\beta$ states decreases as the basis-set quality of the ab initio calculations increases.[42] A most recent ab initio calculation for glucose shows a preference for the $\alpha$-anomer of only 0.4 kcal/mol (Boltzmann average).[21] Thus, it is reasonable to assume that the anomeric effect in D-xylose also is small and favors the $\alpha$-state. However, for methyl and alkyl glycosides the anomeric effect should be larger, and a clear preference for the axial configuration is supported by ab initio data.[43,44] Energy

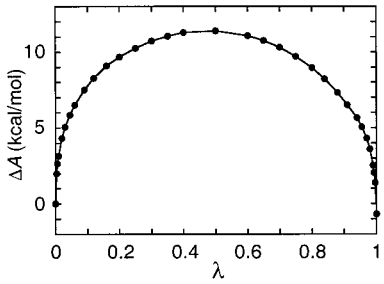



**Figure 3.** Free energy change of **1** in water for the transformation between the α-anomeric configuration and the *β*-anomer as a function of the coupling parameter *λ*. Simulation W1 is denoted by (●) and W2 by a solid line.

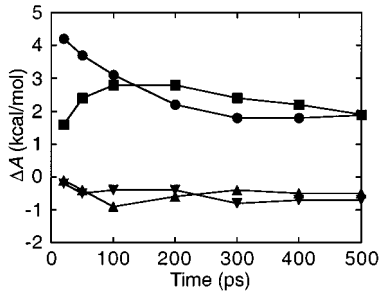



**Figure 4.** Free energy change as a function of simulation time. Δ*A* was calculated using the FEP method; M1 (●), M2 (■), W1 (▼), and W2 (▲).

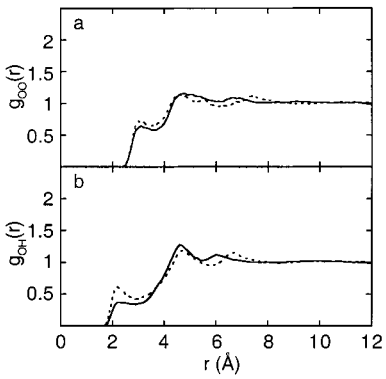
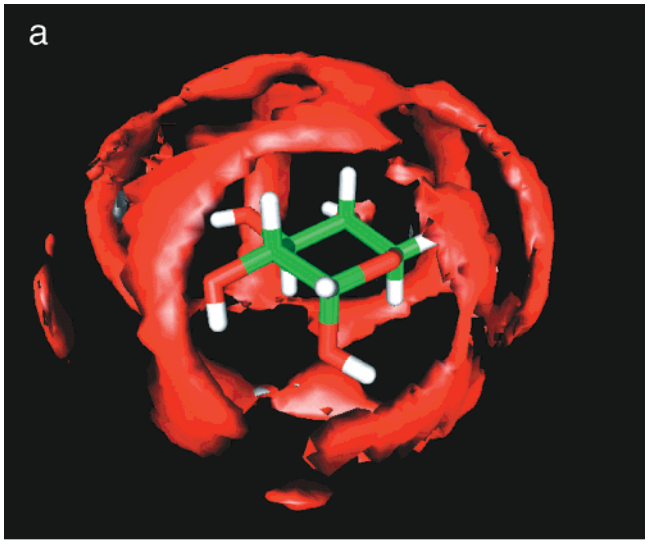



**Figure 5.** Radial distribution functions of the anomeric oxygen in **2** showing differences between the α-anomer (solid line) and the *β*-anomer (dashed line): (a) anomeric oxygen to oxygen in methanol and (b) anomeric oxygen to hydroxyl hydrogen in methanol.

minimization in vacuo with the present CHARMM force field (PARM22) favors for xylose the *β*-anomer, in contrast to the above assumption, whereas for methyl xyloside the α-anomer is preferred, in agreement with conclusions from ab initio data (Table 1).

