JCTC Journal of Chemical Theory and Computation

Evaluating the Accuracy of the Quasiharmonic Approximation

Chia-En Chang,[†] Wei Chen,[‡] and Michael K. Gilson*[,‡]

Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California, and Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received April 6, 2005

Abstract: The quasiharmonic approximation (QH) allows the configurational entropy of a molecule to be estimated on the basis of a molecular dynamics simulation, through construction of a Gaussian probability distribution of conformations with variances equal to those provided by the simulation. At its introduction in 1981, the QH method was successfully applied to simple molecular systems with only one highly occupied energy well, and fluctuations were analyzed in a system of internal bond-angle-torsion coordinates. However, more recent studies have applied the QH method to complex biomolecular systems and have relied upon Cartesian coordinates. The present study evaluates the accuracy of the QH method through comparisons with more detailed methods. The chief findings are that the QH method can markedly overestimate the configurational entropy for systems with multiple occupied energy wells and that such errors tend to be magnified by the use of Cartesian coordinates instead of bond-angle-torsion coordinates.

1. Introduction

Methods of calculating free energies are of central interest in molecular modeling and computational chemistry as a basis for interpretation, prediction, and design. Statistical thermodynamics dictates that free energy is essentially an integral over all configurations of the system that are accessible at ambient temperature. This can be a very difficult quantity to compute, especially for large, flexible molecules. One of the more rigorous approaches to this problem is to use free energy integration (see refs 1−11 and citations in ref 12), which uses molecular dynamics (MD) or Monte Carlo simulations to compute the reversible work, or free energy change, for a molecular process such as a change in conformation. The chief drawback of this approach is that it tends to be computationally demanding[13] and, hence, subject to convergence problems.[14,15] Less rigorous but more efficient approaches include linear interaction energy models[16−18] and the MMPB/SA method (e.g., refs 19−21).

Recently, we have developed a formally rigorous method of computing free energies called the second-generation mining minima (M2) algorithm. This comprises two major steps: identifying the low-energy conformations of the molecular system and calculating local configuration integrals—effectively free energies or standard chemical potentials—within each energy well. These local integrals employ an enhanced form of the harmonic approximation that accounts for anharmonicity via local scans of the energy surface. The contributions from the various energy wells are then combined to provide the full free energy of the system. Thus, a binding free energy, for example, is computed as the difference in the standard chemical potential of the product complex versus the free reactants. This approach requires the use of an implicit solvation model in order to keep the number of energy minima to a manageable number. The M2 method has proven to be both accurate and tractable for a range of host−guest systems, including synthetic receptors and cyclodextrins.[22,23] However, in its present form, at least, the M2 algorithm is not directly applicable to proteins

* Corresponding author phone: (240) 314-6217, fax: (240) 314-6255, e-mail: gilson@umbi.umd.edu.

[†] University of California—San Diego.

[‡] Center for Advanced Research in Biotechnology.

10.1021/ct0500904 CCC: $30.25
Published on Web 08/16/2005

because of the vast number of local energy minima. We are, therefore, exploring adaptations of the M2 method, as well as other approaches to computing molecular free energies.

One such approach is quasiharmonic (QH) analysis, which was introduced by Karplus and Kushick in 1981 as a means of capturing the anharmonicity of the energy surface when computing configurational entropies.[55] Briefly, the QH method involves carrying out MD simulations, computing the covariance matrix of atomic coordinates, interpreting the covariances as resulting from a harmonic energy surface rather than the actual anharmonic energy surface, and computing effective, or "quasiharmonic", force constants on this basis. These effective force constants can then be used to compute approximations to thermodynamic properties such as entropy, mean energy, and free energy (see, e.g., ref 24). A number of variations in technique can be accommodated within this general description; two key issues are the choice of coordinate system and the range of conformations sampled in the MD simulation, as now briefly reviewed.

In Karplus and Kushick's seminal paper, Cartesian coordinates from the MD trajectory were separated into 6 external (translational and rotational) coordinates and $3N - 6$ internal coordinates comprising bond lengths, bond angles, and bond torsions [i.e., bond-angle-torsion (BAT) coordinates], prior to calculation of the covariance matrix, where N is the number of atoms. (See also refs 25−28.) This separation of coordinates is well-suited to the calculation of thermodynamic properties via classical statistical mechanics because external rotations do not affect the molecular conformation. This is in contrast to the separation of coordinates via the Eckart−Sayvetz conditions,[29−31] which results in external rotations that distort the molecular conformation if they are more than infinitesimal.[27,32] Karplus and Kushick focused on changes in configurational entropy associated with the transition of a molecule from one well-defined energy well to another; that is, butane from trans to gauche and decaglycine from an extended to an α-helical conformation. A subsequent paper by Rojas et al,[33] which derived a cubic correction to the QH approximation, similarly focused on individual energy wells and mapped trajectory snapshots to BAT coordinates before computing the covariance matrix.

More recently, a number of groups have used other forms of the QH method to address important problems in molecular and biomolecular function. Typically, in these applications, Cartesian coordinates from the MD trajectory are not explicitly separated into external and internal coordinates before calculation of the covariance matrix. In addition, the MD simulation is not constrained to remain in a single energy well, although proteins at room temperature are believed to access many different energy wells.[34−37] At least two methods of handling the external coordinates have been described, see refs 38−40. In one study, Schafer et al. examine the entropy of a molten globule state of the enzyme α-lactalbumin. In another, Hsu et al. address changes in configurational entropy upon association of two proteins involved in human immunodeficiency virus infection,[41] also evaluating the convergence of the entropy as a function of simulation length. In a yet a third study, Jusuf et al. use the QH method to argue that changes in vibrational entropy are responsible for the cooperativity of ligand binding by glycopeptide antibiotics,[42] and Rinaldo and Field have used the QH method to examine entropy changes associated with a conformational change of the protein transferrin.[43] Several groups have also used a variant of the QH method in which some components of the covariance matrix are neglected. Thus, Luo and Sharp's application of the QH method to compute binding affinities neglects couplings (covariances) among the translational, rotational, and internal coordinates of each molecule, as well as couplings between the internal coordinates of the two molecules,[44] and thereby permits the equilibrium constant to be factorized into translational, rotational, and internal contributions. Swanson et al.'s examination of binding affinities also factorizes the equilibrium constant and applies the QH method only to the translational/rotational component of the covariance matrix.[45]

Thus, the QH method is finding application to molecular systems of increasing complexity. However, the literature to date provides remarkably little information on the numerical accuracy of the approach, that is, on how well it reproduces the entropy or free energy associated with the actual anharmonic energy surface. In particular, we are not aware of any paper comparing the QH approach against accurate reference results for a flexible molecule that accesses multiple energy minima during the MD simulation. Interestingly, as pointed out by Schlitter,[38] the Gaussian probability distribution assumed in the QH method yields the upper limit of the entropy associated with a given coordinate variance.[46] It is, thus, important to ask whether the upper limit is an accurate approximation for the actual entropy.[47] Interestingly, one can readily construct an energy surface for which it is not a good approximation. Consider, for example, a one-dimensional energy function with two equivalent, deep energy wells separated by a distance $2R$ that is much larger than the width of the wells. The variance of the solitary coordinate is clearly R^2, but one can readily show that the entropy of such a system is much less than that of a single harmonic energy corresponding to a variance of R^2. In fact, the entropy of the dual-well system is independent of R, in contrast to the harmonic well for which the variance cannot be increased without increasing the entropy. On the other hand, this dual-energy model is artificial and, therefore, does not bear directly on the real molecular systems to which the QH approximation is usually applied.

