Calculation of Molecular Configuration Integrals

Chia-En Chang, Michael J. Potter,[†] and Michael K. Gilson*

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received: October 7, 2002; In Final Form: November 25, 2002

A method is presented for calculating the conformational free energy of a molecule in all degrees of freedom. The method uses the harmonic approximation with finite integration ranges, along with Mode Scanning, a fast correction for anharmonicity based upon internal bond-angle-torsion coordinates. Mode Scanning accounts for local anharmonicity without the need for expensive Monte Carlo integration. The method is efficient, and comparisons with analytic or highly detailed numerical calculations show excellent accuracy. Similar comparisons for the previously described Mode Integration method show that, although it is computationally demanding, it can be less accurate than the pure harmonic approximation. The inaccuracy of Mode Integration is traceable to its use of a Cartesian coordinate basis set; much more accurate results are obtained when the basis set consists of bond-angle-torsion coordinates.

1. Introduction

Identifying low-energy molecular conformations and evaluating their relative stabilities are essential steps in many applications in computational chemistry, such as computer-aided drug discovery and the design of self-assembling systems. Early methods of conformational analysis focused on the use of gradient-based energy minimization algorithms. However, this approach was ineffective for flexible molecules because gradient-based energy minimizers are poorly suited for sampling a broad range of different conformations. In addition, adequate solvation models were unavailable when these methods were first employed. The introduction of molecular dynamics and Monte Carlo simulations permitted the use of explicit models of the solvent and these methods can, with time, sample a range of conformations, but they are still quite time-consuming, especially when the solute can adopt conformations that are separated by high-energy barriers. Another class of conformational analysis methods uses global optimization algorithms to identify the lowest energy conformations and then evaluates their stabilities by evaluating the configuration integral within each energy well.[1] This "predominant states" approach is incompatible with an explicit representation of solvent molecules because the solvent can adopt innumerable low-energy conformations. However, with the development of fast, accurate implicit solvation models based upon continuum electrostatics (see, e.g., refs 2−5), predominant states methods have become practical and attractive.

Several predominant states methods have been described. Lipkowitz et al. studied host−guest systems by summing the Boltzmann factors associated with a single conformation at each energy minimum, implicitly assuming that the breadths of the various wells are uniform.[6] Wang et al. treated each energy well in more detail via the harmonic approximation.[7] The Mining minima algorithm computes the configuration integral of each energy well via detailed sampling and uses exclusion zones to avoid double-counting of energy minima.[8] This method, combined with modern implicit solvent models, has yielded promising results for conformational analysis[9] and binding.[10,11] However, its applicability is limited because it samples over only freely rotatable bonds, and thus is not suitable for compounds with flexible rings or anchored loops. In addition, it is of concern that useful information might be lost by neglecting "hard" degrees of freedom, such as angle-bends and bond-stretches, even for molecules without flexible rings.

The Mode Integration (MINTA) algorithm,[12,13] seeks to avoid these limitations, evaluating the configuration integral over all degrees of freedom. The calculations are made tractable by diagonalizing the matrix of second derivatives of the energy and using the eigenvectors to guide numerical sampling within each energy well. However, studies of MINTA to date have not fully validated the methodology. Indeed, a recent paper has shown that, by limiting attention to vibrational degrees of freedom, MINTA neglects large free energy contributions associated with rotational moments of inertia and also, in the case of binary association, with translational momenta.[14] There are also questions concerning the numerical accuracy of the vibrational integrals themselves: although the results of MINTA agreed with those of the independent jumping between wells (JBW) method,[12] it was not established that the two calculations ranged over the same set of energy wells. Most importantly, however, MINTA was not tested against analytically solvable cases. A specific concern is that MINTA may be inaccurate for highly flexible molecules. Even the soft eigenvectors found by MINTA are linear combinations of Cartesian coordinates, and there is no way that a linear combination of Cartesian coordinates can trace the circular motion of an atom associated with a bond rotation. Indeed, as diagrammed in Figure 1, a bond-rotation built as a linear combination of Cartesian coordinates will necessarily stretch bonds, and thus will stray from the low-energy basin of the energy well. Finally, although MINTA should in principle be more accurate than the plain harmonic approximation, it has not been compared with the harmonic approximation in any detail.

* Corresponding author. Voice: (301) 738-6217. Fax: (301) 738-6255. E-mail: gilson@umbi.umd.edu.

[†] Current address: VeraChem LLC, 20010 Century Blvd, Suite 102, Germantown, MD, 20874.

10.1021/jp027149c CCC: $25.00
Published on Web 01/07/2003

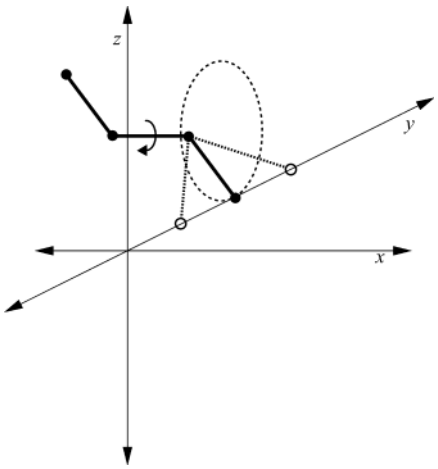

Figure 1. Consequences of describing a bond torsion as a linear combination of Cartesian components. Solid lines and filled circles show the equilibrium conformation of a molecule. Rotation of the indicated torsion turns the right-hand atom along the oval path. A Cartesian representation of this rotation (open circles) proves to be a linear displacement of the atom along the *y*-axis, which stretches the bond and sharply raises the energy.

The present study addresses these and related issues through tests on simple, analytically solvable cases as well as more interesting molecules, using a local implementation of MINTA. In addition, two other integration methods are introduced. The first is a MINTA-like harmonically biased sampling technique based upon internal bond-angle-torsion molecular coordinates; it is conjectured that this coordinate system will yield greater accuracy because its eigenvectors will allow essentially pure bond rotations with minimal coupling to bond lengths and angles. The second method is the harmonic approximation, modified by incorporation of finite integration ranges, along with detection of and adjustment for anharmonicities in specific modes ("Mode Scanning"). Note that the calculations in this paper do not aim to reproduce experimental data, but to evaluate the numerical characteristics of the methods when applied to a single energy well of a flexible molecule. Future publications will address the identification of the low-energy conformations to be integrated and will provide comparisons with experiment.