The free energy change of **1** in water as a function of the coupling parameter *λ* is shown in Figure 3. Excellent agreement between the two simulations is evident and reveals good precision in the simulations. In calculating Δ*A*, the extent to which sufficient sampling has been obtained is crucial to the quality of the data. Figure 4 shows for the FEP method that simulations are needed for several hundreds of ps to obtain a sufficiently good estimate of Δ*A* and stability (convergence) in the transformations, especially in methanol. The free energy differences are compiled in Table 1. The two methods are known to lead to results of comparable accuracy and are considered robust.[45] Comparison of Δ*A*(g) and Δ*A*(solv) show small differences that are revealed directly by the relative free energy differences ΔΔ*A*(solv-g) being no larger than 0.3 kcal/mol. Thus,



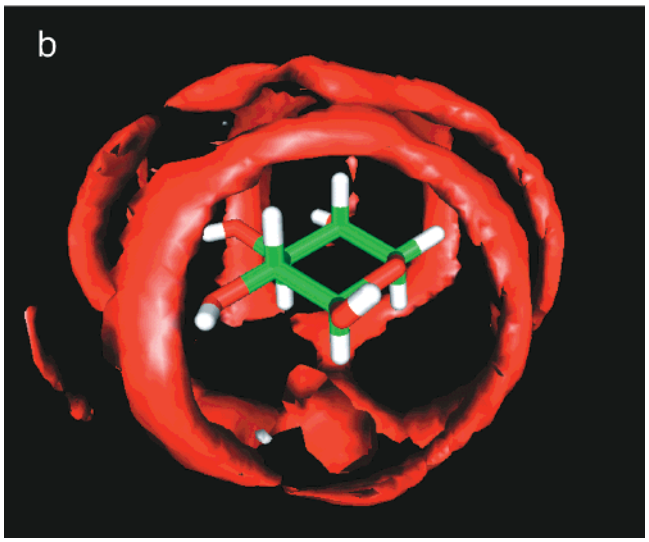




**Figure 6.** Spatial distribution functions around (a) **1**α and (b) **1***β*. Contours show the regions of high positive deviations of the local density (3.0 times higher than the average value). Atomic coloring is green for carbon, red for oxygen, and white for hydrogen.

**TABLE 2: Coordination Numbers of the Anomeric Oxygen Atom**

| compound | $n_{OO}$ | $n_H$ |
|---|---|---|
| **1**α | 5.0 | 2.9 |
| **1***β* | 5.9 | 3.6 |
| **2**α | 0.9 | 0.2 |
| **2***β* | 1.0 | 0.3 |

the solvent effect on the conformational equilibria is, from these simulations, very small. Using a recent development in free energy simulations that comprises "multiple-copy locally enhanced sampling," Simmerling et al.[24] were able to reproduce the solvation effect driving the equilibrium to the *β*-form of glucose (which in vacuo preferred the α form by ∼0.5−1.0 kcal/mol). The strength of this method is due to the fact that barriers to conformational transitions are reduced, thereby significantly improving the sampling and convergence properties of the free energy. It is of interest to note that the study by Simmerling et al.[24] on glucose, those of Brady and co-workers on glucose[22] and xylose,[25] as well as the present simulations used the same water model, viz., TIP3P. Thus, the differences in solvent effects between the simulations should not be sought in the choice of the solvent model but rather in the degree of conformational