The present study addresses this issue by evaluating the accuracy and convergence of the QH method for a range of molecular systems, from a simplified alkane molecule to a host−guest system for which experimental binding data are available. Implementations of the QH method in both Cartesian and BAT coordinates are examined here, because the choice of coordinate system is known to influence accuracy and convergence in normal mode and related calculations.[28] In particular, Cartesian coordinates tend to provide lower accuracy because modes constructed from a Cartesian basis set necessarily consist of linear displacements of atoms, rather than more natural bond rotations. For the simplest system examined here, QH results are compared to analytic results, while for the more complex systems, reference results are obtained from the M2 algorithm (see

above), which has been numerically validated[22,23,28] and has provided good agreement with experimental binding affinities for a range of host−guest systems.

2. Theory

2.1. Chemical Potentials, Configuration Integrals, and Coordinate Systems. The binding constant K_b of a receptor R and ligand L, which form a noncovalent complex RL, is given by[48,49]

$$-RT \ln K_b = \Delta G^\circ \equiv \mu^\circ_{RL} - \mu^\circ_R - \mu^\circ_L \tag{1}$$

where RT is the gas constant multiplied by the absolute temperature and μ° is the standard chemical potential of the species indicated in the subscript. The standard chemical potential of a molecule in solution may be written as[48,50]

$$\mu^\circ = -RT \ln\left(\frac{1}{VC^\circ} Z'\right) \tag{2}$$

$$Z' = \int \mathrm{e}^{-\beta[U(\mathbf{r}')+W(\mathbf{r}')]}\, \mathrm{d}\mathbf{r}' \tag{3}$$

where $\beta = 1/RT$, C° is the standard concentration, V is the volume of the system, Z' is a configuration integral, and U and W are the potential and solvation energies of the molecule, respectively, as a function of the molecule's $3N$ Cartesian coordinates $\mathbf{r}'$. It is convenient to separate the coordinates into 6 external coordinates and $3N - 6$ internal coordinates $\mathbf{r}$ using the criterion that changing the external coordinates does not affect U or W. This allows the integrals over the external coordinates to be carried out trivially, yielding a factor of $8\pi^2 V$, so

$$\mu^\circ = -RT \ln\left(\frac{8\pi^2}{C^\circ} Z\right) \tag{4}$$

$$Z = \int J(\mathbf{r})\, \mathrm{e}^{-\beta[U(\mathbf{r})+W(\mathbf{r})]}\, \mathrm{d}\mathbf{r} \tag{5}$$

where Z is a configuration integral over only the internal coordinates $\mathbf{r}$ and $J(\mathbf{r})$ is the Jacobian determinant for the coordinate transformation.[27,51,52]

It is worth noting that the separation of coordinates afforded by the harmonic oscillator−rigid rotor approximation does not meet the criterion that U and W are independent of the external coordinates because it yields external rotations that correspond to linear displacements of atoms. These alter the conformation and, hence, the energy of the molecule.[27,53] This problem can be avoided by a different approach in which three atoms bonded in sequence are used to establish a molecular frame of reference. The three coordinates of atom 1 then define the translational coordinates of the molecule, the direction of the bond joining atoms 1 and 2 defines two rotational coordinates, and the angle of atom 3 defines the third (body) rotation of the molecule.[27] Thus, atom 1 is the origin of coordinates; the bond joining atoms 1 and 2 represents the z axis; the plane defined by atoms 1, 2, and 3 is the $x-z$ plane; and the y axis is normal to this plane. All of the $3N - 6$ Cartesian coordinates of the molecule can then be referenced to this coordinate system to generate "anchored Cartesian" coordinates.[27] The Jacobian determinant for this coordinate transformation is $b_2^2 b_3 \sin\theta_3$, where b_2 is the bond length of atoms 1 and 2, b_3 is the bond length of atoms 2 and 3, and θ_3 is the bond angle between b_2 and b_3.[26,27] The $3N - 6$ coordinates can be, furthermore, converted from Cartesian to BAT coordinates, in which the position of each atom is defined by one bond length, one bond angle, and one torsion angle, defined relative to three other atoms.[28,51,54] The Jacobian determinant when this transformation is included becomes $b_2^2 \Pi_{i=3}^N b_i \sin\theta_i$, where b_i and θ_i are the bond lengths and bond angles defined by the BAT coordinates.

2.2. Harmonic Approximation. The harmonic approximation is introduced here as a basis for subsequent discussion of the quasiharmonic and M2 methods, below. It is most applicable when a single deep energy well i at location $\mathbf{r}_i$ dominates the energy landscape. The harmonic approximation is established by expressing the energy $E = U + W$ as a Taylor series expansion around the energy minimum at $\mathbf{r}_i$ and retaining terms only up to the second order:

$$E(\mathbf{r}) \approx E(\mathbf{r}_i) + \frac{1}{2}(\Delta\mathbf{r})^T \mathbf{H}(\mathbf{r}_i)\, \Delta\mathbf{r} \tag{6}$$

$$\Delta\mathbf{r} \equiv \mathbf{r} - \mathbf{r}_i \tag{7}$$

where $\mathbf{H}(\mathbf{r}_i)$ is the second derivative, or Hessian, matrix of E at $\mathbf{r}_i$ and the first derivative terms are absent because all first derivatives equal zero at a local minimum. With this approximation, the chemical potential can be written as

$$\mu^\circ_{harm} = -RT \ln\left(\frac{8\pi^2}{C^\circ} J_i \int \exp\{-\beta[E_i + {}^1/_2(\Delta\mathbf{r})^T \mathbf{H}(\mathbf{r}_i)\, \Delta\mathbf{r}]\}\, \mathrm{d}(\Delta\mathbf{r})\right) \tag{8}$$

$$= E_i - RT \ln\left(\frac{8\pi^2}{C^\circ} J_i\right) - \frac{RT}{2} \ln\left\{\left(\frac{2\pi}{\beta}\right)^{3N-6} [\det \mathbf{H}(\mathbf{r}_i)]^{-1}\right\} \tag{9}$$

where $E_i \equiv E(\mathbf{r}_i)$ and $J_i \equiv J(\mathbf{r}_i)$ and we have made the approximation that the Jacobian determinant does not vary significantly within the low-energy part of the energy well. This is reasonable because the Jacobian does not depend on torsion angles, and bond lengths and bond angles are relatively rigid.

2.3. Quasiharmonic Approximation. The harmonic energy expression in eq 6 implies a Gaussian probability distribution of conformations around the energy minimum at $\mathbf{r}_i$, with the covariance matrix $\mathbf{C}$:

$$\mathbf{C} = (\beta\mathbf{H})^{-1} \tag{10}$$

In the QH method, the covariance matrix $\mathbf{C}$ is computed from snapshots of a MD simulation, and eq 10 is inverted to provide an estimate of $\mathbf{H}$,[33,44,45,55−57] which is then substituted into eq 9 and related equations for other thermodynamic quantities, such as the entropy. When the molecular dynamics run used to compute $\mathbf{C}$ is restricted to a single energy well (e.g., refs 55−57), the value of E_i in eq 9 can be obtained by energy minimization within the energy well. When the MD simulation samples multiple energy wells, it is not clear what value to use for E_i. However, the MD simulation provides the Boltzmann average of the energy, $\langle E \rangle$, and this

may be set equal to $E_i + (3N - 6)/2\beta$ by applying the equipartition theorem to the positional degrees of freedom, as is appropriate, given the assumption of harmonicity. From eqs 9 and 10, then

$$\mu^\circ_{\mathrm{QH}} = \langle E \rangle - \frac{3N-6}{2\beta} - RT \ln\left(\frac{8\pi^2}{C^\circ} J_i\right) - \frac{RT}{2} \ln[(2\pi)^{3N-6} \det \mathbf{C}] \quad (11)$$

The sum of the second, third, and fourth terms on the right-hand side is identifiable as $-TS^\circ$, while the sum of the first and second terms can be viewed as an effective energy minimum associated with the QH calculation, although it is not expected to correspond to the actual energy of any particular conformation.