2. Methods

2.1. Theory. The methods described here aim ultimately to compute the standard chemical potential of a flexible molecule in solution. The standard chemical potential of species X can be written as[15−17]

$$\mu_X^\circ = -RT \ln\left(\frac{1}{V_{N,X}C^\circ}\frac{Q_{N,X}(V_{N,X})}{Q_{N,0}(V_{N,0})}\right) \quad (1)$$

Here $Q_{N,X}(V_{N,X})$ is the canonical partition function for a system at volume $V_{N,X}$ containing a large number, N, of solvent molecules and one solute molecule X. C° is the standard concentration, typically 1 mol/L, and $V_{N,X}$ is adjusted to establish the standard pressure of 1 atm. Similarly, $Q_{N,0}(V_{N,0})$ is the canonical partition function for the N solvent molecules without the solute, now at a slightly different equilibrium volume $V_{N,0}$ that also corresponds to standard pressure. (A pressure−volume term that is usually negligible for aqueous systems has been omitted.[17]) In evaluating the chemical potential, it is convenient to separate external (translational and rotational) from internal (conformational) coordinates. The integrals over external degrees of freedom are then carried out analytically, leaving the difficult integral over the internal degrees of freedom to be determined by numerical methods. Predominant states methods estimate the configuration integral (see below) as a sum of contributions from multiple energy wells. For simplicity, the following derivation assumes only one energy well is relevant, but generalization to multiple wells is straightforward.

2.1.1. Rigid Rotor Approximation. As previously discussed,[14] MINTA separates external from internal coordinates via the rigid-rotor approximation; i.e., via the Eckhardt−Sayvetz conditions.[18−21] This results in the following expression for the standard chemical potential μ_X° of species X in solution:[14]

$$e^{-\beta\mu_X^\circ} = \left[\frac{1}{C^\circ}\left(\frac{2\pi M_X kT}{h^2}\right)^{3/2}\right]\left[8\pi^2\left(\frac{2\pi kT}{h^2}\right)^{3/2}(I_{X,a}I_{X,b}I_{X,c})^{1/2}\right]\times \left[\left(\frac{2\pi kT}{h^2}\right)^{(3n_X-6)/2} Z_X^{vib}\right] \quad (2)$$

where M_X is the total mass of X, $I_{X,a}$, $I_{X,b}$, and $I_{X,c}$ are its principle moments of rotational inertia in the selected conformation, n_X is the number of atoms in the molecule, and h and k are Planck's and Boltzmann's constants. The bracketed terms in eq 2 can be interpreted, respectively, as translational, rotational, and vibrational contributions. The vibrational term includes the configuration integral in vibrational coordinates, Z^{vib}, which is given by

$$Z_X^{vib} \equiv \int e^{-\beta(U(\mathbf{q})+W(\mathbf{q}))}\, d\mathbf{q} \quad (3)$$

where U and W represent the potential and solvation energies of the solute as a function of its vibrational coordinates $\mathbf{q}$, which are linear combinations (eigenvectors) of Cartesian atomic coordinates, each of which is associated with a vibrational frequency. The range of integration for a given energy well is discussed below.

2.1.2. Anchored Cartesian Coordinates. The rigid rotor approximation is formally correct only in the limit where the vibrational motions are infinitesimal. For classical statistical thermodynamics, this restriction can be lifted by defining external coordinates that do not affect the conformation and hence the potential energy of the molecule.[22] This is accomplished by setting up a molecular frame of reference that moves and rotates with the molecule and anchoring the internal coordinates of the molecule with respect to this frame, as previously described.[14,17,22] The conformation of the molecule is then specified via $3n_X - 6$ internal Cartesian coordinates $\boldsymbol{r}_{int}$ defined with respect to the molecular frame. Integrating over the external coordinates is trivial because they do not affect the energy (assuming no external field). The resulting expression for the standard chemical potential of species X is

$$e^{-\beta\mu_X^\circ} = \frac{1}{h^{3n_X}}\frac{8\pi^2}{C^\circ} Z_X^{anc}\prod_{i=1}^{n_X}(2\pi kTm_i)^{3/2}$$

$$Z_X^{anc} \equiv \int b_2^2 b_3 \sin\theta_3\, e^{-\beta[U(\mathbf{r}_{int})+W(\mathbf{r}_{int})]}\, d\mathbf{r}_{int} \quad (4)$$

where m_i is the mass of atom i, the factor of $8\pi^2$ results from integration over external rotations, Z_X^{anc} is the internal configuration integral of molecule X in anchored internal coordinates $\boldsymbol{r}_{int}$, and b_2, b_3, and θ_3 are the bond lengths and angle associated with the three linked atoms used to define the molecular coordinate system.[14] (See Figure 2.) The factor of $b_2^2 b_3 \sin\theta_3$ is the Jacobian determinant for the coordinate transformation involved in setting up the external rotations. It depends on the

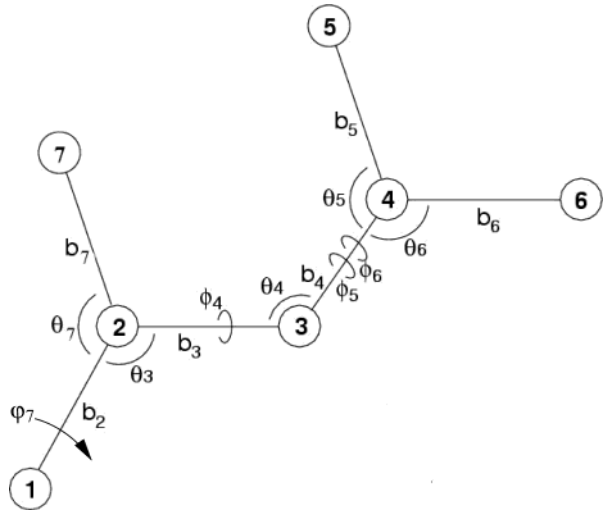

Figure 2. $3n - 6$ internal bond-angle-torsion coordinates for a small molecule.

internal coordinates $\boldsymbol{r}_{\text{int}}$ and hence belongs within the integral but, because the integrand is sharply peaked at the equilibrium values of these bond lengths and angles, this factor can be approximated as constant and removed from the integral. Note that, for a bimolecular complex XY, the internal frame defined for molecule X is also used for molecule Y.

