Use of Locally Enhanced Sampling in Free Energy Calculations: Testing and Application to the α → β Anomerization of Glucose

Carlos Simmerling, Thomas Fox,† and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94148-0446

Received July 21, 1997. Revised Manuscript Received November 5, 1997

Abstract: We present a new approach to calculating conformational equilibria in complex molecules, multiple-copy locally enhanced sampling (LES), and apply this approach to the well-known α → β anomeric equilibrium in glucose. Although a variant of this method has been previously applied to protein stability mutations, it is particularly suited to analyze conformational equilibria in molecules such as glucose, with its many OH groups. This methodology allows us to more rapidly calculate complex equilibria both in vacuo and in solution. We have employed a "generic" force field using 1,1-dimethoxymethane and 1,1-dihydroxymethane as models to derive the torsional parameters associated with O−C−O−C and O−C−O−H fragments. Within this model, we can definitively establish the magnitude of intramolecular and solvation contributions to the α → β equilibrium. Specifically, we find that in vacuo glucose prefers to be in the α configuration by ∼0.5−1.0 kcal/mol (predominantly because of the gauche tendency of the O−C−O−C linkage), and it is the solvation free energy which drives the equilibrium to the β form (Δ*G* in solution = −0.2 kcal/mol calculated, −0.3 kcal/mol experimental). Using the LES approach, which reduces barriers to conformational transitions, we obtain free energies converged (compared to longer calculations) within 0.2 kcal/mol in 200 ps, both in vacuo and in water. Convergence for the single-copy method is considerably slower; variations in the calculated free energy of over 1 kcal/mol in vacuo and 0.5 kcal/mol in water are still observed in simulations of several nanoseconds in length. Thus, it appears that LES can be at least an order of magnitude faster to converge than single-copy methods.

Introduction

Carbohydrate molecules play many important physiological roles in such varied contexts as energy storage, structural units of cell walls and membranes, and cellular components responsible for function and growth. As a result, this class of molecules has received considerable attention in the past, including many experimental and theoretical studies of structure and conformational preferences. The relative populations of the various conformers of many carbohydrates in aqueous solution have been determined experimentally.

The anomeric effect is a key component of the conformational preferences of carbohydrate molecules. In glucose, this stereoelectronic effect makes the OH group much more favorable in the axial (α) configuration than expected based on simple steric effects. The actual ratio of anomers found, however, can be solvent-dependent.[1] For glucose in aqueous solution, 64% is found to be the β-glucopyranose form,[2] in which the anomeric hydroxyl group is in the equatorial position, compared to 36% in the α-glucopyranose form, which has the hydroxyl in the axial position. This ratio corresponds to a free energy difference (α → β) in water of −0.3 kcal/mol. It is difficult to determine experimentally which interactions preferentially stabilize the different configurations and what role the solvent plays in determining the ratio. These issues can be addressed, however, using theoretical methods. Many studies have been reported in the literature, including several quantum mechanical (QM) calculations on glucose.[3−7] While such methods can be highly accurate, they can be applied to only a limited number of conformations. For example, Barrows et al.[6] applied high-quality ab initio calculations with large basis sets and the coupled cluster method for electron correlation to 11 conformers of glucose and found an anomeric free energy difference of 0.4 kcal/mol favoring α in the gas phase. Applying the SM5.4/A[8] continuum solvent model to all 11 conformations resulted in aqueous solvent favoring the β form by 0.6 kcal/mol, with a net anomeric free energy difference in water (α → β) of −0.2 kcal/mol, in excellent agreement with the experimental value.

If one wants to model large, complex systems in the condensed phase with explicit inclusion of solvent, molecular mechanical (MM) methods are an efficient and appropriate choice. Several molecular mechanical force field parameter sets specific to carbohydrates have been developed.[9−14] Such force fields can subsequently be combined with methods such as free

† Current address: Dr. Karl Thomae GmbH, ACF/Computer Aided Drug Design, 88397 Biberach/Riss, Germany.

† To whom correspondence should be addressed. Telephone: (415) 476-4637. Fax: (415) 476-0688. E-mail: pak@cgl.ucsf.edu.

(1) Jaseja, M.; Perlin, A. S.; Dais, P. *Magn. Reson. Chem.* **1990**, *28*, 283−289.

(2) Angyal, S. J. *Aust. J. Chem.* **1968**, *21*, 2737.

(3) Polavarapu, P. L.; Ewig, C. S. *J. Comput. Chem.* **1992**, *13*, 1255−1261.

(4) Brown, J. W.; Wladkowski, B. D. *J. Am. Chem. Soc.* **1996**, *118*, 1190−1193.

(5) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1993**, *115*, 5745−5753.

(6) Barrows, S. E.; Storer, J. W.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *J. Comput. Chem.* **1997**, submitted.

(7) Salzner, U.; Schleyer, P. V. *J. Org. Chem.* **1994**, *59*, 2138−2155.

(8) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385−16398.

(9) Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydr. Res.* **1988**, *180*, 207−221.

S0002-7863(97)02457-8 CCC: $15.00
Published on Web 06/02/1998

energy perturbation (FEP) or thermodynamic integration (TI), which have become powerful tools for the calculation of free energy differences for both conformational and chemical changes.

For example, Ha et al. used TI and found the α form of glucose to be more stable in water than the β anomer (in contrast to experiment) by 0.31 ± 0.43 kcal/mol and reported a large (3.6 kcal/mol) contribution of aqueous solvation preferentially stabilizing the β form.[15] Other researchers used the GROMOS force field[16] and calculated the $\alpha \rightarrow \beta$ anomeric free energy to be −0.86 ± 0.15 kcal/mol, with solvation in this case preferring the α anomer by 0.5 kcal/mol.[17] These two examples demonstrate that results obtained from such methods can be quite sensitive to the particular details of the simulation and force field, and a definitive understanding of the various contributions has not been achieved. These calculations are particularly challenging, however, due to the small free energy difference and the large number of potentially accessible conformations of the CH_2OH and four OH groups groups of the molecule that may need to be explored during the calculation. For this reason several studies have used similar molecules with fewer rotatable bonds (such as 2-methoxytetrahydropyran[18−20]) as models of the anomeric effect in carbohydrates. Schmidt et al.[21] carried out free energy calculations using xylose, which is similar to glucose but lacks the exocyclic hydroxymethyl group, and has a similar experimental α/β ratio in aqueous solution. They reported that the α anomer was again slightly favored (0.15 kcal/mol) with a relatively small solvent contribution of −0.6 kcal/mol favoring β. The solvation effect was suggested to be due to increased hydrogen bonding for the β anomer.

An alternate approach to including many conformations in free energy calculations has recently been reported.[22] The "jumping between wells" (JBW) technique was applied to several carbohydrates[10,23] with the GB/SA continuum solvent model,[24] yielding an $\alpha \rightarrow \beta$ free energy difference for glucose of −0.22 ± 0.02 kcal/mol. A modified AMBER force field was derived, including torsional parameters fit to energy profiles obtained from quantum mechanical calculations on 2-hydroxytetrahydropyran and 2-methoxytetrahydropyran. While this method is very promising and showed excellent convergence properties, the current ability to include only continuum solvent models and exclusive applicability to conformational equilibria somewhat limit its potential.

Elber and co-workers have shown the utility of multiple-conformation/locally enhanced sampling (LES) in studying equilibria in biological molecules.[25−29] They have also applied this methodology to calculations of relative free energy in native and mutant proteins.[30] Simmerling and Elber have shown that LES methods can be aproximately an order of magnitude more efficient than ordinary molecular dynamics.[28] Thus, we felt that LES was an excellent approach to the calculation of conformational free energies for glucose.

One of the key features that developers of molecular mechanical force fields strive for is transferability. It was encouraging to us that the Cornell et al. force field,[31] developed for proteins and nucleic acids, could be extended to the conformational equilibrium of 1,3-dioxanes with a single addition,[32] using ab initio calculations on 1,1-dimethoxymethane[33] (DMM), since there is no OCOC unit in proteins or nucleic acids and it is this fragment that produces the anomeric effect. The torsional potential resulting from reproducing the conformational energies of DMM was used without change in a series of 1,3-dioxanes and led to excellent results. The reason for the better performance of the Cornell et al. force field than MM2 and MM3 was suggested to come from the greater accuracy of the RESP partial atomic charges as compared to the point dipole model in MM2 and MM3. A logical next step is to see whether the new force field could be used to accurately calculate conformational energies in glucose, using the OCOC torsion parameters originally derived for DMM, along with additional parameters for OCOH developed in a similar fashion from 1,1-dihydroxymethane (DHM). We use those parameters without modification in glucose for the calculation of the anomeric free energy difference as well as energies of individual conformations and rotamer populations.

Thus, in our study, presented below, we are addressing two very important issues: (1) conformational sampling and (2) transferability/accuracy of force field parameters. An adequate representation of both sampling and energy function is essential for any quantitative modeling of complex molecules. We demonstrate that LES does offer an improvement in efficiency, particularly in vacuo, where it is easier to become locked into a single hydrogen bonding configuration, and that our force field approach does meet the test of transferability quite well.