2.4. Predominant States Approximation: Summation over Energy Wells. For a molecular system with multiple energy wells that are significantly occupied at ambient temperature, the chemical potential may be estimated via a sum of contributions z_i from the M lowest energy wells:

$$\mu^\circ \approx RT \ln\left(\frac{8\pi^2}{C^\circ} \sum_{i=1}^{M} z_i\right) \quad (12)$$

$$z_i \equiv J_i \int_i \mathrm{e}^{-\beta E(\mathbf{r})}\, \mathrm{d}\mathbf{r} \quad (13)$$

where the subscript on the integral symbol implies an integration domain limited to energy well i. The M2 algorithm (Section 3.1) uses this approach, computing z_i with an extension of the harmonic approximation that accounts for anharmonicity in the shapes of the energy wells.

3. Methods

3.1. Second-Generation Mining Minima Method. The M2 method is based on the predominant states approximation and has two main steps: finding the low energy conformations $i = 1, ..., M$ and evaluating the configuration integral z_i within each energy well i. The first step is carried out with the Tork algorithm.[58] In Tork, the molecule is energy-minimized and the Hessian matrix of the energy with respect to BAT coordinates is computed and diagonalized. The molecule is iteratively distorted along the eigenvectors with lower force constants (eigenvalues) and energy-minimized again to arrive at new energy minima. The method is similar to low mode search;[59−61] perhaps the chief difference is the use of BAT coordinates, which lead to more physically reasonable molecular distortions. Repeat conformations are removed to prevent double-counting; as part of this process, a symmetry-detection algorithm[62] is used to detect topological symmetries[62,63] of the molecule and eliminate conformations that are the same after rotational interchange of chemically equivalent atoms. For example, two conformations that differ by a 120° rotation of a methyl group are considered identical.

The second step, evaluation of the configuration integral z_i in each energy well i, is carried out with the harmonic approximation/mode scanning (HA/MS) method, which uses the harmonic approximation along with mode scanning, a fast correction for local anharmonicity. The Hessian matrix of the energy with respect to BAT coordinates is computed at the energy minimum, and the Hessian is diagonalized. The harmonic approximation of the configuration integral is then computed along each eigenvector with an integration range of ±3 standard deviations of the corresponding Gaussian. For each eigenvector with a force constant less than 10 kcal/mol-A^2, the integral is also evaluated by mode scanning,[28] in which the molecule is distorted stepwise along the eigenvector, the energy is computed at each step, and the integral of the Boltzmann factor is computed numerically. The numerical result for a given eigenvector is substituted for the harmonic result when the numerical integral deviates from the harmonic by >1 kcal/mol, indicating substantial local anharmonicity along the mode. The configuration integral for energy well i becomes

$$z_i \approx J_i \prod_j^{N_{\mathrm{scan}}} S_j \prod_k^{N_{\mathrm{harm}}} \left[\sqrt{\frac{2\pi RT}{K_k}}\, \mathrm{ERF}\left(\frac{w_k}{\sqrt{\frac{2RT}{K_k}}}\right)\right] \quad (14)$$

where j ranges over the N_{scan} numerically integrated modes, S_j is the numerical integral for mode j, k ranges over the N_{harm} modes treated as harmonic, K_k is the eigenvalue of mode k, w_k is the integration range of mode k, and ERF is the error function. The chemical potential is then calculated via a sum of M local integrals z_i, according to eq 12. The mean energy $\langle E \rangle$ can also be computed with the M2 method as

$$\langle E \rangle = \frac{\sum_i^M J_i \int_i E(\mathbf{r})\, \mathrm{e}^{-E(\mathbf{r})/RT}\, \mathrm{d}\mathbf{r}}{\sum_i^M z_i} \quad (15)$$

and the entropy can be computed from the difference between μ° and $\langle E \rangle$.

3.2. Molecular Systems and Computational Methods. This study uses test systems of increasing complexity to evaluate and characterize the QH approximation. Section 4.1.1 considers a united-atom representation of butane in which hydrogens are merged with the carbons to which they are bonded and all nonbonded interactions are turned off, leaving only terms related to bond lengths, bond angles, and the solitary dihedral angle. This model is useful because the configurational integral can be computed to high accuracy by analytic or semianalytic methods. Section 4.1.2 considers linear alkanes in the all-atom representation with the full CHARMM22[64] vacuum potential from the program QUANTA,[65] including nonbonded interactions, but with no solvation model. Finally, Section 4.2 considers the binding affinity of a cyclic urea with a synthetic receptor in chloroform, a system for which experimental binding data are available. The all-atom CHARMM22 representation is used for this system, along with an implicit solvation model, which is implemented as follows. Initial coordinates for all molecules were generated with the program QUANTA[65] and then energy-minimized by the conjugate gradient method and, then, the Newton−Raphson method until the energy gradient was $<10^{-3}$ kcal/mol/Å.

As previously discussed,[22,23] the solvation energy is estimated with a generalized Born (GB) model[66] during the Tork, HA/MS, and MD calculations. In addition, the M2 calculations adjust the free energy of each energy well i toward the Poisson−Boltzmann/surface area (PB/SA) model by using the program UHBD[67] to solve the linearized PB equation and compute the surface area for the energy minimum conformation and replacing the GB contribution with these more detailed solvation terms. Thus, if the initial value of the chemical potential for energy well i is $\mu^{\circ}_{i,\mathrm{init}}$, then the corrected value is

$$\mu^{\circ}_{i,\mathrm{corr}} = \mu^{\circ}_{i,\mathrm{init}} - W_{i,\mathrm{GB}} + W_{i,\mathrm{PB}} + W_{i,\mathrm{SA}} \qquad (16)$$

where the W terms are solvation contributions computed at $\boldsymbol{r}_i$. The PB calculations use a chloroform dielectric constant of 4.8 and a chloroform radius of 2.4 Å.[68] The surface area term, $W_{i,\mathrm{SA}}$, is the work of forming a nonpolar cavity in the solvent with the shape of the solute in conformation i and also includes the van der Waals interactions between the solvent and solute. It is approximated by a linear function of the molecular surface area A_i: $W_{i,\mathrm{SA}} = aA_i + b$. For the chloroform solvent considered here, a and b are set to −0.04 kcal/(mol Å^2) and 1.41 kcal/mol, respectively.[69,70] As noted in Section 2.3, the MD calculation used in implementing the QH method often is not associated with a single, well-defined energy minimum. Therefore, to enable comparison with the M2 calculations, the solvation correction just described is applied on the basis of the energy-minimized conformation used to initiate the MD run.

The QH method is implemented here with the stochastic MD method at 300 K,[71] as implemented in the program UHBD.[67] This approach ensures that the systems are fully thermalized despite their small size. All simulations used a 1 fs time step, and the duration of each simulation is stated in the Results section. In every case, 2000 evenly spaced snapshots were saved and used for the QH analysis, except that 4000 snapshots were used in the case of the 6 ns convergence tests for the urea receptor. The conformations in the snapshots were aligned on the basis of the three arbitrarily selected atoms that define the molecular coordinate system. Both anchored Cartesian coordinates and BAT coordinates were computed, along with the corresponding Jacobian determinants, so that QH results in both coordinate systems could be computed and compared. The resulting coordinates were then used to obtain the covariance matrices for use in eq 11. Because methyl rotations were not considered to generate new conformations when counting energy minima, a ±60° constraint was applied to all methyl groups in the alkanes and the urea receptor during the MD simulations.

For some QH simulations, we wished to restrict the molecular dynamics to conformations in the neighborhood of a single energy well. This was accomplished by restraining all rotatable bonds with added energy terms of the form $E_{\mathrm{rest}} = 0.005[(\phi - \phi_0)/\Delta\phi]$,[14] where ϕ is the rotation to be restrained, ϕ_0 is the center of the restraint range, and $\Delta\phi$, the range of the restraint, was set to 60° for all rotatable bonds, except that a range of 90° was used for rotatable bonds attached to phenyl groups.