2.1.3. Bond-Angle-Torsion Coordinates. The configuration integral in eq 4 can be rewritten as an integral over bond lengths, bond angles, and torsional rotations (BAT coordinates).[22−24] In BAT coordinates, the position of each atom $i > 3$ that is not bound to atom 2 is specified by its bond length (b_i), bond angle (θ_i), and dihedral angle (ϕ_i) with respect to three other atoms that are bonded in sequence and whose positions are already defined. (For atoms $i \leq 3$, one or more coordinates are identified as external.) It is worth pointing out that multiple atoms' torsion angles (ϕ) can be defined with respect to the same set of three other atoms. For example, the BAT coordinates of atom 5 and atom 6 in Figure 2 are (b_5, θ_5, ϕ_5) and (b_6, θ_6, ϕ_6), where ϕ_5 and ϕ_6 are both defined with respect to atoms 2−4. Clearly, however, the torsion angles of these two atoms will be highly correlated with each other because they will tend to rotate as a unit. The dihedral angles of atoms bonded to atom 2 are treated specially because atom 2 is used to define the molecular frame. For example, although the dihedral angle of atom 7 in Figure 2 could be defined relative to atoms 4, 3, and 2, this would produce an unnecessarily complicated set of coordinate dependencies in which torsional motions of atoms 7 and 4 would be strongly correlated with each other. We avoid this problem by instead defining the out-of-plane rotation of atom 7 in terms of an improper dihedral angle φ_i between planes 1−2−7 and 1−2−3. When the appropriate Jacobian determinant is included,[22−24] the anchored configuration integral $Z_{\text{X}}^{\text{anc}}$ (eq 4) becomes

$$Z_{\text{X}}^{\text{anc}} = \int b_2^{\,2} \prod_{i=3}^{n_{\text{X}}} (b_i^{\,2} \sin\theta_i) \mathrm{e}^{-\beta[U(\vec{b},\vec{\theta},\vec{\phi})+W(\vec{b},\vec{\theta},\vec{\phi})]}\, \mathrm{d}\vec{b}\, \mathrm{d}\vec{\theta}\, \mathrm{d}\vec{\phi} \quad (5)$$

Again, the Jacobian determinant (the preexponential factor) can be approximated as independent of conformation and removed from the integrand, if so desired.

2.2. Numerical Methods for Evaluating Configuration Integrals. *2.2.1. Harmonic Approximation and Mode Scanning.* The harmonic approximation can be used to estimate the internal configuration integral with the rigid rotor approximation, anchored Cartesian coordinates, or BAT coordinates, via the second derivative matrix of the energy ($E \equiv U + W$) computed at the base of the energy well of interest. Assuming a fixed Jacobian determinant (see above), the integrand then becomes a multidimensional Gaussian and the integral can be evaluated in terms of the inverse square root of the determinant of the second-derivative matrix.[24] Thus, there is no need to diagonalize the matrix to implement the pure harmonic approximation.

However, diagonalization can yield useful information regarding natural motions, soft modes consisting of correlated motions of multiple degrees of freedom. This information is useful for defining ranges of integration for the energy well (see the following paragraph). It also can be used to guide a search for anharmonicities by gradually distorting the molecule along each mode and evaluating the energy as a function of the distortion. We have found that, for the varied molecules included in the present study, the actual energy scans obtained by this procedure are close to harmonic for most modes, but that an occasional mode is markedly anharmonic. The typical pattern of anharmonicity is a flat-bottomed well in which the eigenvalue (force constant) from the harmonic analysis is very low, suggesting a very wide energy well along the corresponding eigenvector, but the actual energy suddenly rises sharply after an initial flat region. (See, for example, Figure 9.) This phenomenon was previously noted by Kolossvary.[12] The present paper introduces Mode Scanning, in which modes with low force constants are scanned and, for each mode j whose scan reveals significant anharmonicity, the harmonic approximation is replaced by a simple 1-dimensional numerical integral S_j evaluated with the composite trapezoid rule. Thus, the configuration integral is computed partly harmonically and partly via 1-dimensional numerical integrals:

$$Z_{\text{X}}^{\text{anc}} \approx b_2^{\,2}\, b_3 \sin\theta_3 \prod_j^{N_{\text{scan}}} S_j \prod_j^{N_{\text{scan}}} \left[\sqrt{\frac{2\pi kT}{K_i}}\, \mathrm{erf}\left(\frac{w_i}{\sqrt{\frac{2kT}{K_i}}} \right) \right] \quad (6)$$

where the index i ranges over all N_{harm} modes treated as harmonic, j ranges over the N_{scan} numerically integrated modes, K_i is the eigenvalue of mode i, w_i is the integration range of mode i, and erf is the error function. In the present method, modes with harmonic force constants less than 10 (kcal/mol)/Å^2 are scanned for anharmonicity, and the numerical integral is substituted for the harmonic approximation when the numerical integral deviates from the harmonic approximation more than 1 kcal/mol.

A frequently overlooked concern regarding use of the harmonic approximation to estimate the configuration integral is that the common practice of integrating a mode over $[-\infty, \infty]$ can lead to double-counting of regions of conformational space when the integrands of neighboring energy wells both deviate significantly from zero for some range of conformations, as illustrated in Figure 3. In particular, when a normal mode in BAT coordinates is a nearly pure bond rotation, the Gaussian integrand can remain large for multiple rotations around the bond. This "bond spinning" problem can lead to significant numerical errors, as reported in the Results. Finally, because MINTA-like sampling methods cannot in practice integrate over $[-\infty, \infty]$, use of unbounded harmonic integrals would make it impossible to compare the two methods in detail. These problems are avoided here by imposing a finite integration range on the harmonic approximation. In Cartesian coordinates, the Gaussian integrands associated with the harmonic ap-

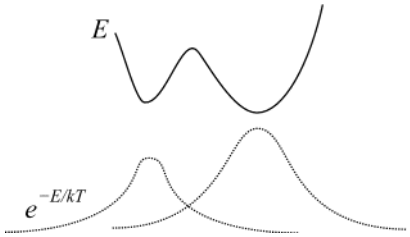

Figure 3. Overlap of configuration integrals for neighboring energy wells with the harmonic approximation. Solid line: energy. Dotted lines: Gaussian Boltzmann factors associated with the harmonic approximations in the two energy wells. Note the overlap between the two Gaussians.

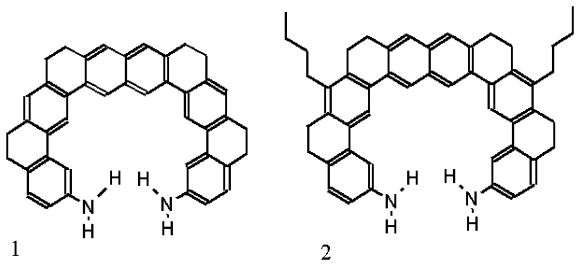

Figure 4. Truncated urea receptor (1) and urea receptor (2).

proximation are integrated to the three-standard deviation (3σ) point, as previously done.[12] However, the 3σ limit still leads to spinning of rotatable bonds in BAT coordinates. This problem is addressed here by limiting the integration ranges in BAT coordinates to the lesser of 3σ and $\pm 60°$ in any torsional coordinate. The $\pm 60°$ limit is not applied in Cartesian coordinates where modes never correspond to pure bond rotations, for reasons discussed in Results and diagrammed in Figure 1. To provide meaningful comparisons, the same limits are used for the harmonic approximation, Mode Scanning, and harmonic-biased Monte Carlo integration in BAT coordinates.