Methods

Figure 1 depicts a model of β-glucopyranose, with the atom numbering that will be used in the text. We label the rotatable bonds $\tau_1-\tau_6$, with $\tau_{1\alpha}$ representing τ_1 in the α anomer (which has the hydroxyl in the axial position) and $\tau_{1\beta}$ for the β anomer (with the hydroxyl group in the equatorial position). The model is shown in the $^4C_1$ (chair) conformation with the ring hydroxyl groups hydrogen bonded in a counterclockwise pattern. The exocyclic hydroxymethyl group con-

(10) Senderowitz, H.; Parish, C.; Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 8985−8985.
(11) Reiling, S.; Schlenkrich, M.; Brickmann, J. *J. Comput. Chem.* **1996**, *17*, 450−468.
(12) Glennon, T. M.; Zheng, Y. J.; Legrand, S. M.; Shutzberg, B. A.; Merz, K. M. *J. Comput. Chem.* **1994**, *15*, 1019−1040.
(13) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraserreid, B. *J. Phys. Chem.* **1995**, *99*, 3832−3846.
(14) Kouwijzer, M.; van Eijck, B. P.; Kroon, J. *Acta Crystallogr. B* **1995**, *51*, 209−220.
(15) Ha, S. H.; Gao, J. L.; Tidor, B.; Brady, J. W.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 1553−1557.
(16) van Gunsteren, W. F. University of Groningen, 1987.
(17) van Eijck, B. P.; Hooft, R. W. W.; Kroon, J. *J. Phys. Chem.* **1993**, *97*, 12093−12099.
(18) Zheng, Y. J.; Legrand, S. M.; Merz, K. M. *J. Comput. Chem.* **1992**, *13*, 772−791.
(19) Wiberg, K. B.; Marquez, M. *J. Am. Chem. Soc.* **1994**, *116*, 2197−2198.
(20) Jorgensen, W. L.; Detirado, P. I. M.; Severance, D. L. *J. Am. Chem. Soc.* **1994**, *116*, 2199−2200.
(21) Schmidt, R. K.; Karplus, M.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 541−546.
(22) Senderowitz, H.; Guarnieri, F.; Still, W. C. *J. Am. Chem. Soc.* **1995**, *117*, 8211−8219.
(23) Senderowitz, H.; Still, W. C. *J. Phys. Chem. B* **1997**, *101*, 1409−1412.
(24) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(25) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(26) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277−9287.
(27) Li, H. Y.; Elber, R.; Straub, J. E. *J. Biol. Chem.* **1993**, *268*, 17908−17916.
(28) Simmerling, C.; Elber, R. *J. Am. Chem. Soc.* **1994**, *116*, 2534−2547.
(29) Simmerling, C. L.; Elber, R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3190−3193.
(30) Verkhivker, G.; Elber, R.; Nowak, W. *J. Chem. Phys.* **1992**, *97*, 7838−7841.
(31) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(32) Howard, A. E.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 243−261.
(33) Wiberg, K. B.; Murcko, M. A. *J. Am. Chem. Soc.* **1989**, *111*, 4821.

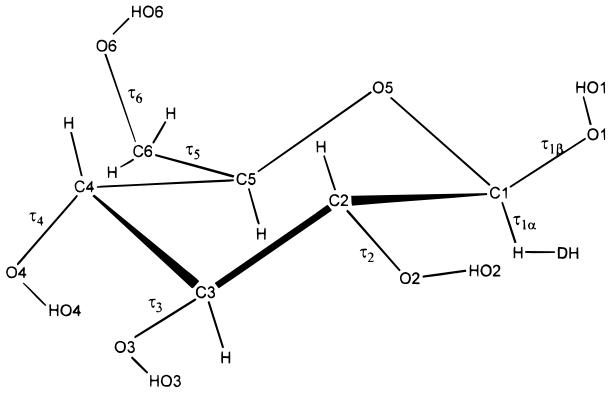

Figure 1. Model of the β anomer of glucose used in the perturbation. In the α anomer, the O1 hydroxyl group is in the axial, rather than equatorial, position. The dummy atom, corresponding to HO1 in the α form, is designated DH. Rotatable bonds are labeled $\tau_1-\tau_6$.

Table 1. Lennard-Jones AMBER Atom Types and RESP Partial Atomic Charges for Glucose[a]

atom	AMBER atom type	charge, A	charge, B
C1	CT	0.2677	0.3364
H1	H2	0.0770	0.0530
O1	OH	−0.5380	−0.6013
HO1	HO	0.3836	0.4293
C2	CT	0.0258	0.1270
H2	H1	0.1348	0.0972
O2	OH	−0.5477	−0.6760
HO2	HO	0.3671	0.4588
C3	CT	0.0878	0.1759
H3	H1	0.1323	0.0894
O3	OH	−0.5441	−0.6817
HO3	HO	0.3684	0.4620
C4	CT	0.0314	0.0093
H4	H1	0.1294	0.0987
O4	OH	−0.5426	−0.6238
HO4	HO	0.3689	0.4085
C5	CT	0.0079	0.0467
H5	H1	0.0929	0.1118
C6	CT	0.1683	0.1525
1H6	H1	0.0510	0.0291
2H6	H1	0.0372	0.1001
O6	OH	−0.5639	−0.6678
HO6	HO	0.3690	0.4481
O5	OS	−0.3639	−0.3831

[a] Model A is the set obtained from 12 conformations, and model B is the set obtained from only two conformers.

formation is defined by convention using the O5−C5−C6−O6 and C4−C5−C6−O6 dihedrals. Three low-energy rotamers for this group exist: GT, in which O6 is gauche with respect to O5 and trans to C4; TG, which has O6 trans to O5 and gauche to C4; and GG, which is gauche in both cases.

Force Field Parameters. The parameters for glucose were chosen as follows. Lennard-Jones atom types were assigned using the standard AMBER force field[31] and are listed in Table 1. Partial charges were independent of conformation and configuration with a single set employed for both α and β anomers. Previous work[32] has shown that a single set of charges provided optimal results. To find the appropriate conformations to use in deriving the partial charges, conformational search was carried out in vacuo (with a distance-dependent dielectric constant of $4R_{ij}$) using a torsional grid spacing of 60° (resulting in 719 unique minimized structures for α and 721 for β) with an initial partial charge of −0.4 assigned to each hydroxyl oxygen atom and 0.4 to the corresponding hydrogen atoms, with all other atoms without charge. The global energy minima for α and β were selected and optimized using Gaussian94 at the Hartree−Fock level with the STO-3G basis set. Five additional minimized conformations each for α and β were chosen randomly with the restraint that each rotatable bond was represented by more than one rotamer, as suggested from the studies by Cornell et al.[34] An RHF/6-31G* single-point calculation was carried out for each optimized structure, and the resulting electrostatic potentials were used for two-stage multiple-conformation RESP charge fitting[34−36] (model A, Table 1). In this method, a single set of charges is simultaneously fit to multiple conformations, as compared to averaging multiple charge sets obtained from single conformations. To test the sensitivity of the results to the charge model, an additional set of partial charges was derived. In this case, only the lowest energy conformations of the α and β anomers were used (model B, Table 1). These two charge sets were used for all further calculations, both in vacuo and in solution.

Table 2. Torsional Parameters Derived from Dimethoxymethane[32] and Dihydroxymethane

fragment	periodicity	phase (deg)	force constant (V_n/2, kcal/mol)
O−C−O−C	1	180	1.60
O−C−O−H	2	0	0.85
O−C−O−H	3	0	0.90

The choice of the number of conformations used in model A is somewhat arbitrary, but it was felt that this number should lead to a reasonable sampling of conformational and configurational heterogeneity in the charge set. As one can see, and as noted by Cornell et al.,[34] this conformational heterogeneity also leads to reduced charges on both oxygen and hydroxyl hydrogen atoms.

In addition to the charges, additional internal parameters were required for fragments that were not presented by Cornell et al. and for which specific electronic effects are likely to be important. For glucose, this is the case for torsional parameters for OCOC and OCOH units. Since the former were already derived by Howard et al.,[32] only the OCOH parameters need to be derived. We thus carried out ab initio calculations and optimized the energy of the g,g, g,t, and t,t geometries of 1,1-dihydroxymethane at the 6-31G* level and then calculated single-point energies of these and a number of other representative structures at MP2/6-31G*. We then used the RESP charges derived for the lowest energy g,g conformation and evaluated how well a molecular mechanical model with no torsional potential reproduced these relative energies. We then considered V_1, V_2, and V_3 torsional potentials for the O−C−O−H unit and found that $V_2 = 0.85$ and $V_3 = 0.9$ kcal/mol did a good job of reproducing the quantum mechanical energies, as well as fitting the MP2/6-31G* ab initio energies of a series of other conformations of this molecule, better than alternative sets of torsional parameters. For example, the ab initio (molecular mechanical) energies, relative to the g,g energy of 0.0, were 4.7(4.8) for g,t(60,180) and 9.8(11.2) for t,t. We had to use a larger V_3 torsional term than found in the Cornell et al. X−CT−OH−HO (X = any atom) term to ensure that the g,t conformation was a local minimum, as it is at the ab initio level. All parameters for this and the O−C−O−C fragment are presented in Table 2.

Free Energy Calculations. Since free energy is a state function, and therefore independent of the perturbation pathway, we are free to choose whatever path is convenient. Several different pathways have been used in the past to study this equilibrium. Among these has been the generation of an equilibrium distribution of the two anomers and using the populations to calculate the free energy difference.[10] Others have calculated the potential of mean force while slowly changing the chirality around the anomeric center.[17] Another approach is to simply mutate the anomeric hydroxyl group into a hydrogen and convert the other hydrogen into a hydroxyl.[15] We chose the latter aproach, since the same methodology can later be applied to chemical changes, in addition to the configurational change presently considered.

The mutation was carried out using the single topology method, in which all of the atoms exist in both states. In Figure 1 we show the topology of glucose used for the calculations. A dummy atom DH is

(34) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.
(35) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(36) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357−1377.

attached to the anomeric H1 hydrogen. During the perturbation of β to α, the H1 hydrogen changes into oxygen O1 and DH is converted into hydrogen HO1. The hydroxyl in the alternate anomeric position is perturbed into a hydrogen and dummy atom.