All calculations were carried out on Linux computers with Athlon 1.67 GHz processors. The parameters for Tork conformational search are as described previously.[58]

4. Results

This section compares various methods of computing chemical potentials, including the QH approximation in both Cartesian and BAT coordinates and the M2 method. All results are based upon integrals over $3N - 6$ internal degrees of freedom, defined with respect to the molecular frame of reference, and include the appropriate Jacobian determinants and the factor of $8\pi^2/C^{\circ}$, which results from integrating over external coordinates. The values of $\langle E \rangle$ reported for the QH calculations are mean potential energies obtained from the MD calculations.

4.1. Linear Alkanes. This section compares QH results for a simplified representation of butane for which analytical results are available and then compares QH results for more fully represented linear alkanes with the results of the M2 method, whose numerical accuracy has previously been validated.[28] We begin by establishing that the QH calculations have converged adequately for this series of systems. This is done by examining convergence for the most complicated one, heptane in an all-atom representation with no conformational restraints. Figure 1 graphs the entropic contribution to the chemical potential computed with the QH method as a function of the number of sequential snapshots included from simulations of 2 ns total duration, to a maximum of 2000 snapshots. Four of these convergence graphs are shown. Two use a BAT basis set and are based upon simulations starting from two different conformations of heptane, while the other two use an anchored Cartesian basis set and are based upon the same simulations used for the BAT results. The two BAT graphs plateau at values agreeing to within a few tenths of kilocalories per mole about $^3/_4$ of the way through the 2 ns simulation, while the anchored Cartesian result plateaus about halfway through the simulation. These results document adequate convergence for the most complex alkane studied here using 2000 snapshots from a 2 ns MD trajectory.

4.1.1. Simplified Butane. The simplest system considered here is united-atom butane with all nonbonded interactions artificially turned off. This system possesses only four particles and six energy terms (three bond stretches, two angle bends, and one torsion), and its configuration integral can be factorized and computed semianalytically to high precision. It possesses three equivalent energy wells because of the 3-fold periodicity of the dihedral energy component, and Table 1 presents results for a single energy well and for all three energy wells taken together. The QH results for a single well were obtained by adding a restraining function during the MD calculation, as described in the Methods section.

For a single energy well, chemical potentials computed by QH in BAT coordinates and by the M2 method agree with the semianalytic reference results to within 0.05 kcal/mol. However, QH in Cartesian coordinates yields a chemical potential that is about 0.5 kcal/mol lower than the reference result. For integration over all three energy wells, the M2 method still agrees with the reference results to within 0.03

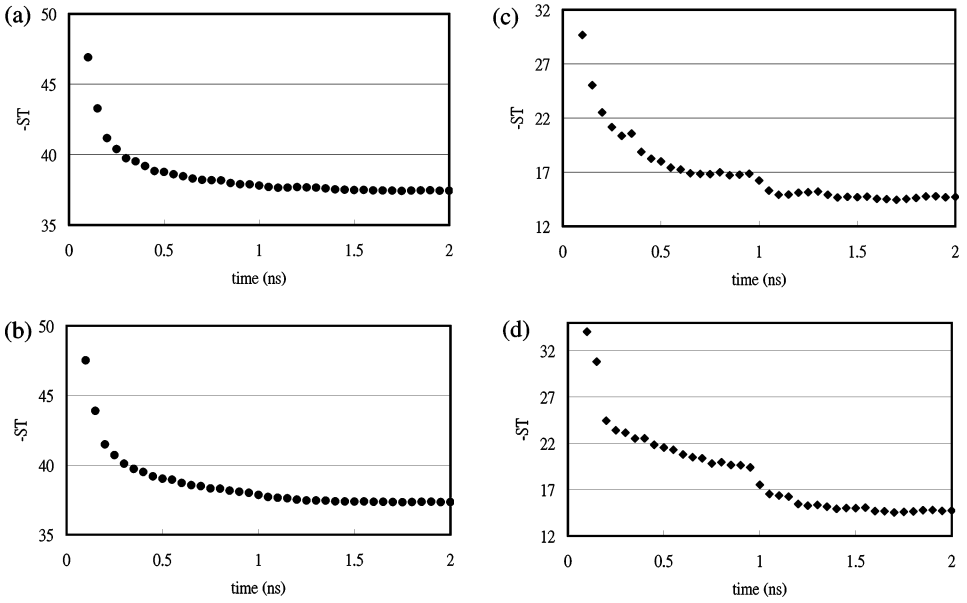

Figure 1. Convergence analysis of QH calculations, showing the entropy contribution to the chemical potential (kcal/mol) as a function of the duration of MD simulations of heptane, which sample all energy wells. Top (a, c) and bottom (b, d) graphs show results for simulations starting from the QUANTA-generated conformation and the least-stable local energy minimum found by a Tork conformational search. Left graphs (a, b): BAT coordinates. Right graphs (c, d): Cartesian coordinates.

Table 1. Computed Chemical Potential μ° and Average Potential Energy $\langle E \rangle$ (kcal/mol) of Simplified Butane Obtained Analytically, from the M2 Method, and from QH Calculations Based upon Simulations of 100 ps and 2 ns for One and Three Energy Wells, Respectively[a]

	analytic	M2		QH		
	μ°	$\langle E \rangle$	μ°	$\langle E \rangle$	μ°_{BAT}	μ°_{Cart}
single energy well	−1.71	1.74	−1.69	1.84	−1.76	−2.19
three energy wells	−2.37	1.74	−2.34	1.85	−2.94	−4.53

[a] μ°_{BAT} and μ°_{Cart}: QH results based upon BAT and Cartesian coordinates.

kcal/mol, but the QH calculations yield significantly larger errors, −0.57 kcal/mol and −2.16 kcal/mol, respectively, for BAT and Cartesian coordinates. Another way to look at these data is to recognize that the chemical potential associated with three equivalent energy wells must be $-RT \ln 3 = -0.66$ kcal/mol lower than that associated with only one of the wells. The semianalytic and M2 results yield this difference, but the QH calculations with BAT and Cartesian coordinates yield differences of −1.18 and −2.34 kcal/mol, respectively.

The cause of the error can be elucidated by additional calculations focusing exclusively on the variation of dihedral energy with butane's solitary dihedral degree of freedom. First, Figure 2 shows the exact dihedral potential for one of the three dihedral-energy energy wells (dots), along with the harmonic (bold solid) and the QH approximations (light solid) to the shape of this energy well. The approximations match the exact curve well, and the chemical potentials associated with the harmonic and QH approximations, 0.40 and 0.37 kcal/mol, respectively, agree well with the analytic result, 0.36 kcal/mol. Thus, the QH treatment accounts well for the small amount of anharmonicity in the single energy well. For comparison, Figure 3 shows the exact potential

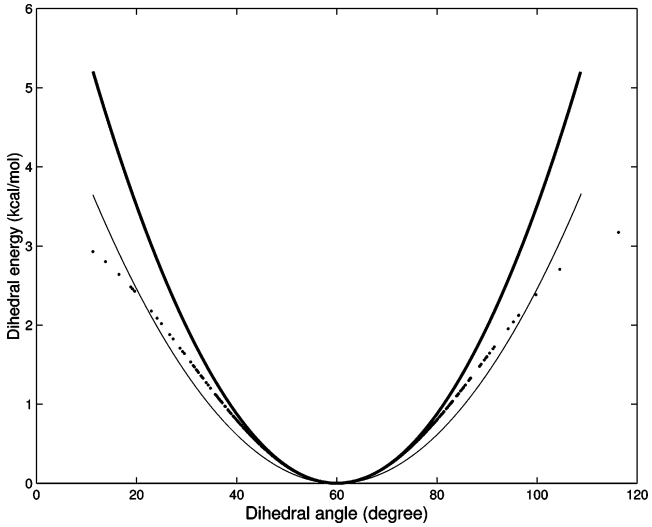

Figure 2. Dihedral energy of simplified butane (see Methods) as a function of the central dihedral angle, in a single energy well. Dots are results from MD trajectory snapshots during a 2 ns MD simulation. Bold line: harmonic approximation. Thin line: QH approximation based on the MD results (dots).

for all three dihedral energy wells (dots), along with the QH approximation (solid). It is evident that the QH approximation effectively merges the three relatively narrow wells into a single broad well. Not surprisingly, then, it yields a significantly lower chemical potential, −0.8 kcal/mol, than the analytic result, −0.29 kcal/mol. On the other hand, using the harmonic approximation to approximate the integral in each well and summing the three resulting integrals, as in the M2 method, yields an accurate chemical potential of −0.26 kcal/mol. Thus, the QH method generates significant errors because it merges multiple narrow energy wells into one broad energy well.