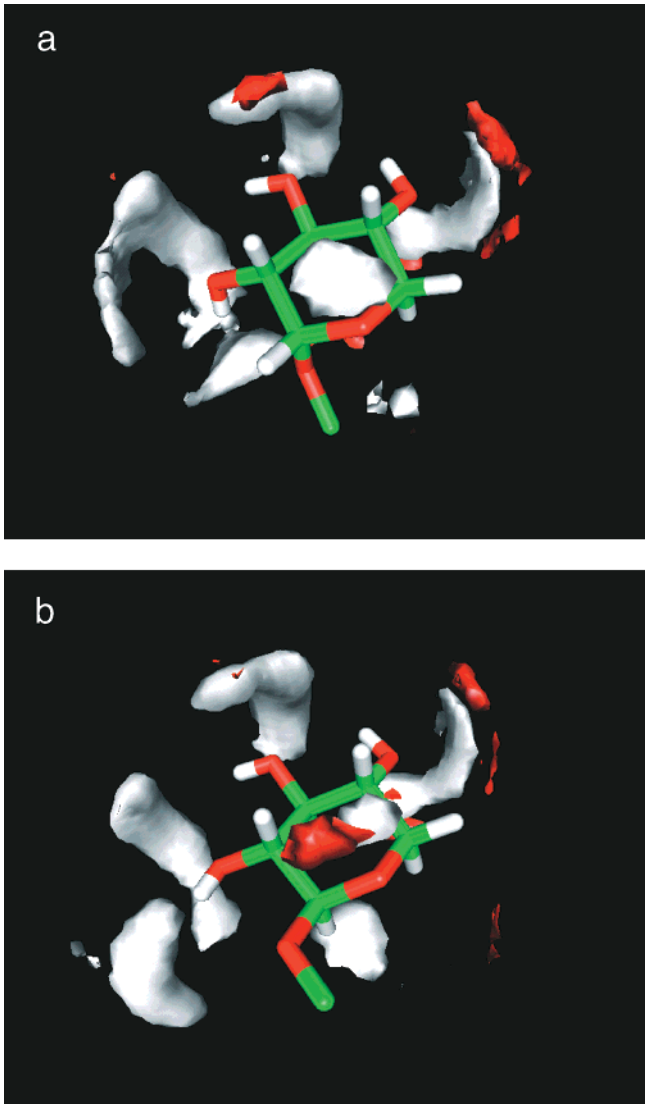



**Figure 7.** Spatial distribution functions around (a) **2**α and (b) **2**$\beta$. Contours show the regions with a local density 3.4 times higher than the average value. Atomic coloring as in Figure 6.

sampling obtained in the simulations. Also, by necessity, small differences do exist between the different systems, because the solute molecules are described by somewhat different potentials, and thereby the effective interaction potential between solvent and the solute is changed.

The solvent structure around **1** has previously been investigated by MD simulations.[26] In the $\beta$-anomer O1 had a larger amount of hydrogen bonds than the α-anomer. This is consistent with our simulations where the coordination numbers are larger for **1**$\beta$ than for **1**α (Table 2). The radial distribution functions (RDFs) for the anomeric oxygen in **2** to the methanolic oxygen and hydroxyl proton are shown in Figure 5. The $g_{OO}(r)$ are very similar for the two anomeric forms, whereas a small difference is present for $g_{OH}(r)$ as shown for the hydrogen coordination numbers (Table 2).

In comparison to RDFs, additional information about the three-dimensional solvation structure can be obtained using spatial distribution functions (SDFs) of atomic density around a solute[46] given by

$$g_{AB}(\vec{r}) = \frac{\rho_B(\vec{r}|\vec{r}_A = 0)}{\rho_B} \qquad (3)$$

where the first index A refers to a solute atom, and the second index B refers to a solvent atom, $r$ is a vector in Cartesian coordinate space between the atoms, and $\rho_B$ is the bulk solvent density of B in the solvent. The SDFs give radial as well as angular dependence of the solvation structure around the solute fixed in the local coordinate frame. Typical band-shaped structures of water molecules, especially oxygen, surrounding the solute are shown in Figure 6 for **1**α and **1**$\beta$. However, in methanol regions of high local density of the hydroxyl protons of methanol surround the regions of the sugar hydroxyl groups where hydrogen bond donation can occur from solvent hydrogens to solute oxygens (Figure 7). An interesting difference between **2**α and **2**$\beta$ is the solvent structure around O2 as a result of the configurational change at the anomeric position.

In conclusion, free energy simulations of the monosaccharides xylose in water and methyl xyloside in methanol show that the influence of solvent on the α:$\beta$ ratio is small. The energy difference in vacuo between the anomeric forms for free sugars such as glucose and xylose is anticipated to be very small. However, the modest effect of solvent is nevertheless extremely important. The present simulations, although sampled for 500 ps at each $\lambda$-point, are still lacking sufficient sampling of conformational space for an accurate determination of $\Delta A$. The solvent structure around xylose is in agreement with previous MD simulations and that of methyl xyloside shows, in particular, local hydrogen bond donations from the hydroxyl group in methanol to the oxygen atoms in the sugar.

**Acknowledgment.** This work was supported by a grant from the Swedish Natural Science Research Council. We thank the Center for Parallel Computers, KTH, Stockholm, for putting computer facilities at our disposal.