All of the simulations presented here were carried out using modified versions of the AMBER 4.1 suite of programs[37] and force field.[31] Free energy calculations were carried out using both FEP and TI. Obtaining converged free energy results using either method depends on calculating converged ensemble averages for *each* of the many windows that constitute the perturbation. For flexible solutes, the requirement of properly sampling all relevant conformations using molecular dynamics can be especially problematic. Transitions between conformations may occur as rarely as several hundred picoseconds apart, requiring nanoseconds of simulation during each of the windows to obtain correct equilibrium distributions. In the case of glucose, considering a single chair conformation and three rotamers for each of the six rotatable bonds results in 729 possible conformations of each anomer. Previous calculations[6] have suggested that many of these conformers are populated at room temperature in vacuo. With current computational resources, however, exhaustive sampling for such systems is often unfeasible. It may be impractical to determine not only which conformations are populated (especially with explicit solvent) but also how the calculated free energy may be affected by insufficient sampling.

In addition to the FEP and TI calculations, free energies were also calculated employing the partition functions of the two anomers:

$$\Delta G_{\beta\rightarrow\alpha} = RT \ln(Q_\alpha/Q_\beta)$$

$$Q_\alpha = \sum_i e^{(-E_i/kT)}$$

where Q_α is the partition function for anomer α. The partition function for each anomer was estimated as follows: a conformational search was performed, using a torsional search with a grid size of 120°. Each of the 729 conformations of the anomer was minimized in vacuo with constant dielectric constant. A sum over all unique structures (typically ~80) with Boltzmann weighting provided a value for the approximate partition function. Since only minimized structures contributed to the partition function, the free energy includes conformational, but not vibrational, entropy. This number is therefore only an estimate of the actual free energy difference. In addition, our use of explicit, rather than continuum representation of the solvent prevents us from using this method to calculate the free energy difference in solution.

To be consistent with the TIP3P water model,[38] the Cornell et al. force field does not include a van der Waals term for hydroxyl hydrogen atoms. This can be problematic when performing free energy perturbations. The hydrogen atom is normally located within the repulsive sphere of the oxygen atom, preventing atoms with negative charge from approaching the positive hydrogen too closely. During the course of creating/removing the hydroxyl group, the oxygen radius is sufficiently reduced that the hydrogen becomes accessible, resulting in potentially unstable simulations.

Electrostatic decoupling is often employed to avoid such difficulties. Two simulations are performed; the first involves removal of the charge from the disappearing atoms, followed by changes in the van der Waals parameters. In the case of glucose, however, when the hydroxyl in the β position is being removed, the corresponding group in the α position is appearing. Decoupling in this case requires at least three separate simulations: mutating the β hydroxyl charges, followed by the vdW changes, and finally changing the α hydroxyl charges. In addition, removing the extra LES copies of the hydroxyl groups would have required decoupling for the correction legs of the free energy cycle, necessitating seven simulations to calculate the $\alpha \rightarrow \beta$ free energy. We therefore sought an alternate method that would provide stability without requiring numerous simulations.

One approach that we investigated involved simply adding a Lennard-Jones parameter to the hydroxyl hydrogen atoms. The ϵ value was chosen to be 0.0157 kcal/mol, the same as for the AMBER nonpolar hydrogen atom type. A range of r^* values was used to calculate the energies of a water dimer and the density and heat of vaporization of a periodic box of water. A value of 0.2 Å was the largest radius that did not significantly affect any of these properties and was therefore adopted for all simulations. In addition, since this parameter was applied in both the α and β conformations, it is unlikely that it would significantly affect the free energy difference between the two. Test simulations without this modification, employing the lengthy decoupling procedure or the potential energy function changes described below, provided comparable free energies.

We also explored the use of a soft-core potential energy function.[39] For an atom disappearing at $\lambda = 1$, this function has the form:

$$V_{nb} = (1-\lambda)\sum_{i<j}\left\{\left[\frac{A_{ij}}{(\alpha_{LJ}\sigma_{ij}\lambda^2 + r_{ij}^6)^2} - \frac{B_{ij}}{(\alpha_{LJ}\sigma_{ij}\lambda^2 + r_{ij}^6)}\right] + \frac{q_i q_j}{4\pi\epsilon(\alpha_C\lambda^2 + r_{ij}^2)^{1/2}}\right\}$$

where α_{LJ} and α_C are the soft core parameters for Lennard-Jones and electrostatic terms, respectively. Previous tests[25] found that suitable values for these parameters using AMBER were 0.5 Å^5 for α_{LJ} and 15.0 Å^2 for α_C.

This altered potential function is identical to the standard AMBER functional form at the two end points of the perturbation and should therefore not affect the calculated free energies. At intermediate λ values, however, the interaction at short interatomic distances is smoothed and the singularity present at the original end points is removed. Simulations using this function were stable and did not require decoupling.

Locally Enhanced Sampling Protocol. Locally enhanced sampling[25] (LES) is a mean-field method that has been applied to a variety of problems, including the study of ligand diffusion pathways in proteins,[27,40,41] optimization of conformations of peptides[28,29] and protein side chains,[26] protein loop modeling,[42] small molecules,[43] and ligand docking.[44,45] LES provides the ability to focus effort on a portion of the system, making additional copies of those parts where conformational sampling may be critical. In the present case, we employ multiple copies of those parts of the molecule where multiple conformers are possible. In LES, each of the copies does not interact with any other copies of the same group, and each sees the other groups as an average of all copies. An important point is that an atom sees the mean force from all of the copies, not a force from an average conformation of the copies. The copies may simultaneously explore alternate regions of conformational space, providing multiple trajectories for the duplicated portion at a much lower computational expense than repeated simulation of the entire original system.

An additional advantage of LES is that the barriers to conformational transition are reduced, resulting in a smoother potential energy surface, with more rapid transitions and efficient sampling of the various accessible minima. LES is therefore well suited to free energy calculations, where it is important to obtain the many required ensemble averages at a minimal computational cost. The use of LES for molecular dynamics and free energy calculations has been described previously;[26,30] hence, the derivation will not be presented in full detail. We do note, however, that free energies calculated directly for the smoother LES energy surface are not the same as those from the original

(37) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(39) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529−539.

(40) Czerminski, R.; Elber, R. *Proteins* **1991**, *10*, 70−80.

(41) Verkhivker, G.; Elber, R.; Gibson, Q. H. *J. Am. Chem. Soc.* **1992**, *114*, 7866−7878.

(42) Zheng, Q.; Kyle, D. *J. Mol. Biol.* **1994**, *19*, 324−329.

(43) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *Biopolymers* **1996**, *39*, 103−114.

(44) Miranker, A.; Karplus, M. *Proteins* **1991**, *11*, 29−34.

(45) Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisis, C. *J. Mol. Biol.* **1993**, *234*, 515−521.

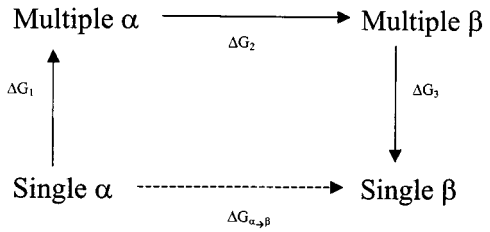

Figure 2. Free energy cycle used with LES. The dashed line represents the original perturbation. The solid lines represent the path taken with LES. Since free energy is a state function, and independent of path, both methods should converge to the same value.

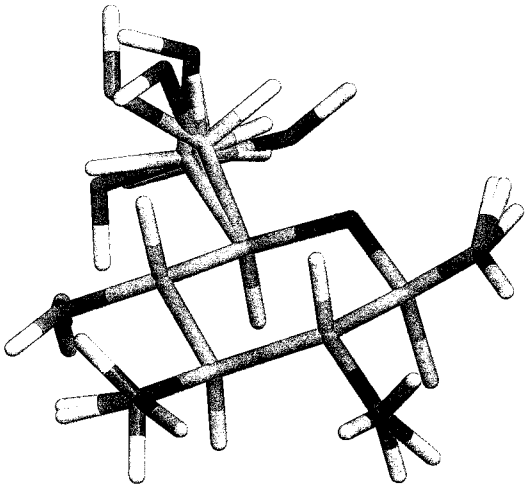

Figure 3. β-Glucose with five LES copies of each hydroxyl hydrogen and the hydroxymethyl group.

energy function. We therefore need to correct the free energies that are calculated using the modified system.

This correction is calculated through the construction of a thermodynamic cycle for our calculation (Figure 2). We refer to this as the LES1 model. In the first leg, we create the additional copies of groups for which we desire enhanced sampling. In the case of glucose, this corresponds to a perturbation of each of the hydroxyl hydrogen atoms from single to multiple copies, resulting in multiple copies of each of the hydroxyl dihedral terms. We also create additional copies of the entire exocyclic hydroxymethyl group and the dummy atom corresponding to the alternate anomer (Figure 3). At the beginning of the copy creation phase, four of the copies are dummies and have no nonbonded interactions, while one "real" copy has the full parameters found in the original system. The real and dummy copies for all groups are then gradually mutated until all five copies are identical, with each possessing $1/N$ (for N copies) of the Coulomb and van der Waals interactions of the original atoms. For copy removal, this process is simply reversed. In addition to increasing simulation stability during annihilation of polar hydrogen atoms, the soft core treatment described above was also found to be useful during removal of the extra copies of the relatively large hydroxymethyl group, again eliminating the need for decoupled simulations. The masses of copied atoms were always the same as the original atoms and did not change during the perturbations.