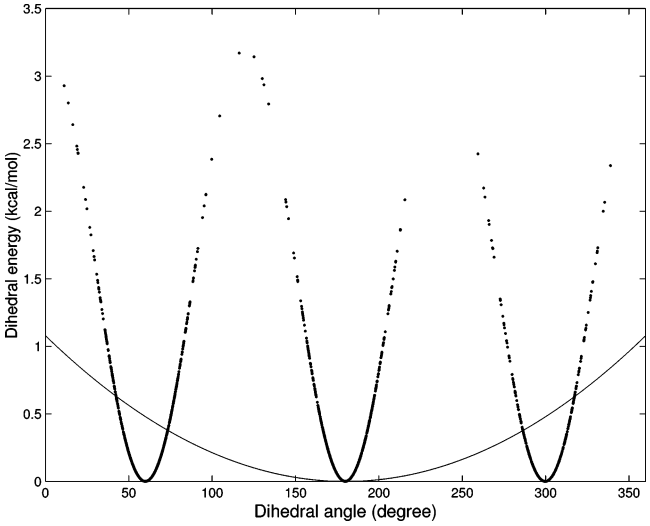

Figure 3. Dihedral energy of simplified butane across all three energy wells. Dots and thin line have the same meaning as in Figure *2*.

It is also of interest that the QH results are particularly inaccurate when they are based upon Cartesian coordinates. We conjecture that this is because the Cartesian basis set effectively magnifies the fluctuations associated with bond rotations. Rotating one bond causes an atom to move along a curved path, producing changes in at least two Cartesian degrees of freedom and, thus, generating excessive contributions to the covariance matrix and, hence, to the entropy obtained by the QH approximation. This error is especially large when large bond rotations occur; thus, going from BAT to Cartesian coordinates produces a much larger error when all three energy wells of simplified butane are considered than when only one is considered, as shown in Table 1.

4.1.2. All-Atom Alkanes. The results for simplified butane suggest that the QH approximation is particularly inaccurate for a system with multiple distinct energy wells. If so, then its accuracy should decrease for linear alkanes of increasing length, because they possess more energy wells. This expectation is evaluated here by applying the QH approximation to linear alkanes from butane through heptane. All hydrogens are treated explicitly and nonbonded interactions are included, to provide a more physically realistic test than simplified butane. Analytic reference results are not available for these cases, so reference results are obtained by the M2 method, on the basis of validation of its accuracy for simplified butane (above) and other systems.[28] Data are presented for unconstrained sampling over all energy wells and also for sampling only in the all-trans global energy minimum of each alkane.

The results for sampling over all energy wells, in Table 2a, confirm that the QH results become more inaccurate with increasing chain length. The errors result primarily from an excessively large entropy, consistent with overestimation of motional freedom by the QH method. This is also consistent with the fact that the Gaussian distribution yields the maximum entropy for a given variance.[46] As for simplified butane, the errors are most severe when Cartesian coordinates are used as the basis set. Thus, for heptane, the QH chemical potentials with BAT and Cartesian basis sets differ from the reference M2 results by −1.9 and −24.7 kcal/mol, respectively.

The QH method yields considerably more accurate results when it is used to sample over only the single all-trans energy well of each alkane, as shown in Table 2b. QH in BAT coordinates now agrees closely with the reference M2 calculations: chemical potentials computed with BAT and Cartesian coordinates are accurate to within 1.15 and 4.85 kcal/mol. It is worth remarking on the fact that going from multiple energy wells (Table 2a) to a single energy well (Table 2b) does not raise the entropy by $-RT \ln N$, where N is the number of energy wells. This is because the all-trans conformation studied here is the global energy minimum, so the other energy wells are not as highly populated and,

Table 2. Standard Chemical Potential μ°, Average Potential Energy $\langle E \rangle$, and Configurational Entropy $-TS^\circ_{\text{config}}$ (kcal/mol, 1 mol/L Standard Concentration) of All-Atom Linear Alkanes, Computed with the M2 Method and the QH Method Using 2 ns MD Simulations[a]

(a) All Energy Wells

		M2			QH				
						bond-angle-torsion		Cartesian	
	number of wells	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$-TS^\circ_{\text{config}}$	μ°
butane	3	10.95	20.81	31.55	11.33	19.86	31.19	15.13	26.46
pentane	6	13.42	27.48	40.90	14.09	25.82	39.82	14.38	27.47
hexane	15	15.92	34.04	49.96	16.90	31.53	48.43	14.51	31.41
heptane	52	18.53	40.72	58.95	19.58	37.43	57.01	14.71	34.27

(b) Single Energy Well

						bond-angle-torsion		Cartesian	
	number of wells	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$-TS^\circ_{\text{config}}$	μ°
butane	1	10.71	21.06	31.77	10.40	21.26	31.64	19.50	29.88
pentane	1	13.18	27.86	41.04	13.59	27.99	41.58	23.93	37.52
hexane	1	15.54	34.65	50.19	16.07	34.40	50.47	29.45	45.72
heptane	1	17.87	41.48	59.35	18.59	41.90	60.50	35.95	54.54

[a] (a) Results computed for all energy wells. (b) Results for the all-trans energy well only.

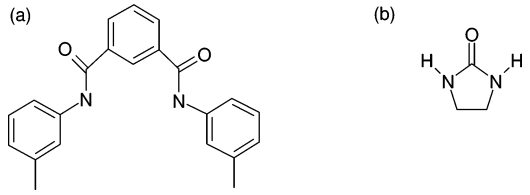

Figure 4. Host−guest system. (a) Receptor and (b) ethylenurea ligand.

Table 3. Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea from the M2 Method[a]

	initial energy	number of wells	$\langle E \rangle$	$-TS^{\circ}_{\mathrm{config}}$	μ°	$\mu^{\circ}_{\mathrm{corr}}$
ethylenurea	−12.1	1	−12.4	19.3	6.9	4.1
receptor	−38.5	15	−2.9	104.2	101.3	83.0
complex	−74.6	87	−28.0	134.3	106.3	83.7
change on binding			**−12.7**	**10.8**	**−1.9**	**−3.4**

[a] Last row shows changes in these quantities upon binding. Results are based upon the GB solvation model except that $\mu^{\circ}_{\mathrm{corr}}$ has been corrected toward the PB/SA solvent model, as detailed in the Methods section. See Table 1 for other symbols. Experimental binding free energy is −4.1 kcal/mol.

therefore, make smaller contributions to the configuration integral.

It is also of interest that the values of average energy $\langle E \rangle$ obtained by the M2 method agree with the QH values to within 1 kcal/mol. This agreement is striking because the methods are so different. The M2 method computes $\langle E \rangle$ with the harmonic approximation in multiple energy wells, correcting anharmonicity via mode scanning if needed, while the QH values are obtained directly from MD simulations. The agreement found here supports the validity of viewing the molecular configuration integral as a sum of contributions across distinct energy wells.

Note that the BAT results are insensitive to the choice of the three root atoms that define the molecular frame of reference and to the four-atom sets that define torsion angles in the BAT coordinate system. For example, changing to a different set of root atoms in heptane, and consequently also changing the torsion definitions, yields the same QH and M2 entropies to within 0.2 and 0.01 kcal/mol, respectively.