The validity of these limits was tested for the truncated urea receptor, compound **1** in Figure 4. A complete set of 130 energy minima was generated, and one thousand random, uniformly distributed conformations were generated within the integration range—the lesser of 3σ and $\pm 60°$ in BAT coordinates—of each energy well. It was found that none of these trial conformations lay in the integration range of any other energy well, indicating that double-counting is not a significant concern with these integration ranges. Similar results were obtained for several other test compounds not shown here.

2.2.2. Harmonic-Biased Sampling. As shown in the Results, Mode Scanning can increase accuracy relative to the pure harmonic approximation. However, Mode Scanning still ignores higher order correlations among degrees of freedom. For example, it is possible that additional anharmonicities could be detected by scanning along linear combinations of the low-eigenvalue eigenvectors. MINTA addresses this possibility by detailed sampling of the energy surface around the energy minimum, using the harmonic approximation to focus sampling into the low-energy neighborhood of the energy minimum. We have implemented MINTA as previously described,[12] but in anchored Cartesian internal coordinates rather than vibrational coordinates. Because both methods are Cartesian-based, they should have the same strengths and weaknesses. In particular, we expect that correlated motions built as linear combinations of Cartesian coordinates will not accurately track the low-energy region of an energy well because bond rotations projected onto Cartesian coordinates will stretch bonds, as discussed above. This concern is addressed here by implementation of harmonic-biased sampling in BAT coordinates, which can accurately capture pure bond rotations. The second derivative matrix of the energy with respect to BAT coordinates, which is required for this approach, is computed as

$$\mathbf{H}_{\mathrm{BAT}} = \mathbf{C}^{-1}\cdot\mathbf{H}_{\mathrm{AC}}\cdot(\mathbf{C}^{-1})^{\mathrm{T}} \tag{7}$$

Here, $\mathbf{C}$ is a matrix with elements $C_{ij} = \partial x_{\mathrm{BAT},j}/\partial x_{\mathrm{AC},i}$, where $x_{\mathrm{BAT},j}$ and $x_{\mathrm{AC},i}$ are coordinates j and i in the BAT and anchored Cartesian representations, respectively; and $\mathbf{H_{AC}}$ is the Hessian matrix in anchored Cartesian coordinates. Note that C_{ij}, $\mathbf{H}_{\mathrm{AC}}$, and $\mathbf{H}_{\mathrm{BAT}}$ are of dimension $3n_x - 6$ because the external coordinates have been separated out, as described in section 2.1.2.[14] The sampling ranges used for harmonic-biased sampling are the same as the integration limits used for the harmonic approximation and Mode Scanning; see previous section. With harmonic-biased sampling, the configuration integral is computed with conformational samples near the energy minimum that are distributed with a Gaussian probability distribution computed on the basis of the harmonic approximation. Thus,

$$Z_{\mathrm{X}}^{\mathrm{anc}} \approx \frac{J_0}{N_{\mathrm{smp}}}\sum_{j=1}^{N_{\mathrm{smp}}}\frac{\mathrm{e}^{-\beta[U(\mathbf{x}_j)+W(\mathbf{x}_j)]}}{P(\mathbf{x_j})}$$

$$P(\mathbf{x}) \equiv \sum_{i=1}^{3n_x-6}\frac{\mathrm{e}^{-(K_i/2kT)x_i^2}}{\int_V \mathrm{e}^{-(K_i/2kT)x_i^2}\,\mathrm{d}x_i} \tag{8}$$

where J_0 is the Jacobian determinant for the energy minimum, $\mathbf{x_j}$ defines conformation j in terms of the normal coordinates (x_1, x_2, ..., $x_{3\mathrm{nx}\,-\,6}$); the $j = 1, 2, ..., N_{\mathrm{smp}}$ conformations are sampled with the Gaussian probability distribution $P(\mathbf{x})$; and K_i is the eigenvalue of mode i.

2.3. Computational Details. Initial coordinates for all molecules were generated with the program QUANTA;[25] linear alkanes were built in the all-trans conformation. Potential energy was computed with the CHARMM 22 parameter set. No detailed solvent model was used, but a distance dependent dielectric constant ($\epsilon_{ij} = 4r_{ij}$) or a fixed dielectric constant $\epsilon = 40$, were used to avoid grossly unrealistic in vacuo Coulombic interactions. A conformational search method which will be described in a separate paper was used to generate different minima for compound 1 in Figure 4 in order to test for overlapping energy minima (section 2.2.1) and for the cyclophane in Figure 10. (Note that any other conformational search method that generates low-energy local minima could have been used instead.) Each molecular system was energy-minimized by the conjugate gradient and then the Newton−Raphson method until the energy gradient was $<10^{-5}$ (kcal/mol)/Å. All calculations were done on a PC with a Pentium III 733 MHz processor.

3. Results

This section evaluates the following methods for computing the molecular configuration integral in a single energy well for a series of molecules: harmonic-biased Monte Carlo integration (HB) in bond-angle-torsion (BAT) coordinates; HB in anchored Cartesian coordinates (MINTA); the pure harmonic approximation (HA) in both BAT and Cartesian coordinates; and harmonic approximation with Mode Scanning (HA/Scan) in both BAT and Cartesian coordinates. Note that, in HA/Scan, all modes with low force constants are scanned, but the harmonic approximation is replaced by a numerical integration only for modes that show significant anharmonicity, as defined above.

TABLE 1: Calculated Standard Chemical Potentials (kcal/mol; $C^\circ = 1$ mol/L) of United-Atom Linear Alkanes in All-Trans Conformations and without Nonbonded Interactions[a]

		bond-angle-torsion		anchored Cartesian		rigid rotor
	analytic	HB	HA	HB	HA	HA
propane	−11.22	−11.22	−11.22	−11.22	−11.22	−11.22
butane	−11.19	−11.19	−11.16	−11.12	−11.16	−11.16
pentane	−11.16	−11.16	−11.10	−11.02	−11.10	−11.10
hexane	−11.13	−11.13	−11.04	−10.84	−11.04	−11.04
heptane	−11.10	−11.10	−10.99	−10.66	−10.99	−10.99
octane	−11.07	−11.07	−10.93	−10.46	−10.93	−10.93

[a] See the first paragraph in the Results for abbreviations. Note that the harmonic approximation in the rigid-rotor oscillator approximation is in Cartesian coordinates. Each HB calculation used 500 000 samples and was executed 10 times with different random number seeds; the arithmetic means are reported. The potential energy at the base of these all-trans energy wells is zero.