In the next leg of the cycle, each of the copies of the anomeric hydroxyl is mutated to hydrogen, while the copies of the alternate anomer are created from the dummy atoms. Copies of other parts of the molecule are unchanged during this step. This simulation corresponds to the normal $\alpha \rightarrow \beta$ FEP calculation, with the distinction that multiple copies are present. The final leg involves removal of all of the extra copies, returning to the original (single-copy) state, but in the alternate anomer. There are several possible options for treating the covalent terms of the force field for the disappearing atoms. To maintain a reasonable topology of the molecule, we leave the bond and angle force constants at the full original value for all copies, at all stages of the perturbation. We scale the dihedral force constants by $1/N$ (N copies) for the states when all copies are equivalent and restore the original force constant for the real copy when the others are removed. The dihedral force constants for the dummy copies are set to zero at the end point where the atoms disappear. Several alternate treatments for the dummy copies were tested and not found to significantly affect the final results.

Since the end points of the new perturbation are identical to the original system and free energy is a state function, following this alternate path to the free energy calculation should provide exactly the same result. Moreover, since we have enhanced the sampling of conformational space, we expect the LES method to converge to this final value with substantially reduced computational effort.

The optimal number of copies is a compromise between two factors. First, we require enough copies to effectively sample the accessible conformations. The second is that additional copies will further alter the energy surface, increasing the difficulty in calculating the correction portions of the LES free energy cycle. Previous LES free energy calculations[46] on molecules of similar size using a variety of numbers of copies found five to be a reasonable compromise. Accordingly, five copies of each group were employed for all LES calculations reported in this work. Since the hydroxyl hydrogen and hydroxymethyl group were each copied separately and interact with other groups in an average way, at each step of the simulation, 15 625 (5^6) points on the original energy surface contribute to the LES energy and forces.

The original perturbation involving the calculation of the anomeric free energy difference is relatively modest—one hydroxyl hydrogen atom is removed and another is grown, with a corresponding mutation of oxygen into hydrogen and vice versa. In contrast, the legs involving the creation and removal of copies are much more elaborate—36 particles are changed. It is reasonable to ask whether this creation of copies may be even more troublesome than the original perturbation. However, since it is possible for the copies to overlap, cavity creation (which can be problematic) is not required. No changes in phase or periodicity of dihedral terms are required (in contrast to the α to β perturbation). In addition, the dihedral force constants and nonbonded interactions are partially reduced during the majority of the copy removal simulation, so much of the net $\alpha \rightarrow \beta$ perturbation benefits from considerably increased sampling of conformational space. Another potential drawback is that the cycle end points, where only one copy remains, are still in a single conformation. The free energy obtained in this way is equivalent to perturbing a single conformation of glucose in the α state to a single β state, which may not reflect the actual ensemble that contributes to the experimental result.

In an attempt to minimize the impact of these potential limitations, we propose an alternate free energy cycle. In this approach, we do not remove *all* of the LES copies, but only the extra copies of the anomeric hydroxyl hydrogen. This corresponds to performing the original perturbation of α to β, but with the anomeric hydroxyl surrounded by a distribution of possible conformations for the other groups in the molecule. The end points of this new cycle are not the same as in the original system, and the free energy is therefore no longer guaranteed to be the same as the original perturbation. However, this strategy may be a more realistic representation of the actual ensemble of states that contribute to the experimentally measured free energy difference. Moreover, the magnitude of the perturbation is reduced, with fewer mutations performed and enhanced sampling is achieved for all but the anomeric center during the entire cycle. For the remainder of this manuscript, this approach will be referred to as the LES2 model.

We tested one further variation of the LES method in which copies were made of all of the atoms described previously with the exception of the anomeric hydroxyl group, which was never copied. Since the charges on these atoms are reduced during the mutation, weakening of hydrogen bonds relative to the other hydroxyl groups may reduce the tendency to remain in a single conformation. Since there are no copies of the original perturbed group, the end points of this simulation correspond exactly to those described in the previous paragraph (LES2), without the requirement of separate correction simulations. Instead, a single simulation directly provides us with the (approximate) anomeric

(46) Simmerling, C. Unpublished results.

free energy difference. The free energies obtained from this approach (designated LES3) and the other methods are presented in the Results.

It is important to note that, since LES only provides approximate dynamic behavior, the actual conformer populations are not necessarily correct.[47,48] One must therefore use caution when interpreting such populations. Modified multiple-copy schemes which do furnish Boltzmann-weighted populations have been proposed.[43,49] It is unclear, however, whether these techniques are suitable for use in explicit solvent or provide all of the advantages of LES. However, any population differences are accounted for in the correction legs of the free energy cycle. The actual free energy calculation (in principle) converges to the correct value, and the LES1 method involves no approximation relative to the standard approach.

Molecular Dynamics Protocol. The time step used for all simulations was 2 fs. A residue-based cutoff of 8 Å was applied to nonbonded interactions in solution. A value of 10 Å was tested in solution and yielded nearly identical results. A neighbor list was used and updated each 10 steps. All bonds were constrained to their equilibrium lengths using the SHAKE[50] algorithm, and the bond PMF correction[51] was employed. A constant dielectric constant of 1.0 was used. Scaling factors for 1−4 electrostatic and VDW interactions were 1.2 and 2.0, respectively. For the simulations in water, glucose was solvated by 357 TIP3P38 water molecules in a periodic box approximately 25 Å in each dimension. Pressure was maintained at 1 atm.

The temperature was coupled to an external bath at 300 K. For calculations in vacuo, the lack of collisions with water molecules resulted in poor redistribution of energy, often resulting in a significant accumulation of kinetic energy in a single rotation (especially for dummy atoms) or rigid body motion of the entire molecule. Although the average kinetic energy for all atoms corresponded to a temperature of 300 K, parts of the system were essentially frozen. To avoid this problem, an alternate coupling algorithm was used in vacuo, where velocities for all atoms were reassigned[52] at the start of each window, prior to the equilibration phase. After reassignment, the standard algorithm[53] in AMBER was employed to maintain constant temperature throughout the remainder of the window. Multiple copies were initially assigned identical coordinates but moved apart after velocity assignment and equilibration.

Except where noted, each simulation was carried out using FEP with the soft core energy function and consisted of multiple (51 or 101) windows, with 2 ps of equilibration followed by varying amounts of data collection to determine the total simulation length. Free energies were calculated using all interactions, including those between perturbed atoms. Uncertainties were calculated as one-half the difference between $\alpha \rightarrow \beta$ and $\beta \rightarrow \alpha$ (or copy creation/removal) simulations of the same length. For free energies that are the sum of values obtained from several simulations, net uncertainties were calculated as the square root of the sum of squares of the individual uncertainties. As others have noted in the past, this uncertainty is only a lower bound to the actual error in the calculation.

Potential of mean force (PMF) calculations were also performed, in which the free energy as a function of dihedral angle was calculated. For these simulations, a rotation of 360° was carried out over 51 or 101 windows, each consisting of 1 ps of equilibration and varying amount of data collection. Free energies were calculated using TI and the constraint forces method.[54]

Results and Discussion

This section will be divided into several parts. First, we present the final free energy results obtained with each model, for simulation both in gas phase and in solution, calculated using FEP or the approximate partition function. All free energies are reported for the $\alpha \rightarrow \beta$ direction. We also compare individual conformational energies and populations of the hydroxymethyl group rotamers with those obtained from quantum mechanical calculations and experiment. Finally, we discuss the specific details of the conformational sampling and convergence of each of the different models.

Table 3. Anomeric Free Energy Differences in Vacuo Calculated Using FEP with Standard AMBER and after Addition of Torsional Parameters for the OCOC and OCOH Fragments[a]

	$\Delta G\ \alpha \rightarrow \beta$	
force field	FEP	Boltzmann
Cornell et al. without anomeric terms	-2.33 ± 0.02	-0.37
Cornell et al. with anomeric terms	0.63 ± 0.05	0.36

[a] A single copy was used with soft-core potential modifications and the 12 conformer charge model.

Free Energies. In Table 3 we list the results of single-copy calculations with and without the specific anomeric dihedral terms that were described in the Methods. The charge set fit to 12 conformers was used. The free energy obtained in the absence of the specific dihedral terms (−2.3 kcal/mol) strongly favors the β anomer, in disagreement with high-level ab initio calculations (0.4 kcal/mol)[6] which we consider our reference value. This is not unexpected, since the anomeric effect in glucose favors the α form and was not explicitly accounted for in this model. However, including the specific dihedral terms for OCOC and OCOH results in excellent agreement (0.6 kcal/mol) with the reference value. The free energies calculated from the approximate partition function do not compare well with those from FEP calculations. The importance of the vibrational entropy contribution to the anomeric free energy in glucose has been reported in the past.[6,55]

We next proceeded with the LES simulations in vacuo, which required approximately twice the CPU time of single-copy simulations of the same length. We list the free energies obtained for each of the three legs of the LES cycles in Table 4a for simulations in vacuo. In each case, the magnitude of the net free energy correction from the extra copies ($\Delta G_1 + \Delta G_3$) is relatively small, under 0.3 kcal/mol. In both models the α anomer is favored in vacuo, with each model near the reference value[6] (0.4 kcal/mol). Removing all copies (LES1) slightly reduces the preference for the α anomer compared to the models in which some copies remained (LES2 and LES3). However, the LES1 model also results in a larger hysteresis, and more difficulty in achieving convergence was anticipated due to the magnitude of the perturbation involved. The larger perturbation is also reflected in the increased value for the free energy of removing the copies. Specific details of the convergence properties of the free energies are presented in a later section.