4.2. Urea Receptor. Because the QH method can provide standard chemical potentials, it can be used to compute the binding free energy of two molecules via the expression $\Delta G_{\mathrm{bind}} = \mu^{\circ}_{\mathrm{RL}} - \mu^{\circ}_{\mathrm{R}} - \mu^{\circ}_{\mathrm{L}}$, where R and L indicate a receptor and a its ligand, respectively. This section tests such an application for a simple host−guest system, a synthetic receptor and its urea guest (Figure 4) that interact primarily via hydrogen bonding and bind in chloroform with a standard free energy change of −4.1 kcal/mol.[72] As shown in Table 3, the M2 method yields a binding free energy $\Delta G^{\circ}_{\mathrm{corr}}$ of −3.4 kcal/mol, in good agreement with the experimental result, the error being less than 1 kcal/mol. However, the QH method yields binding free energies $\Delta G^{\circ}_{\mathrm{corr}}$ that are about 8 kcal/mol too strong, with either BAT or Cartesian coordinates, as shown in Table 4. For QH in Cartesian coordinates, this error would be even greater if not for partial cancellation of errors, since the chemical potentials from the QH calculations in Cartesian coordinates deviate from the M2 results by on the order of 30 kcal/mol. QH in BAT coordinates yields more accurate chemical potentials, with maximal errors of only about 9 kcal/mol

It is worth asking whether the inaccuracy of the QH method results from the somewhat arbitrary method by which the solvation correction is applied. As noted in the Methods section, there is no definite prescription for selecting the conformation(s) for which it should be computed, unlike for the M2 method; here, the QH correction is obtained for the first conformation of the MD calculation. It is, thus, of interest to consider the accuracy of the M2 and QH methods when the solvation correction is neglected; the only solvation model then is the GB model (see Methods). As shown in Tables 3 and 4, the uncorrected M2 results are still much closer to those of the experiment than the uncorrected QH results: ΔG° is −1.9 kcal/mol from M2 and −9 kcal/mol from QH with both BAT and Cartesian coordinates. These results suggest that the errors in the QH calculations result from inaccuracy in the configuration integral rather than the solvation model.

It is also important to assess the convergence of the QH results, which are based upon 2000 snapshots from a 2 ns MD simulation. Convergence was assessed by recomputing the QH results with a new MD calculation starting from the global energy minima of the receptor and the complex obtained from the M2 calculations. (The free ligand is rigid and, therefore, has only one highly occupied conformation.) The new QH calculation also overestimates the binding affinity, as shown in Table 5, but by a different amount. This difference indicates that the MD calculations are not adequately converged, even though they are rather lengthy and the molecular systems are simple. Examination of the two MD simulations indicates that both sample the same energy minima but spend different amounts of time in them.

Table 4. QH Calculations of Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea, Based upon 2 ns MD Simulations Started from QUANTA-Generated Conformations[a]

	initial energy	$\langle E \rangle$	bond-angle-torsion			Cartesian		
			$-TS^{\circ}_{\mathrm{config}}$	μ°	$\mu^{\circ}_{\mathrm{corr}}$	$-TS^{\circ}_{\mathrm{config}}$	μ°	$\mu^{\circ}_{\mathrm{corr}}$
ethylenurea	−21.1	−12.1	19.4	7.3	4.3	18.4	6.3	3.5
receptor	−38.5	−2.1	101.1	99.0	81.1	78.4	76.4	58.4
complex	−74.6	−26.5	123.9	97.4	74.3	99.9	73.4	50.3
change on binding		**−12.3**	**3.4**	**−8.9**	**−11.1**	**3.1**	**−9.3**	**−11.6**

[a] See Table 1 for symbols.

Table 5. QH Calculations of Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea Based upon 2 ns Simulations Started from Conformations of Lowest Potential Energy as Found by a Tork Conformational Search

			bond-angle-torsion			Cartesian		
	initial energy	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	μ°_{corr}	$-TS^\circ_{\text{config}}$	μ°	μ°_{corr}
ethylenurea	−21.1	−12.1	19.4	7.3	4.3	18.4	6.3	3.5
receptor	−39.4	−2.2	103.0	100.8	82.2	82.3	80.1	61.5
complex	−75.6	−26.5	122.5	96.0	73.5	97.1	70.7	48.1
change on binding		**−12.2**	**0.1**	**−12.1**	**−13.0**	**−3.6**	**−15.7**	**−16.9**

Table 6. QH Calculations for the Free Urea Receptor and the Bound Complex Using BAT Coordinates and Based upon 6 ns MD Simulations Started from QUANTA-Generated Conformations (Simulation 1) and from the Global Energy Minima (Simulation 2)[a]

	initial energy	$\langle E \rangle$	μ°
	Simulation 1		
receptor	−38.5	−2.0	97.3
complex	−74.6	−26.5	95.5
	Simulation 2		
receptor	−39.4	−2.1	97.5
complex	−75.6	−26.5	93.9

[a] Results are based on the GB solvation model only. See Table 4 for symbols.

We sought to improve convergence by extending the 2 ns simulations of the complex and the free receptor to 6 ns. QH analysis in BAT coordinates of these longer simulations still leaves a 1.6 kcal/mol discrepancy between the two MD runs for the complex, though the results for the free receptor now agree to within 0.2 kcal/mol. (See Table 6.) The binding free energies computed from these QH calculations are still inaccurate, at −6.1 and −7.9 kcal/mol for Simulations 1 and 2, respectively. Presumably, the free receptor results converge more readily because it is simpler than the complex. The 0.2 kcal/mol difference between the two 6 ns results for the free receptor is traceable to slightly different probabilities of the two main conformations of the receptor. In conformation A, the dihedral angle marked in Figure 5 is near 0, and in conformation B, this angle is near π radians. As illustrated in Figures 6 and 7, both MD runs find conformations A and B. However, simulation 1 assigns a 50% probability to conformation A, while simulation 2 assigns 35.7% probability to conformation A. The implications of this conformational transition are highlighted by the graphs of computed free energy versus time in Figure 6 and 7, which show sudden drops in free energy (top graph) when the receptor shifts from conformation A to conformation B (bottom graph).

It is also of interest to test a prior suggestion for achieving well-defined results from the QH method, that is, using short MD simulations in order to limit the number of transitions among different energy wells.[73] We tested this approach by repeating the QH calculations with 100-fold-shorter MD runs, 20 ps in length. As shown in Table 7, this method does not improve the agreement with the experiment results or with the M2 calculations. Similar results were obtained when