3.1. Simplified Linear Alkanes: Comparison with Analytic Results. Simplified linear alkanes form a useful test set because their configuration integrals can be computed accurately by analytic or semianalytic methods. The alkanes are simplified by merging hydrogens into carbons (united-atom representation) and by artificially eliminating all nonbonded interactions and hence all coupling among the bond-stretch, bond-angle, and torsional degrees of freedom. Thus, the configuration integral is factorized and the integrals over bond lengths, bond angles, and torsions can be evaluated separately and very accurately as 1-dimensional analytic or numerical integrals. (Tests show, not surprisingly, that the configuration integrals change negligibly when the Jacobian determinants (eq 5) are approximated as constant and moved outside the integrand.) Table 1 summarizes the results. The numbers include the momentum integrals[14] to facilitate comparison with the rigid rotor harmonic oscillator approximation (RRHO), from which the momentum integrals cannot be separated. The integration limits are as described in the Methods.

The harmonic approximation gives virtually identical results when implemented in BAT, anchored Cartesian, and rigid-rotor (Ekhart−Sayvetz) coordinates: see the "HA" columns in Table 1. This agreement validates the formulas used, such as the Jacobian determinants and the transformation matrixes. The harmonic approximation also agrees quite well with the analytic results (first column of the table), though modest deviations are present and increase with the length of the alkane. Mode Scanning shows that all the modes are highly harmonic, so HA/Scan uses the pure harmonic approximation and the HA/Scan results are identical to the HA results. The HA/Scan results therefore are not listed in the table.

The results of harmonic-biased sampling in BAT coordinates (numerical column 2 in the table) are highly accurate, agreeing with the analytic results to four significant digits. However, MINTA is less accurate, as shown in the fourth column, and the error rises with the length of the alkane, as graphed in Figure 5. MINTA is actually less accurate than the harmonic approximation in these tests. Its errors presumably result from the fact that no linear combination of Cartesian coordinates can reproduce a bond rotation accurately for more than a few degrees of rotation (Figure 1). BAT coordinates do not have this problem and so should do a better job of focusing sampling in low-energy conformations. If this is correct, then integrals over BAT coordinates should also converge with fewer samples than integrals over Cartesian coordinates, and Figures 6 and 7 document that this is indeed the case; note the difference in the scale of the ordinates in the two graphs. (The large downward excursions in the Cartesian results occur when a rare low-energy conformation is found.) It is also of interest to document that sampling without any harmonic bias at all converges slowly, as shown for BAT coordinates in Figure 8.

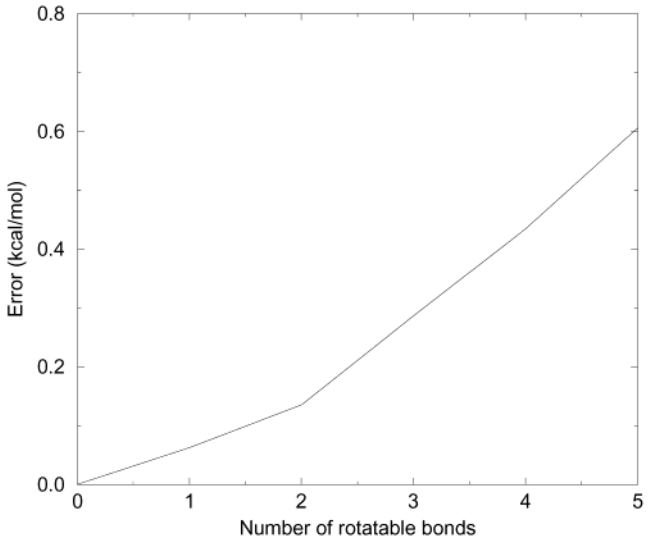

Figure 5. Errors in chemical potentials for linear alkanes from harmonic-biased sampling in Cartesian coordinates (MINTA), as a function of the number of rotatable bonds.

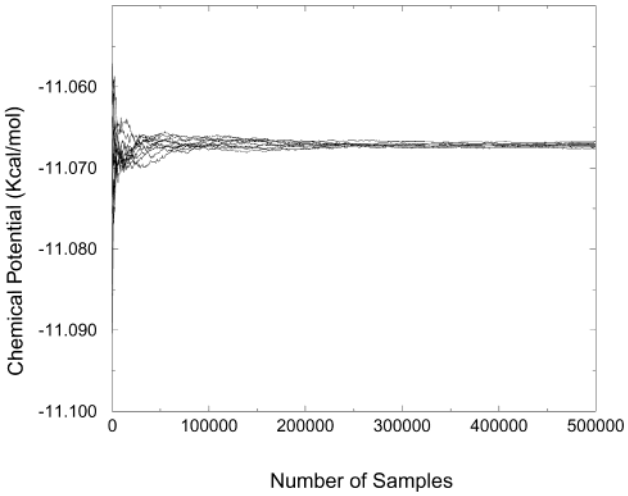

Figure 6. Convergence of chemical potential evaluated by harmonic-biased sampling (HB) in BAT coordinates, for simplified octane. Each graph shows one of 10 independent runs with 500 000 samples each.

3.2. Cyclic Alkanes. If freely rotatable bonds are the chief source of error for MINTA, then the method should be more accurate for cycloalkanes. To test this hypothesis, calculations are reported here for cyclopropane to cyclodecane, again with the united-atom model, but now including nonbonded interactions because no analytic approximation is available in any case. Table 2 shows that, as expected, harmonic biased (HB) sampling in BAT and anchored Cartesian coordinates agree quite well with each other, suggesting that both are accurate.