In Table 5 we provide the calculated free energies in solution, for both the single-copy and LES systems. Values for the individual legs of the LES cycles are provided in Table 4b. In this case, the LES simulations required only 30% additional computational effort. The single-copy and LES1 models both slightly favor the α anomer in solution, in contrast to the experimental finding that β is favored ($\Delta G = -0.3$ kcal/mol). However, the deviation is only 0.5 kcal/mol, near the limit of the accuracy of this method. Additionally, these simulations are not well converged, despite the low hysteresis. This number

(47) Straub, J. E.; Karplus, M. *J. Chem. Phys.* **1991**, *94*, 6737−6739.
(48) Ulitsky, A.; Elber, R. *J. Chem. Phys.* **1993**, *98*, 3380−3388.
(49) Huber, G. A.; McCammon, J. A. *Phys. Rev. E* **1997**, *55*, 4822−4825.
(50) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(51) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532−4545.
(52) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.
(53) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(54) Pearlman, D. A. *J. Chem. Phys.* **1993**, *98*, 8946−8957.
(55) Senderowitz, H.; Parish, C.; Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 2078−2086.

Table 4. Free Energies (kcal/mol) for Each of the Three Legs of the LES Cycles (a) in Vacuo and (b) in Water[a]

	(a) in vacuo		(b) in water	
	LES1	LES2	LES1	LES2
ΔG_1: copy creation, α anomer	4.44 ± 0.09	0.49 ± 0.01	12.28 ± 0.09	3.10 ± 0.04
ΔG_2: α → β	0.52 ± 0.01	0.52 ± 0.01	−0.02 ± 0.04	−0.02 ± 0.04
ΔG_3: copy removal, β anomer	−4.67 ± 0.03	−0.21 ± 0.03	−12.08 ± 0.03	−3.25 ± 0.01
net ΔG α → β	0.29 ± 0.10	0.80 ± 0.03	0.18 ± 0.10	−0.18 ± 0.06

[a] The 12 conformer charge model was used with FEP and soft-core modifications. The LES3 model is not included since-copy removal simulations are not required.

Table 5. Anomeric Free Energies (α → β) Calculated Using FEP, the 12 Conformer Charge Model, and the Soft Core Energy Function[a]

method	ΔG, water	ΔG, vacuum	$\Delta\Delta G$
single	0.13 ± 0.09	0.90 ± 0.15	−0.77 ± 0.17
LES1	0.18 ± 0.10	0.29 ± 0.10	−0.11 ± 0.14
LES2	−0.18 ± 0.06	0.80 ± 0.03	−0.98 ± 0.07
LES3	−0.13 ± 0.09	0.85 ± 0.02	−0.98 ± 0.09

[a] The different LES models are described in the text.

Table 6. Anomeric Free Energies (α → β) Calculated Using FEP, the 2 Conformer Charge Model and the Soft Core Energy Function[a]

method	ΔG, water	ΔG, vacuum	$\Delta\Delta G$
single	0.08 ± 0.01	0.43 ± 0.16	−0.35 ± 0.16
LES1	0.25 ± 0.40	0.60 ± 0.18	−0.35 ± 0.44
LES2	0.17 ± 0.07	1.01 ± 0.07	−0.84 ± 0.10

[a] The different LES models are described in the text.

should be considered only a lower bound of the actual uncertainty. Results obtained from TI and FEP calculations differed by at least several tenths of a kilocalorie for both cases, demonstrating that the small hysteresis is misleading. This will be discussed in more detail in a following section.

The free energies obtained from the LES2 and LES3 models favor the β anomer in solution, and both are in excellent agreement with the experimental value of −0.3 kcal/mol, which is within the uncertainty range of the calculated values. In addition, results obtained from LES2 using TI and FEP differ by less than 0.1 kcal. These simulations converge much more rapidly than those from the single-copy or LES1 models, most likely because more than one conformation is represented during all stages of the perturbation. However, the LES3 model converged less rapidly than the LES2 model, in part due to the presence of only a single copy of the anomeric hydrogen atoms during the α → β stage.

The degree to which solvation contributes to the true relative populations of α and β is not well understood. Subtracting the free energy value in solution from that obtained in a vacuum provides the solvation contribution for our model. These $\Delta\Delta G$ values are presented in Table 5. In all of the methods used, solvent favors the β anomer by approximately 0.5 to 1 kcal/mol. This is in contrast with results determined by previous calculations,[15] which found a larger effect (−3.03 ± 0.5 kcal/mol, preferring β), or that solvation favored the α anomer by 0.5 kcal/mol.[17] It is, however, in excellent agreement with results from high-level ab initio calculations combined with a continuum solvent model which found the solvation contribution to be −0.6 kcal/mol.[6]

Finally, in Table 6, we present free energies obtained for a similar set of simulations in which the charge set derived from only the lowest energy conformations of α and β (model B) was used. For both charge models, the α conformer is significantly favored in vacuo and solvation favors β by ∼1 kcal/mol using the most reliable LES2 approach to calculating the free energy. Although model A leads to ΔG in better agreement with experiment, the ΔG value for model B is only 0.35 kcal/mol different in solution. Given that the number of conformations to be used in a RESP fit is somewhat arbitrary, it has been important to show that the number of conformations included in the charge fitting procedure does not significantly affect the results. Encouragingly, the model that includes more conformational heterogeneity seems to fit better with experiment.

Barrows et al. performed high-level quantum mechanical calculations on a set of 11 conformations of glucose which differed in the hydrogen bond pattern, hydroxymethyl conformation, and anomer.[6] The authors used the relative energies to evaluate various QM and MM models, including a previous AMBER carbohydrate force field[12] which provided an average unsigned error for six available conformations of 1.9 kcal/mol. We minimized all 11 conformations using our new parameters and the 12 conformer charge model, obtaining an average unsigned error of 0.6 kcal/mol, with the largest deviation being 1.3 kcal/mol. Using the two-conformer charge model resulted in an average error of 0.7, with a maximum error of 1.8 kcal/mol. This is in agreement with our observations that the 12 conformer model also does a slightly better job of reproducing the reference values for the anomeric free energy difference.

Two conformations of β-glucose with an alternate chair conformation ($^1C_4$) were studied using the same level of ab initio calculations.[56] We manually constructed conformations for β-glucose on the basis of the figures in that manuscript and minimized them using the same procedure as described above for the 11 $^4C_1$ conformations. While our model structures did not retain all of the initial hydrogen bonds, we find that these structures have energies ∼8 kcal/mol higher than the lowest energy $^4C_1$ conformations of β-glucose, similar to the "consensus energy" difference of ∼7 kcal/mol reported by those authors.

These results demonstrate the validity of the AMBER philosophy of force field development—using high-quality RESP charges, standard Lennard-Jones atom types, and dihedral parameters transferred from small molecule models. The results obtained, both in vacuo and in aqueous solution, are quite close to the reference values. Of course the force field will need to be tested on other carbohydrates to determine if it is uniformly accurate or requires further improvement. The issue of which method is most efficient for calculating such free energies is the subject of the next section.

Sampling and Convergence. Since simulations in the absence of solvent molecules are less computationally demanding, we performed more extensive testing for each method in vacuo. The results of these simulations will be discussed first. However, since the solvent environment can have a significant effect on the simulations, we also point out differences when appropriate. Before examining the details of the perturbation simulations, it is useful to evaluate the dynamics at each of the end points—in this case, the individual anomers. In Figure 4 we show the dihedral values of the six rotatable bonds in α

(56) Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydr. Res.* **1995**, *276*, 219−251.

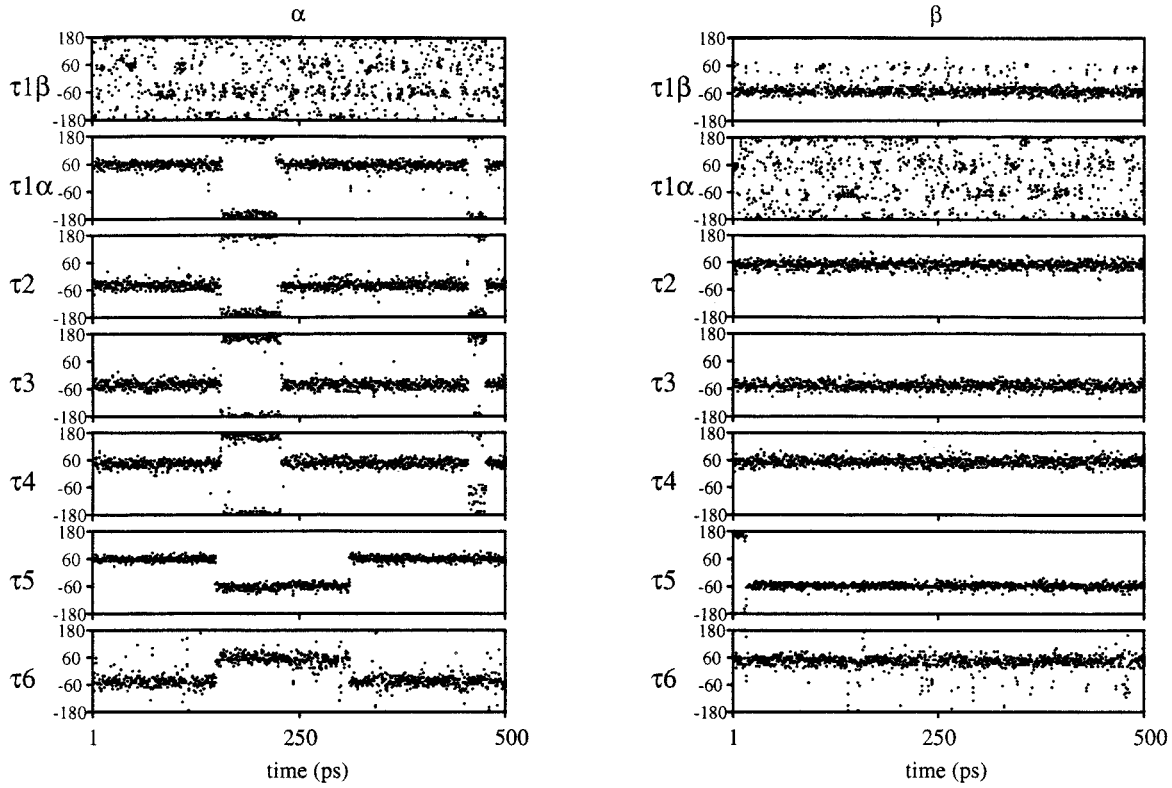

Figure 4. Torsion angle sampling during 500 ps of dynamics for both anomers, using a single copy. Conformational transitions are separated by several hundred picoseconds. Since the dummy atom does not form hydrogen bonds, transitions for the dummy dihedral are rapid.