## References and Notes

(1) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564−568.
(2) Fleischman, S. H.; Brooks, C. L., III *J. Chem. Phys.* **1987**, *87*, 3029−3037.
(3) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.
(4) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461−469.
(5) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graphics Mod.* **1998**, *16*, 213−225.
(6) Guo, Z.; Brooks, C. L., III; Kong, X. *J. Phys. Chem. B* **1998**, *102*, 2032−2036.
(7) Schäfer, H.; van Gunsteren, W. F.; Mark, A. E. *J. Comput. Chem.* **1999**, *20*, 1604−1617.
(8) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487−1493.
(9) Sen, S.; Nilsson, L. *J. Comput. Chem.* **1999**, *20*, 877−885.
(10) Zacharias, M.; Straatsma, T. P.; McCammon, J. A.; Quiocho, F. A. *Biochemistry* **1993**, *32*, 7428−7434.
(11) Hu, H.; Shi, Y. Y.; Wang, C. X. *Proteins* **1996**, *26*, 157−166.
(12) Liang, G.; Schmidt, R. K.; Yu, H.-A.; Cumming, D. A. *J. Phys. Chem.* **1996**, *100*, 2528−2534.
(13) Pathiaseril, A.; Woods, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 331−338.
(14) Sen, S.; Nilsson, L. *Biophys. J.* **1999**, *77*, 1801−1810.
(15) Banba, S.; Guo, Z.; Brooks, C. L., III *J. Phys. Chem. B* **2000**, *104*, 6903−6910.
(16) Lybrand, T. P. *Rev. Comput. Chem.* **1990**, 1, 295−320.
(17) Jorgensen, W.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(18) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(19) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *Encyclopedia of Computational Chemistry*, van Gunsteren, W. F., Schleyer, P. v. R., Eds.; John Wiley & Sons: Chichester, 1998; Vol. 1, pp 271−277.
(20) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(21) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300−313.
(22) Barrows, S. E.; Storer, J. W.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *J. Comput. Chem.* **1998**, *19*, 1111−1129.
(23) Ha, S.; Gao, J.; Tidor, B.; Brady, J. W.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 1553−1557.

(24) van Eijck, B. P.; Hooft, R. W. W.; Kroon, J. *J. Phys. Chem.* **1993**, *97*, 12093−12099.
(25) Simmerling, C. S.; Fox, T.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 5771−5782.
(26) Schmidt, R. K.; Karplus, M.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 541−546.
(27) Senderowitz, H.; Still, W. C. *J. Phys. Chem. B* **1997**, *101*, 1409−1412.
(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(29) Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydr. Res.* **1988**, *180*, 207−221.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J. *J. Chem. Phys.* **1983**, *79*, 926−935.
(31) Nóse, S. *J. Chem. Phys.* **1984**, *81*, 511−519.
(32) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.
(33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(34) Bennet, C. H. *J. Comput. Phys.* **1976**, *22*, 245−268.
(35) Beveridge, D. L.; DiCapua, F. M. In *Computer Simulations of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989; pp 1−26.
(36) Brooks, C. L., III In *Computer Simulations of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989; pp 73−88.
(37) Lyubartsev, A.; Laaksonen, A. *Comput. Phys. Commun.* **2000**, *128*, 565−589.
(38) Laaksonen, L., "gOpenMol", version 1.31, Center of Scientific Computations, Espoo, Finland. http://laaksonen.csc.fi/gopenmol/gopenmol.html.
(39) Angyal, S. J. *Adv. Carbohydr. Chem. Biochem.* **1984**, *42*, 15−68.
(40) Bishop, C. T.; Cooper, F. P. *Can. J. Chem.* **1963**, *41*, 2743−2758.
(41) *The anomeric effect and associated stereoelectronic effects*, Thatcher, G. R. J., Ed.; American Chemical Society: Washington, DC, 1993.
(42) Molteni, C.; Parrinello, M. *J. Am. Chem. Soc.* **1998**, *120*, 2168−2171.
(43) Tvaroska, I.; Carver, J. P. *J. Phys. Chem.* **1994**, *98*, 9477−9485.
(44) Odelius, M.; Laaksonen, A.; Widmalm, G. *J. Phys. Chem.* **1995**, *99*, 12686−12692.
(45) Chipot, C.; Kollman, P. A.; Pearlman, D. A. *J. Comput. Chem.* **1996**, *17*, 1112−1131.
(46) Kusalik, P. G.; Laaksonen, A.; Svishchev, I. M. In *Molecular dynamics: From classical to quantum methods*; Balbuena, P. B., Seminario, J. M., Eds.; Elsevier Science: Amsterdam, 1999; *Theoretical and Computational Chemistry*, Vol. 7, pp 61−97.