The harmonic approximations also are in good agreement with the sampling methods for all molecules except cyclopentane, for which the harmonic approximations yield a much lower estimate of the configuration integral. In the HA/Scan method, the mode scans show that only the lowest mode of cyclopentane is markedly anharmonic—flat-bottomed on very local scale, but actually quite narrow as shown in Figure 9. (A similar anharmonicity was reported previously for cycloheptane[12] but

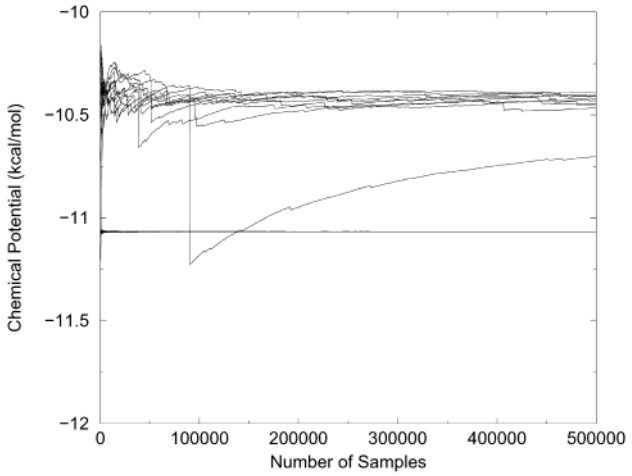

Figure 7. Convergence of chemical potential evaluated by harmonic-biased sampling (HB) in anchored Cartesian coordinates (top) and BAT coordinates (middle, chemical potential ∼−11.1 kcal/mol), for simplified octane. The data for BAT coordinates are the same as in Figure 6, but the scale of the ordinate is different. Each graph shows one of 10 independent runs with 500 000 samples each.

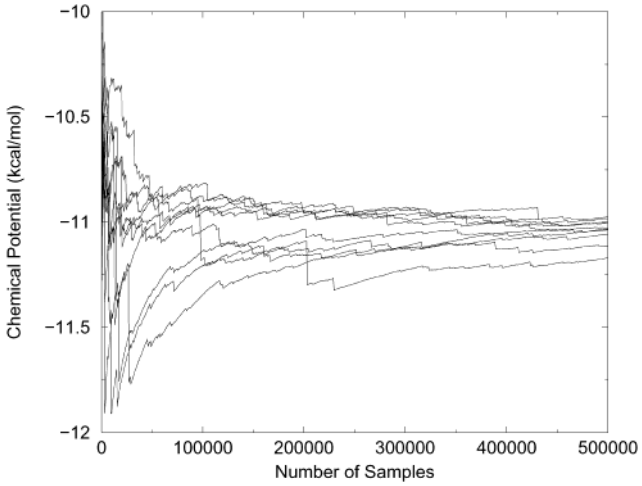

Figure 8. Convergence of chemical potential by nonbiased sampling in BAT coordinates for simplified octane. See Figure 6 for further details.

TABLE 2: Calculated Standard Chemical Potentials (kcal/mol; C° = 1 mol/L) of United-Atom Cyclic Alkanes in a Single Energy Well[a]

	bond-angle-torsion		anchored Cartesian		rigid rotor	
	HB	HA	HB	HA	HA	PE
cyclopropane	−10.26	−10.26	−10.26	−10.26	−10.26	125.94
cyclobutane	−9.70	−9.76	−9.69	−9.76	−9.76	43.64
cyclopentane	−9.20	−10.36	−9.18	−10.92	−10.92	13.87
cyclohexane	−7.51	−7.51	−7.52	−7.51	−7.51	1.15
cycloheptane	−7.28	−7.40	−7.35	−7.40	−7.40	5.79
cyclooctane	−6.88	−7.43	−7.18	−7.67	−7.67	13.18
cyclononane	−5.77	−5.86	−5.85	−5.85	−5.82	18.63
cyclodecane	−5.43	−5.66	−5.66	−5.66	−5.49	13.52

[a] The potential energy at the base of the well (PE), which is excluded from the free energies to facilitate comparison, is listed in the final column. See Table 1 for further details.

is not observed with the force field used here.) In HA/Scan, this anharmonicity is corrected by numerically integrating this solitary anharmonic mode and combining the result with the harmonic approximation for the remaining modes, yielding corrected free energies of −8.96 and −8.85 kcal/mol, respectively, for BAT and anchored Cartesian coordinates, and

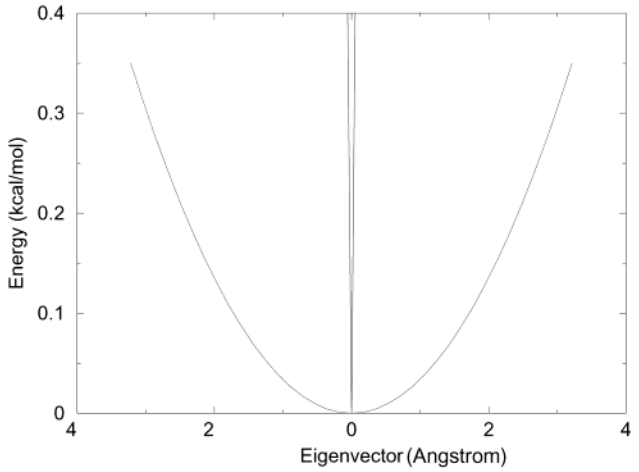

Figure 9. Anharmonicity in cyclopentane: true potential energy shape along the first mode (narrow) vs the harmonic approximation (wide).

TABLE 3: Standard Chemical Potentials (kcal/mol; C° = 1 mol/L) of the Urea Receptor and the Truncated Urea Receptor, Calculated with Various Methods[a]

	bond-angle-torsion			anchored Cartesian			
	HB	HA	HA/Scan	HB	HA	HA/Scan	PE
urea rcptr (**1**)	352.17	352.01	352.01	356.05	351.61	353.91	33.17
trunc rcptr (**2**)	274.08	274.19	274.19	277.42	272.43	278.14	35.39

[a] See the first paragraph of the Results for abbreviations. The potential energy at the base of the well, PE, is excluded from these data. Calculations were done with no solvation model except a dielectric constant of 40. Each Monte Carlo calculation in the sampling methods uses 10^6 samples.

reducing the error relative to the sampling methods to less than 0.35 kcal/mol. The computational cost of this HA/Scan method is far less than that of detailed sampling. For BAT coordinates, the CPU times are as follows: HA/Scan 0.14 s, HA 0.05 s, and HB 201 s. Thus, the HA/Scan method in this case efficiently corrects for the deviation from harmonicity, suggesting that time-consuming sampling may not be needed in general.

3.3. Urea Receptors. Table 3 presents tests on reasonably complex molecules, the "truncated" and full urea receptors, 1 and 2 in Figure 4, which are based upon compounds that have been studied experimentally.[26] No analytic results are available but, on the basis of the alkane results, harmonic-biased sampling in BAT coordinates is used as the accurate reference results to which the other methods are compared.