and β glucose in vacuo as a function of time, over 500 ps of simulation—much longer than would typically be used for the data collection phase of a single perturbation window. It is clear that on this time scale very few, if any, conformational transitions are observed. If more than one conformation should contribute to the actual ensemble, the calculated free energy difference may be incorrect.

Hydrogen Bond Network. The β anomer exhibits only a single conformation (counterclockwise) of the hydrogen bond network during the entire 500-ps simulation. The α anomer, however, undergoes several transitions, each time spending a short period in the clockwise arrangement before reverting back to counterclockwise. To estimate the thermal accessibility of these conformations, we performed simple energy minimization on both patterns for α and β, retaining the GT conformation for the hydroxymethyl group. We found an energy difference of 4.5 kcal/mol in β, corresponding to a 0.0005 relative population of the clockwise arrangement. In α, however, the energy difference was reduced to only 2.3 kcal/mol, with a population ratio of 0.02. Similar quantum mechanical calculations for TG conformers also resulted in a lower energy penalty for the clockwise pattern in the α anomer relative to that for β ($\Delta\Delta E = 1.4$ kcal/mol).[6] This may explain why this conformation is observed in the α simulations but not those of the β anomer. However, the high barriers to rearranging these hydrogen bonds, and the resulting infrequency of transitions, make obtaining an equilibrium distribution during individual windows difficult. In fact, the system often becomes stuck and spends an inappropriate amount of time in the less favorable clockwise pattern. A specific example will be presented in a later section.

Hydroxymethyl Group. Data from quantum mechanical calculations[4,6] and NMR experiments[57] suggest that more than one conformation of the exocyclic hydroxymethyl group should be populated in vacuo and in aqueous solution. However, this is not reflected in the single-copy MD simulations, which rarely undergo transitions (Figure 4). This should be contrasted with the simulation using LES copies (Figure 5). The reduction in barrier heights and simultaneous exploration of multiple minima lead to significantly improved sampling of the accessible conformational space, for rotation about both C−O and C−C bonds.

Two separate issues affect the populations obtained from MD simulations. First, the potential energy function may be a poor model for the actual energy surface. Furthermore, the barriers between the minima may preclude achieving the proper distribution even with an accurate force field. We attempt to separate these issues by calculating the PMF for rotation about the C5−C6 bond in the α and β anomers, using single-copy and LES systems both in vacuo and in solution. Similar calculations (without LES) have been carried out for glucose in the past.[17] Since the relative well depths in LES are not necessarily the same as those of the original system, the populations obtained are not expected to be in quantitative agreement with the original energy surface. However, we explore only barrier heights and whether qualitative agreement and reasonable ordering of the various conformations in glucose are obtained using the multiple-copy approach.

In the LES system, only a single copy was rotated, leaving the other copies unrestrained. If all copies were moved simultaneously, the PMF would be similar to the single-copy system. An important advantage to LES, however, is that the copies can be moved independently, crossing several smaller barriers rather than a single large one. Several simulations were performed for each system, with total simulation ranging from 100 ps to 3 ns. The results for the α anomer in vacuo are presented in Figure 6. It is immediately apparent that the

(57) Nishida, Y.; Ohrui, H.; Meguro, H. *Tett. Lett.* **1984**, *25*, 1575.

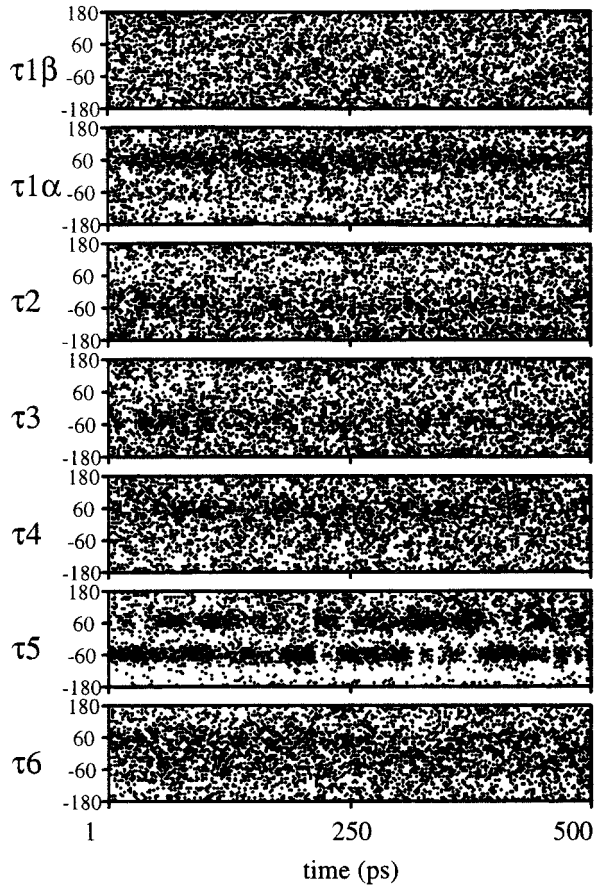

Figure 5. Dihedral angle sampling for five LES copies of the α anomer in vacuo. The combination of reduced barriers and multiple copies results in extensive sampling of conformational space.

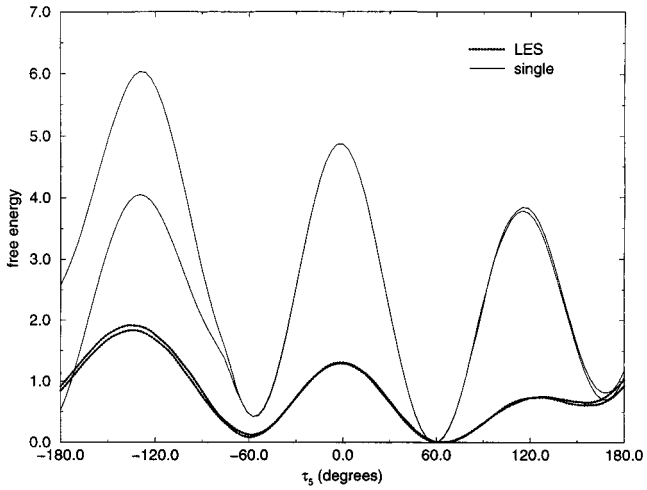

Figure 6. PMF curves for rotation about the OCCO torsion for the α anomer in vacuo. The solid lines represent forward and backward simulations of 3.2 ns each using a single copy, showing nearly 2 kcal/mol difference. The dotted lined represents values obtained from 400-ps simulations using LES. The barriers are significantly reduced and convergence is improved.

barriers to rotation about this bond, 4−6 kcal/mol in the non-LES system, have been reduced to only 1−2 kcal/mol through LES.

The net free energy for a rotation of 360° should be zero. Both single and LES systems typically showed decreasing error as simulation time was extended. The LES simulations, however, converged to under 0.1 kcal within 400 ps, while single-copy simulations continued to result in errors above 1 kcal/mol despite extending simulation times to 3 ns.

Table 7. Hydroxymethyl Rotamer Populations Obtained from Methods Described in the Text along with Examples of Values Reported in the Literature[a]

	α			*β*		
method	GT	GG	TG	GT	GG	TG
			Vacuum			
QM[6]	0.37	0.41	0.22	0.41	0.44	0.16
QM[4]	0.03	0.82	0.15			
single MD	0.32	0.68	0.00	0.00	0.97	0.03
single PMF	0.42	0.31	0.26	0.44	0.32	0.24
LES MD	0.43	0.42	0.15	0.46	0.34	0.24
LES PMF	0.40	0.33	0.28	0.40	0.33	0.28
			Solution			
NMR[57]	0.45	0.55	0.00	0.46	0.54	0.00
QM + continuum[6]	0.57	0.39	0.04	0.79	0.15	0.06
single MD	0.97	0.00	0.03	0.97	0.00	0.03
single PMF	0.44	0.42	0.14	0.47	0.38	0.15
LES MD	0.60	0.27	0.13	0.55	0.33	0.12
LES PMF	0.42	0.31	0.26	0.47	0.29	0.25

[a] Populations are given as separate ratios for each anomer; relative ratios of the two anomers were not determined.