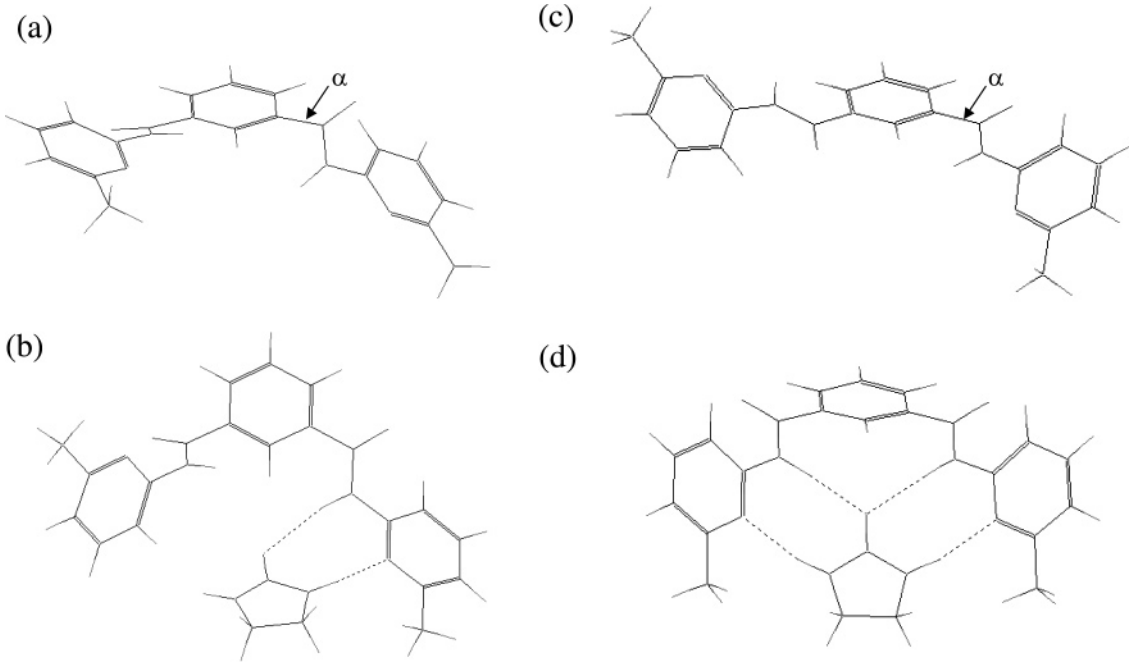

Figure 5. Conformations of free host and host−guest complex. Top (a, c): free receptor. Bottom (b, d): receptor−ligand complex. Left (a, b): QUANTA-generated conformations. Right (c, d): global energy minima obtained from a Tork conformational search. A key dihedral angle, α, is noted. (See Results and Figures 6 and 7.)

Table 7. QH Calculations of Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea Based upon Only the First 20 ps of the MD Simulations

			bond-angle-torsion			Cartesian		
	initial energy	$\langle E \rangle$	$-TS^\circ_{config}$	μ°	μ°_{corr}	$-TS^\circ_{config}$	μ°	μ°_{corr}
ethylenurea	−21.1	11.9	19.5	7.6	4.7	17.8	6.9	4.1
receptor	−38.5	−1.4	102.8	101.5	83.5	82.6	81.2	63.3
complex	−74.6	−26.4	126.2	99.8	76.7	102.4	76.0	52.9
change on binding		**−13.1**	**3.9**	**−9.3**	**−11.5**	**2.0**	**−12.1**	**−14.5**

another set of 20 ps simulations was started from the global energy minimum conformation of each species (results not shown). Inspection of the first trajectory shows that, although the free receptor does stay in one energy well during the short simulation, the more flexible complex still samples multiple energy minima during the 20 ps simulation. Thus, shortening the simulation to 20 ps does not accomplish the goal of eliminating conformational transitions. More importantly, failure to sample over multiple energy wells that are occupied at the ambient temperature is expected to produce its own errors. Thus, using artificially short simulations does not appear to be a helpful approach to improving the QH method.

It is encouraging that the change in mean energy $\langle E \rangle$ upon binding computed with the M2 method (−12.7 kcal/mol) agrees well with that from the 2 ns MD simulations used in the QH calculations (−12.3 and −12.2 kcal/mol, as shown in Tables 3−5). This result is consistent with the alkane test cases and further supports the validity of the equations and approximations used in the M2 method, which does not involve any MD calculations. It also indicates that the difference in the binding free energies from the QH and M2 methods results primarily from differences in the entropies they yield. In particular, the QH approximation overestimates entropies, especially for relatively flexible systems. Thus, QH with BAT coordinates and M2 yield very similar values of $-TS^\circ$ for the rigid ethylenurea ligand (19.3 and 19.4 kcal/mol, respectively), but the deviation is larger for the more flexible receptor (104.2 versus 101.1 or 103.0 kcal/mol) and larger still for the complex (134.3 versus 123.9 or 122.5 kcal/mol). As a consequence, the QH calculations markedly overestimate the binding affinity.

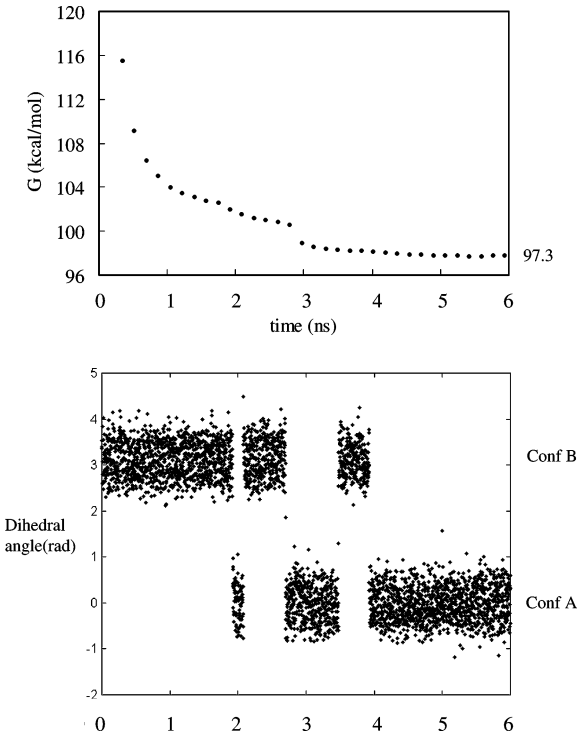

Figure 6. Analysis of QH calculations for the urea receptor, based upon simulation 1 (Table 6), which starts from the QUANTA-generated conformation. Top: convergence of μ°. Bottom: values of dihedral angle α (Figure 5) defining conformations A ($\alpha = 0$) and B ($\alpha = \pi$).

5. Discussion and Conclusions

This study has characterized the accuracy of the QH approximation by studying molecular systems that are small enough to allow comparison with accurate reference calculations, including a host−guest system for which experimental binding data are available. We find that the QH approximation can be accurate for a system with a single, highly occupied energy well, when the BAT coordinate system is

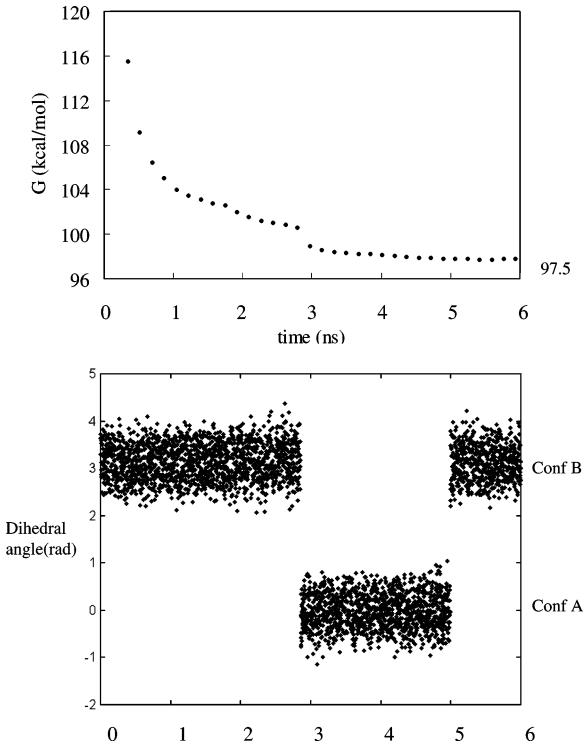

Figure 7. Same as Figure 6, but for simulation 2 (Table 6), which starts from the global energy minimum found by a Tork conformational search.

used as a basis set. Somewhat surprisingly, however, the use of a Cartesian basis set can introduce errors of several kilocalories per mole, even if the system is as simple as a linear alkane and attention is limited to a single energy well. Furthermore, for systems that explore multiple conformations (energy wells), the QH method can readily generate errors in the tens of kilocalories per mole. As in the case of a single energy well, the errors are much greater when a Cartesian basis set is used, rather than a BAT basis set. Although some cancellation of error occurs when a binding free energy is computed, the residual errors for the simple host−guest system studied here are still substantial. In contrast, the M2 method, which treats each energy well separately and uses BAT coordinates, agrees with experimental results to within 0.7 kcal/mol. It is expected that the QH method will also produce large errors for proteins, because proteins also access multiple distinct energy wells at room temperature.[34−36]

The chief reason for the inaccuracy of the QH method appears to be that it effectively merges multiple narrow energy wells and represents them, instead, as one wide energy well whose entropy is much greater than that associated with the individual energy wells (see Figure 3). This explanation is consistent with the fact that the entropy from the QH method is an upper limit to the actual entropy associated with the observed conformational variance.[46] That Cartesian coordinates yield particularly large errors, even for fluctuations within a single energy well (Table 2b), seems to result from the fact that fluctuation of a single torsional degree of freedom produces fluctuations of multiple Cartesian degrees of freedom and, hence, leads to higher computer entropies. This problem is least severe for small fluctuations, and the errors tend to cancel when taking free energy differences. However, there is no guarantee of adequate cancellation in any given calculation.