Interestingly, the pure harmonic approximations in BAT and Cartesian coordinates do not agree with each other, especially for the truncated form of the receptor, where the harmonic approximation in Cartesian coordinates yields a free energy 1.76 kcal/mol lower than that in BAT coordinates. This difference results from the imposition of the additional ±60° integration limit in BAT coordinates (see section 2.2.1). Here, the $\pm 3\sigma$ range corresponds to torsional ranges of up to ±480°, an example of "spinning" as discussed in section 2.2.1. Integrating over $\pm 3\sigma$ range yields integrals that agree with the Cartesian results to high precision, but that clearly are in error because they reflect a torsional rotation of greater than 360°. Interestingly, distorting along the low modes in Cartesian coordinates does not lead to spinning because the bond rotations are coupled with marked bond stretching, as noted above; as a consequence, there is no way to impose the ±60° limit when the harmonic approximation is used in Cartesian coordinates. Note that the

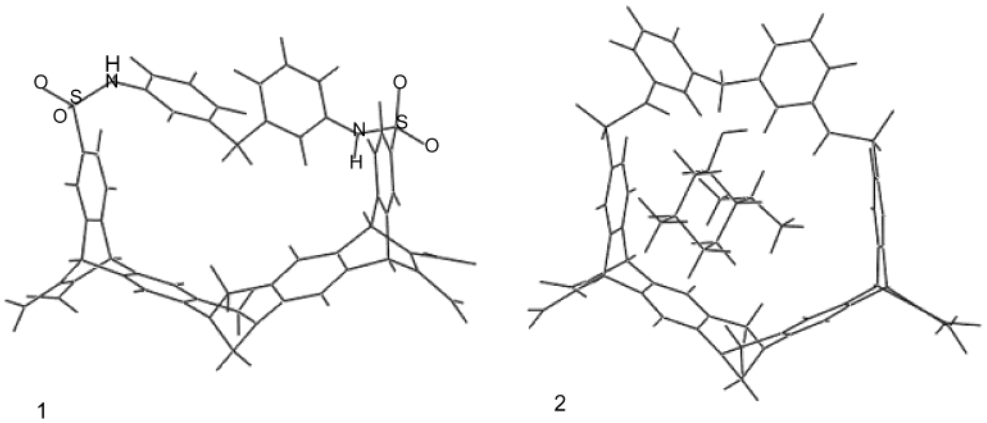

Figure 10. Free cyclophane (1) and the complex of menthol with cyclophane (2).

harmonic-biased sampling calculations in BAT coordinates also use the additional ±60° limits.

Mode Scanning reveals significant anharmonicity in the five lowest modes of the truncated urea receptor in anchored Cartesians, but not in BAT coordinates. However, HA/Scan, which uses the Mode Scanning correction, in Cartesian coordinates leads to excessively high chemical potentials, presumably because Cartesian distortions produce stretched bonds and therefore make energy wells look narrower than they really are. For the same reason, MINTA (HB in Cartesian coordinates) deviates substantially from the reference BAT results. In fact, MINTA is much less accurate than the pure harmonic approximation, even though it requires much more computer time. Thus, the same sort of error observed for the simplified linear alkanes (section 3.1) occurs in this more interesting molecule.

In summary, the results for the urea receptors and the linear alkanes indicate that Cartesian coordinates yield consistently less accurate integrals than BAT coordinates, whether they are used for harmonic-biased sampling (MINTA; HB) or for Mode Scanning in HB/Scan. Fortunately, using BAT coordinates does not impose excessive computational demands. For example, the CPU time needed for the harmonic approximation in BAT coordinates is about 1.9 and 4.8 s for the truncated and full urea receptors, compared with 0.9 and 2.3 s, respectively, in Cartesian coordinates. In general, the use of Cartesian coordinates for calculating molecular configuration integrals can lead to substantial errors, whereas BAT coordinates generate more accurate results.

3.4. Cyclophane. This section describes computational tests for the synthetic cyclophane shown in Figure 10, which binds neutral alicyclic compounds.[27] Table 4 presents the standard chemical potentials of 10 distinct conformations of the free receptor and of its complex with (−)-menthol (Figure 11), computed with the HB, HA, and HB/Scan methods, all in BAT coordinates. The table also shows the energy at the base of each conformational energy minimum; the reported free energies are not scaled by these energies, so the free energies indicate only the shape of the energy wells, not their depths.

Interestingly, the chemical potentials of the free receptor are rather uniform across conformations, despite significant differences in the depths of the energy wells. For the complex, the chemical potentials reported in Table 4, which exclude the potential energies, tend to rise as the potential energies fall, indicating that the deeper wells are narrow − an example of entropy-enthalpy compensation. The pure harmonic approximation, HA, agrees remarkably well with the reference results, harmonic-biased Monte Carlo sampling (HB) for both the free receptor and the complex. When the HA/Scan is used, Mode Scanning shows that most of the soft modes (force constant <10 (kcal/mol)/Å^2) are quite harmonic; only one mode of conformation 2 of the free receptor is sufficiently anharmonic that its numerical integrals deviate from the harmonic approximation by >1 kcal/mol. Substituting numerical integrals for the harmonic approximations for this conformation improves the accuracy relative to the reference results.

TABLE 4: Calculated Standard Chemical Potentials (kcal/mol; C° = 1 mol/L) of Ten Conformations of Cyclophane and the Menthol−Cyclophane Complex in BAT Coordinates, Ordered by Potential Energy at the Base of the Energy Wells[a]

conformation	HB	HA	HA/Scan	PE
	Cyclophane Alone			
1	386.36	385.93	385.93	149.86
2	**385.80**	**384.77**	**385.84**	149.88
3	386.23	386.00	386.00	150.18
4	385.94	385.40	385.40	150.41
5	386.37	386.00	386.00	150.79
6	386.09	386.14	386.14	151.74
7	386.61	386.04	386.04	151.82
8	386.60	386.25	386.25	151.93
9	386.26	385.57	385.57	152.12
10	386.37	385.81	385.81	154.11
	Cyclophane−Menthol Complex			
1	491.13	490.66	490.66	142.25
2	491.41	491.13	491.13	142.71
3	489.91	490.24	490.24	142.87
4	490.70	490.07	490.07	143.62
5	489.36	488.12	488.12	144.94
6	488.03	488.23	488.23	145.11
7	488.18	487.82	487.82	146.59
8	487.72	487.82	487.82	146.59
9	489.60	489.65	489.65	147.40
10	487.67	487.77	487.77	147.83

[a] See the first paragraph in the Results for abbreviations. The potential energies, which are excluded from the free energies to facilitate comparison, are listed in the final column. Calculations were done with a dielectric constant $\epsilon_{ij} = 4r_{ij}$. Boldface text indicates the one conformation that possessed a significantly anharmonic mode and thus for which the HA/Scan result differs from HA.

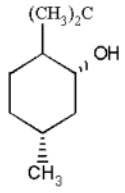

Figure 11. (−)-Menthol.