Relative populations can be obtained from the PMF curves through a ratio of the sum of Boltzmann weighted energies over segments of the curve. The results are provided in Table 7, along with populations determined by high-level ab initio calculations[4,6] and experiment. Despite the approximate nature of the LES populations, single-copy and LES calculations are in excellent agreement, with the same relative order and average errors of only 3% in vacuo and 7% in solution. These are quite reasonable when we consider the uncertainty in the values obtained from the single-copy simulations and the minor free energy changes that such differences represent.

The populations for the α anomer obtained from the AMBER force field, both in vacuo and in solution, are in reasonable agreement with those from ab initio calculations of Barrows et al.[6] but deviate significantly from those of Brown and Wladkowski.[4] Nearly equal amounts of GT and GG are present, along with a lesser amount of TG. Our results, however, do not show a trend toward an increased population of the GT rotamer in the *β* form, in disagreement with Barrows et al. Senderowitz et al.[55] also found a increased preference for the GT conformer, with an overall GT:GG:TG ratio in GB/SA continuum water of 65:25:10. The experimental results, however, suggest that nearly equal populations of the GT and GG rotamers are present for each anomer,[57] in accord with our data.

An alternate method for obtaining such populations is directly from the ensemble generated during a MD simulation. Comparison with those obtained from the converged PMF assesses the extent of the sampling during the simulation. These values are also presented in Table 7, using 1 ns of dynamics for each anomer. Both in vacuo and in solution, the single-copy MD transitions were rare and populations differ significantly from those obtained from the PMF for the same energy surface. However, the populations for the LES system are in excellent agreement with those calculated from the corresponding PMF and, by 300 ps, showed a deviation of only a few percent from their final values. This demonstrates that MD simulations rapidly converge to the correct distribution *for the LES energy surface*. Despite the differences between this surface and the original one, we feel that better sampling of the approximate surface is preferable to inferior sampling of the correct one. Additionally, use of the complete free energy cycle can correct these differences.

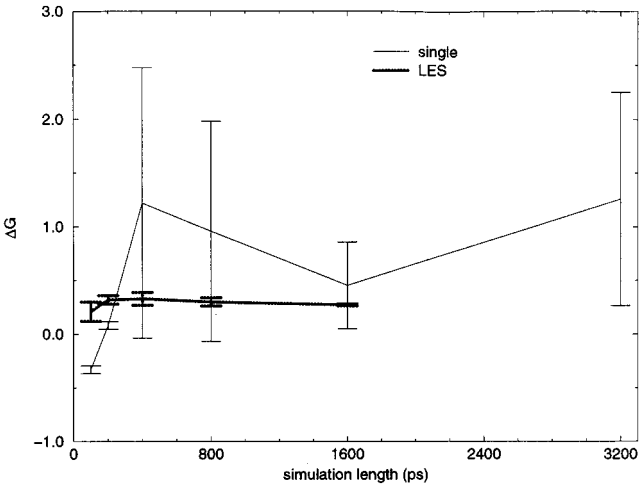

Figure 7. Net free energy for $\alpha \rightarrow \beta$ in vacuo as a function of the simulation length. The LES2 model converges rapidly and shows decreasing hysteresis, while the single copy is not well converged even at several nanoseconds.

Free Energy Convergence. The calculations described in this section were carried out with a preliminary force field (without the anomeric term for the HOCO fragment) and therefore do not provide exactly the same free energies as reported in Tables 5 or 6. However, conformational sampling was not significantly affected by subsequent modifications and the knowledge gained from this large set of calculations was used to determine initial simulations lengths for later models, thus avoiding calculations which were expected to be poorly converged.

In Figure 7 we show the free energies obtained as a function of simulation length, for a set of simulations in which 51 windows were used, combined with 1 ps of equilibration and varying amounts of data collection. Data are shown for both single-copy and LES2 models. The LES2 free energy includes the copy removal correction value of 0.20 kcal/mol, a value which varied less than 0.1 kcal over perturbations of 100−800 ps. These data clearly demonstrate that the LES2 perturbations converge much more rapidly than those using a single copy. The poor convergence can be directly attributed to insufficient conformational sampling.

An excellent example of this inadequate sampling is demonstrated in Figure 8. Again we show the dihedral values as a function of simulation time, but for two separate perturbation ($\alpha \rightarrow \beta$) simulations of 400 ps. Both employed a single copy and identical simulation parameters but were initiated with coordinate sets obtained from different equilibration simulations. Not only do these plots demonstrate the rarity of conformational transitions during the perturbation, but perhaps more disturbing is the lack of similarity in the conformations explored. As described earlier, in the α state, the clockwise hydrogen bond network becomes thermally accessible, though ∼2 kcal higher in energy than the counterclockwise form. However, due to the high energy barriers, a single transition can result in this unfavorable pattern persisting throughout many windows. Similar effects have been observed by other researchers.[17] The calculated free energies differ by nearly 3 kcal/mol—clearly indicating the effect that an incomplete ensemble can have on this type of calculation. Examination of ΔG as a function of λ during the simulation (Figure 9) demonstrates that no single window contributes to this difference, but that the entire free energy curve is altered. A smooth curve is therefore not a reliable indicator of simulation quality. Determining which conformation should be used to initiate simulation of a flexible molecule can be nontrivial and have may a significant impact on the results obtained.

In addition to sensitivity to the initial coordinate set, results obtained from the single-copy simulations varied by similar amounts depending on the method used (FEP vs TI), as well as the relationship between number and length of windows for simulations of the same total time (data not shown). In some cases, the free energy hysteresis was deceptively small due to sampling over similar subsets of conformational space, demonstrating once again that this number is not sufficient evidence for convergence, nor does it always decrease with longer calculations. Only when simulation lengths were extended adequately to allow exploration of alternate minima did the hysteresis become a reliable indicator of the quality of the ensemble averages.

Finally, we observe that the LES simulations converged rapidly (within 200 ps) and demonstrated relatively small hysteresis, which decreases as simulation lengths are increased. Although alternate hydroxyl conformations were extensively sampled, at the conclusion of every LES1 copy removal simulation, only the lower energy (counterclockwise) form was observed, both for α and β anomers. Despite the existence of a single conformation at the end points of the LES1 model, this copy consistently finds the lowest energy conformation, in contrast to non-LES simulations which often become stuck in higher energy local minima. For all LES models, the calculated free energy was insensitive to the choice of initial coordinates.

To determine whether the free energy calculated using LES is correct for this particular force field, however, we need to compare to a well-converged value for the single-copy system. Only by extending the single-copy simulations to 100 ns in each direction were we able to obtain a reliable value of -0.04 ± 0.03 kcal/mol, in excellent agreement with the value of -0.20 ± 0.05 kcal/mol obtained rapidly and consistently with LES.

For simulations in solution, the barriers to rearrangement of the internal hydrogen bonds are significantly reduced, since alternate hydrogen bonds can be made with water molecules. This is reflected in increased transition frequency for several hydroxyl rotamers during 1 ns of single-copy MD (Figure 10). However, such transitions for the anomeric hydroxyl are less common, and the hydroxymethyl group remains in a single conformation throughout the entire simulation. It is difficult to tell a priori whether such limited sampling will affect the calculated free energy. LES calculations again show extensive sampling, similar to that observed in vacuo.

We again compare the convergence properties of the free energy for single-copy and LES2 models. In Figure 11 we plot the anomeric free energy difference as a function of time for both models. We observe that the single-copy system still demonstrates variations of several tenths of a kilocalorie even for simulations of nanoseconds in length. The LES2 model, however, shows fairly rapid convergence, with less than 0.1 kcal variation over simulations ranging from 400 ps to 1.6 ns. As observed in vacuo, the single-copy hysteresis is a poor indicator of actual convergence. Results from longer simulations often fall outside of the uncertainty range suggested by shorter simulations. The hysteresis obtained using LES, however, does correlate well with the variation in free energy seen by extending simulation length and appears to be a valid indicator of convergence. Similar to what was observed in vacuo, the LES1 and LES3 models did not converge as rapidly as LES2. The LES1 model involves a larger perturbation, with the correction term being a small difference between two very

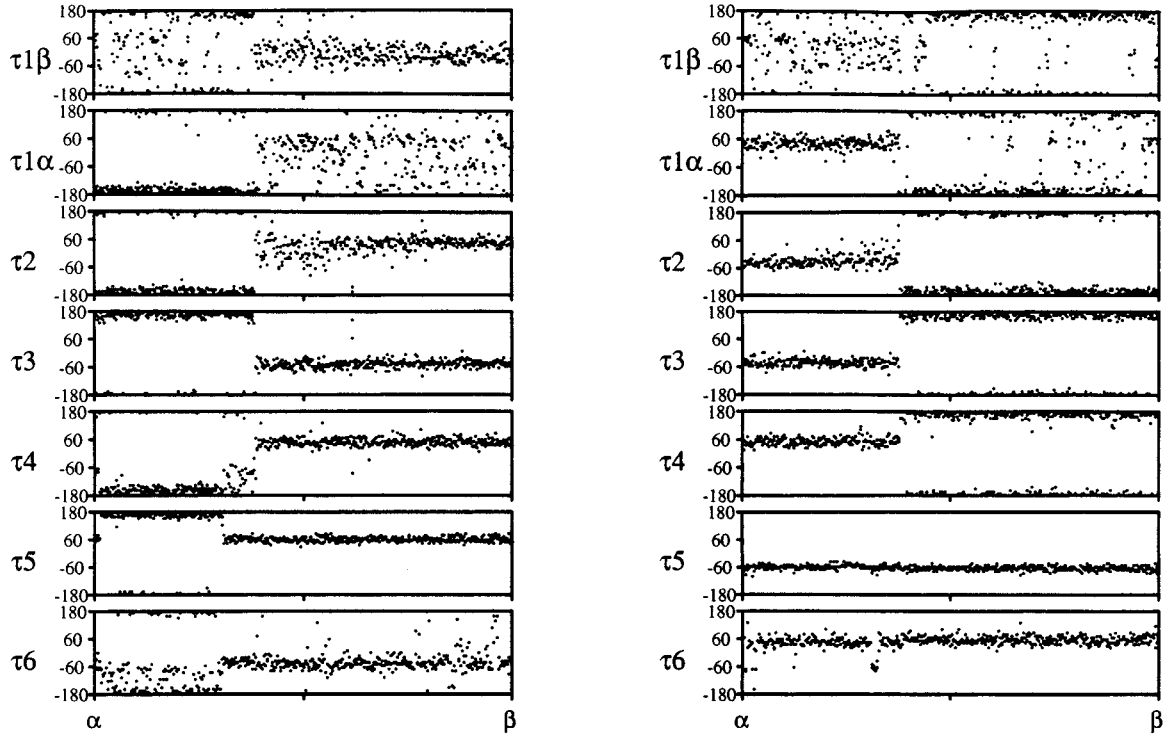

Figure 8. Torsion angle sampling during two independent 400-ps α → *β* perturbation simulations in vacuo. Both undergo a change in hydrogen bond network, but sample different patterns during equivalent portions of the perturbation.