The quasiharmonic method also poses significant convergence challenges. Thus, two separate 2 ns MD calculations of the free and bound host−guest systems were found to yield binding free energies that differed by about 3 kcal/mol, and even 6 ns simulations are not fully converged for the receptor ligand complex. These are rather simple systems, so much longer calculations are presumably needed for a system as complicated as a protein. This expectation is consistent with Rinaldo and Field's QH results for the protein transferrin; Figure 15 in their paper shows that the computed entropy rises monotonically by 7 kcal/K mol during the course of a 1.5 ns simulation and is clearly still not converged at the end.[43] One reason convergence is difficult to achieve is that merely visiting all relevant energy wells is not enough; it is also necessary for the relative populations of the various wells to have converged fairly well. In contrast, converging the M2 method only requires visiting and integrating each occupied energy well a single time. Indeed, for the systems studied here, the M2 method can be run to convergence in about the time required to carry out a 2 ns MD calculation, and it yields more accurate and better converged results.

Overall, the present study indicates that the QH method can be accurate for a molecular system that remains in a single energy well but that it substantially overestimates the configurational entropy of systems that access multiple energy wells. The errors are particularly severe when the covariance matrix is computed in Cartesian coordinates. It is also found that the cancellation of errors does not suffice to provide accurate entropies for changes of state and that achieving good convergence can be surprisingly difficult even for a simple system, if it accesses multiple energy wells at ambient temperature.

Acknowledgment. This publication was made possible by Grant GM61300 from the National Institute of General Medical Sciences (NIGMS) of the NIH. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS. The authors thank Dr. Ioan Andricioaei for a helpful conversation.

References

(1) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281−283.

(2) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(3) Miyamoto, S.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226−245.

(4) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830−3832.

(5) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564−568.

(6) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−92.

(7) Lybrand, T. *Rev. Comput. Chem.* **1990**, *1*, 295−320.

(8) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.

(9) Warshel, A.; Tao, H.; Fothergill, M.; Chu, Z.-T. *Isr. J. Chem.* **1994**, *34*, 253−256.

(10) Marrone, T. J.; Briggs, J. M.; McCammon, J. A. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71−90.

(11) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430−437.

(12) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.

(13) Reddy, M. R.; Erion, M. D. *J. Enzyme Inhib.* **1998**, *14*, 1−14.

(14) Mitchell, M. J.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 271−275.

(15) Balbes, L. M.; Mascarella, S. W.; Boyd, D. B. *Rev. Comput. Chem.* **1994**, *5*, 337−379.

(16) Aqvist, J.; Medina, C.; Samuelsson, E. J. *Protein Eng.* **1994**, *7*, 385−391.

(17) Wang, W.; Wang, J.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 395−402.

(18) Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 10388−10397.

(19) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(20) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. *Biopolymers* **2003**, *68*, 16−34.

(21) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891−913.

(22) Chang, C.-E.; Gilson, M. K. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.

(23) Chen, W.; Chang, C.-E.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035−3049.

(24) Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.

(25) Go, N.; Scheraga, H. *J. Chem. Phys.* **1969**, *51*, 4751−4767.

(26) Go, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535−542.

(27) Potter, M. J.; Gilson, M. K. *J. Phys. Chem. A* **2002**, *126*, 563−566.

(28) Chang, C.-E.; Potter, M. J.; Gilson, M. K. *J. Phys. Chem. B* **2003**, *107*, 1048−1055.

(29) Eckart, C. *Phys. Rev.* **1935**, *47*, 552−558.

(30) Wilson, E. B., Jr.; Howard, J. B. *J. Chem. Phys.* **1936**, *4*, 260−268.

(31) Sayvetz, A. *J. Chem. Phys.* **1939**, *7*, 383−389.

(32) Kolossvary, I. *J. Phys. Chem. A* **1997**, *101*, 9900−9905.

(33) Rojas, O. L.; Levy, R. M.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 1037−1043.

(34) Ansari, A.; Berendsen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E. T.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5000−5004.

(35) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318−321.

(36) Kitao, A.; Hayward, S.; Go, N. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 496−517.

(37) Darian, E.; Hnizdo, V.; Fedorowicz, A.; Singh, H.; Demchuk, E. *J. Comput. Chem.* **2005**, *26*, 651−660.

(38) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617−621.

(39) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809−7817.

(40) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289−6292.

(41) Hsu, S.-T. D.; Peter, C.; van Gunsteren, W. F.; Bonvin, A. M. J. J. *Biophys. J.* **2005**, *88*, 15−24.

(42) Jusuf, S.; Loll, P. J.; Axelsen, P. H. *J. Am. Chem. Soc.* **2002**, *124*, 3490−3491.

(43) Rinaldo, D.; Field, M. J. *Biophys. J.* **2003**, *85*, 3485−3501.

(44) Luo, H.; Sharp, K. *PNAS* **2002**, *99*, 10399−10404.

(45) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.

(46) Shannon, C. E.; Weaver, W. *University of Illinois Press* 1949.

(47) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620−630.

(48) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.

(49) Mihailescu, M.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 23−26.

(50) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986; Chapter 15.

(51) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652−1661.

(52) Boresch, S.; Karplus, M. *J. Chem. Phys.* **2001**, *105*, 5145−5154.

(53) Kolossvary, I.; McMartin, C. *J. Math. Chem.* **1992**, *9*, 359−367.

(54) Pitzer, K. S. *J. Chem. Phys.* **1946**, *14*, 239−243.

(55) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325−332.

(56) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370−1374.

(57) Levy, R. M.; Rojas, O. L.; Friesner, R. A. *J. Phys. Chem.* **1984**, *88*, 4233−4238.

(58) Chang, C.-E.; Gilson, M. K. *J. Comput. Chem.* **2003**, *24*, 1987−1998.

(59) Kolossvary, I.; Guida, W. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019.

(60) Kolossvary, I.; Guida, W. *J. Comput. Chem.* **1999**, *20*, 1671−1684.

(61) Keserû, G. M.; Kolossvary, I. *J. Am. Chem. Soc.* **2001**, *123*, 12708−12709.

(62) Chen, W.; Huang, J.; Gilson, M. K. *J. Chem. Inf. Comput. Sci.* **2004**, *44*(4), 1301−1313.

(63) Ivanov, J.; Schûûrmann, G. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 728−737.

(64) *CHARMm*, version 22; Molecular Simulations Inc.: Waltham, MA, 1992.

(65) *QUANTA*; Accelrys, Inc.: San Diego, CA.

(66) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(67) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187−197.

(68) Luo, R.; Head, M. S.; Given, J. A.; Gilson, M. K. *Biophys. Chem.* **1999**, *78*, 183−193.

(69) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.

(70) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101−1109.

(71) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173−185.

(72) Goswami, S.; Mukherjee, R. *Tetrahedron Lett.* **1997**, *38*, 1619−1622.

(73) Nola, A. D.; Berendsen, H. J. C.; Edholm, O. *Macromolecules* **1984**, *17*, 2044−2050.

CT0500904