In real applications, one would usually be interested in the overall chemical potentials of the free and complexed species. These can be computed as

$$\mu^\circ = -RT \log \sum_j^{n_i} e^{-\mu^\circ_j/RT} \quad (9)$$

where μ°_j is the chemical potential of energy well j, now accounting for its potential energy. Applying this formula yields, for the free receptor, standard chemical potentials of 535.22 kcal/mol from harmonic-biased sampling (HB), 534.44 kcal/mol from the harmonic approximation (HA), and 534.94 kcal/mol from the harmonic approximation with Mode Scanning (HA/Scan). For the complex, no conformations possess significantly anharmonic modes, and the overall chemical potentials are 632.23 and 632.11 kcal/mol from HB and HA. Thus, the harmonic approximation is remarkably accurate, and the HA/Scan gives only somewhat better accuracy. The CPU timings for the free receptor are 7320, 5.3, and 66 s, respectively for HB, HA, and HA/Scan.

4. Discussion and Conclusions

This study has evaluated the accuracy of several numerical methods for computing the molecular configuration integral in a single energy well: MINTA (harmonic-biased Monte Carlo integration in Cartesian coordinates); harmonic-biased sampling in bond-angle-torsion (BAT) coordinates; the pure harmonic approximation with finite integration ranges; and the harmonic approximation modified by substitution of numerical integrals only for markedly anharmonic modes (Mode Scanning). Several important conclusions may be drawn from the results.

First, the use of Cartesian coordinates leads to significant errors for systems with freely rotatable bonds. In fact, MINTA is sometimes markedly *less* accurate than the pure harmonic approximation; it is also much slower. The inaccuracy results from the fact that distortions along normal modes built of Cartesian coordinates are unable to generate pure bond-rotations: they also stretch bonds and alter bond angles. As a consequence, the harmonic approximation in Cartesian coordinates leads the sampling algorithm away from the low-energy conformations near the base of the energy well. This problem does not arise with bond-angle-torsion coordinates, so integrals guided by the harmonic approximation in BAT coordinates are much more accurate.

Second, the pure harmonic approximation performs remarkably well in most cases, so long as the integration range is appropriately limited. However, occasional modes that are strongly anharmonic can generate significant errors, as in the case of cyclopentane in this paper (Table 2). Such anharmonicities were a primary motivation for the development of MINTA,[12] but the present study shows that these cases can be corrected more efficiently and accurately by Mode Scanning in BAT coordinates; i.e., by numerically integrating individual anharmonic modes. When Cartesian coordinates are used for molecules with freely rotatable bonds, Mode Scanning fails to increase the accuracy of the harmonic approximation. Again, this is because scanning along Cartesian modes leads to high-energy conformations due to stretching of bonds.

Third, approximating configuration integrals via the harmonic approximation with an integration range of $[-\infty, \infty]$ can lead to significant errors due to double-counting of regions of conformational space. This problem can occur even when the integral is limited to three standard deviations of the Gaussian integrand from the energy minimum, but it is readily addressed in BAT coordinates by further restricting the integration range to prevent rotation of any bond beyond $\pm 60°$. The problem is harder to correct in Cartesian coordinates, which gives the same unbounded harmonic integral but does not map naturally to bond rotations.

In summary, integration methods based upon harmonic modes expressed in Cartesian coordinates are not suitable for compounds with freely rotatable bonds, whereas methods based upon harmonic modes in BAT coordinates are accurate and need not be inordinately time-consuming. The most accurate method tested here is harmonic-biased sampling in BAT coordinates, but in most applications the gain in accuracy is unlikely to be worth its high computational cost, given the success of two faster approaches. Thus, the present study establishes the accuracy and speed of the harmonic approximation with appropriate integration limits in BAT coordinates and further introduces Mode Scanning to identify and correct the occasional mode that is markedly anharmonic. More accurate but slower harmonic-biased sampling in BAT coordinates, may be most useful as a meaning of spot-checking the accuracy of the much faster HA/Scan calculations. Subsequent publications will describe how these methods can be use in practical applications.

Acknowledgment. This work was supported by a grant from NIH (GM61300).

References and Notes

(1) Gilson, M. K.; Given, J. A.; Head, M. S. *Chem. Biol.* **1997**, *4*, 87−92.
(2) Gilson, M. K.; Honig, B. *Protein Struct. Funct. Genet.* **1988**, *4*, 7−18.
(3) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101−1109.
(4) Still, W. C.; Tempczyk, S.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(5) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(6) Lipkowitz, K. G.; Zegarra, R. *J. Comput. Chem.* **1989**, *10*, 595−602.
(7) Wang, J.; Purisima, E. O. *J. Am. Chem. Soc.* **1996**, *118*, 995−1001.
(8) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem. A* **1997**, *101*, 1609−1618.
(9) David, L.; Luo, R.; Head, M. S.; Gilson, M. K. *J. Phys. Chem. B* **1999**, *103*, 1031−1044.
(10) Luo, R.; Gilson, M. K. *J. Am. Chem. Soc.* **2000**, *122*, 2934−2937.
(11) Mardis, K. L.; Luo, R.; Gilson, M. K. *J. Mol. Biol.* **2001**, *309*, 507−517.
(12) Kolossvary, I. *J. Phys. Chem. A* **1997**, *101*, 9900−9905.
(13) Kolossvary, I. *J. Am. Chem. Soc.* **1997**, *119*, 10233−10234.
(14) Potter, M. J.; Gilson, M. K. *J. Phys. Chem. A* **2002**, *126*, 563−566.
(15) McMillan, W. G.; Mayer, J. E. *J. Chem. Phys.* **1945**, *13*, 276−305.
(16) Hill, T. L. *Cooperativity Theory in Biochemistry*, 1st ed.; Springer Series in Molecular Biology; Springer-Verlag: New York, 1985; p 8ff.
(17) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(18) Eckart, C. *Phys. Rev.* **1935**, *47*, 552−558.
(19) Wilson, Jr., E. B.; Howard, J. B. *J. Chem. Phys.* **1936**, *4*, 260−268.
(20) Sayvetz, A. *J. Chem. Phys.* **1939**, *7*, 383−389.
(21) Wilson, Jr., E. B.; Decius, J. C.; Cross, P. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; McGraw-Hill Book Co.: New York, 1955.
(22) Pitzer, K. S. *J. Chem. Phys.* **1946**, *14*, 239−243.
(23) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652−1661.
(24) Go, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535−542.
(25) Molecular Simulations Inc., Walthan, MA.
(26) Bell, T. W.; Hou, Z. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1536−1538.
(27) Glagovich, N. M.; Webb, T. H.; Suh, H.; Geib, S.; Wilcox, C. S. *Proc. Indian Acad. Sci.* **1994**, *106*, 955−970.