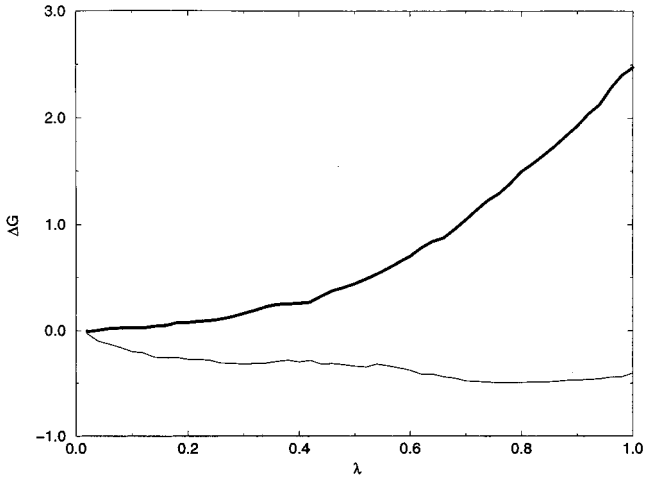

Figure 9. Free energy as a function of the coupling parameter *λ* during two independent 400-ps single-copy α → *β* perturbations in vacuo. The calculated free energies differ by nearly 3 kcal/mol due to differences in dihedral angle sampling.

large numbers. The LES3 model did not employ copies of the anomeric hydroxyl group, which can be seen in Figure 10 to have inferior sampling as compared to other hydroxyl groups, invalidating the assumption made in this model. The single-copy simulations converged most slowly.

Conclusions

We derived new torsional parameters for the OCOH fragment, using dihydroxymethane as a model. These parameters, along with those determined previously[32] for an OCOC fragment, were added to the Cornell et al. force field and combined with RESP multiple-conformation charge fitting for glucose. MD and FEP/TI simulations were carried out both in vacuo and in an explicit model of aqueous solution. Energies of individual conformations and populations of hydroxymethyl rotamers are in excellent

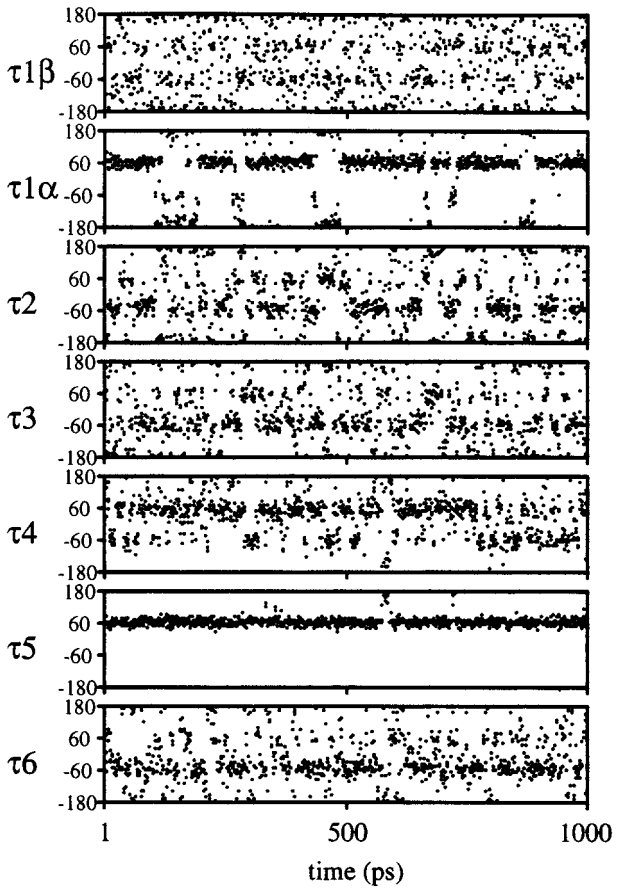

Figure 10. Dihedral angle sampling during single-copy dynamics of the α anomer in solution. Hydrogen bonds to water molecules result in increased frequency of conformational transitions, with the exception of the C−C bond τ_5 and the anomeric hydroxyl group τ_1, which are still slow on the time scale of individual windows during free energy calculations.

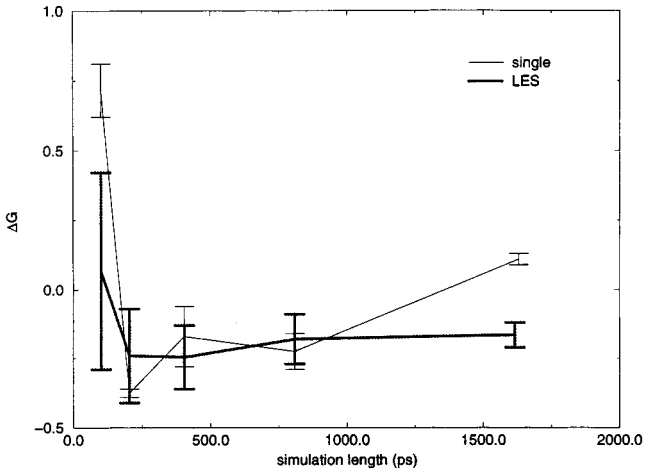

Figure 11. Net anomeric free energies ($\alpha \rightarrow \beta$) as a function of simulation length for single-copy and LES2 models in solution. Similar values are obtained for each model, but LES2 converges within 200 ps, while the single copy shows larger fluctuations even at 1.6 ns. Only the LES2 simulations show decreasing hysteresis with added simulation time. In all LES2 simulations, the uncertainty range includes the final converged value.

agreement with reference values. Calculating the anomeric free energy using either LES or a single copy provided values within 0.5 kcal/mol of the reference free energies obtained from the highest level gas-phase ab initio calculations and the experimental value in aqueous solution.

As important as reproducing the experimental and ab initio data is the definitive insight these studies give us on the contribution to the anomeric equilibrium in glucose. Because we have used a simple, generic force field here,[58] it is possible to dissect the factors governing this equilibrium more easily than when using force fields that use a more complex model and a larger set of torsional potentials or when using only ab initio calculations. For this force field *without* the additional anomeric torsional terms (OCOC and OCOH), the β anomer is favored by 2.3 kcal/mol. This is of the same magnitude but even larger than the 1.75 kcal/mol preference for equatorial over axial methyl cyclohexane,[59] despite the smaller size of the OH than methyl group, presumably because intramolecular electrostatic/H-bonding effects favor the β anomer. The anomeric effect, when added to the force field changes this 2.3 kcal/mol preference for β to a 0.6 kcal/mol preference for α. When one now considers this equilibrium in aqueous solution, better intermolecular hydrogen bonding in the β anomer stabilizes it relative to α by $\sim$0.5−1.0 kcal/mol, a result that is consistent with the findings of Barrows et al.[6] and Schmidt et al.,[21] but not with those of Ha et al.[15] and van Eijck et al.[17] Of course it will be of great interest to see how well this force field approach works on other aldopyranoses[60] and alternate chair forms of glucose,[56] and LES gives us the tools to do this in a definitive way.

The most important result here is a validation of the LES approach to free energy calculations. While exploration of conformational space for the single-copy hydroxyl groups was increased somewhat in aqueous solution, transitions between hydroxymethyl rotamers were rare. We showed that, with LES, the barriers to conformational transition are reduced, and the sampling and convergence properties of the free energy are significantly improved. In addition, the hysteresis obtained from LES simulations provides a more reasonable indicator of convergence, with increasing precision as simulation length is extended. We have presented a new version of the LES approach that converged the most rapidly and provided results closest to the reference values of the variants of LES, in which the additional copies of the perturbed group (the anomeric hydroxyl hydrogen) were removed, but multiple copies of all other groups were allowed to remain. This model is appropriate for systems where conformational sampling in the nonperturbed part of the molecule is important and may be the best representation of the ensemble of conformations which actually contribute to the experimental equilibrium.

Acknowledgment. This research was supported by Grants NIH-CA-25644 and NIH-GM-29072. The authors thank C. Cramer for providing a preprint of ref 6.

JA972457N

(58) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461−469.

(59) Booth, H.; Everett, J. R. *J. Chem. Soc., Perkin Trans. 2***1980**, 255.

(60) Dowd, M. K.; French, A. D.; Reilly, P. J. *Carbohydr. Res.* **1994**, *264*, 1−19.