Evaluation of the Molecular Configuration Integral in All Degrees of Freedom for the Direct Calculation of Conformational Free Energies: Prediction of the Anomeric Free Energy of Monosaccharides

István Kolossváry

Department of Chemistry, Columbia University, New York, New York 10027, and Department of Chemical Information Technology, Technical University of Budapest, Szt. Gellért tér 4, 1111 Budapest, Hungary

Received: August 27, 1997[⊗]

A novel theoretical approach is presented for the direct calculation of conformational free energies without the need for expensive free energy simulations and the use of computational alchemy. It is shown that conformational search combined with advanced Monte Carlo integration can tackle the daunting problem of solving the molecular configuration integral in all degrees of freedom, affording a direct method to calculate accurate conformational free energies. The new algorithm termed mode integration (MINTA) was applied here to explain anharmonic effects in cycloheptane and to predict the *A* value of cyclohexane derivatives and the anomeric free energy of carbohydrates (tetrahydropyran derivatives and pyranose monosaccharides) using a continuum solvation model.

Introduction

Chemical stability can always be formulated in terms of conformational free energy (CFE) differences. There can be various levels envisioned at which approximations to CFE differences can be made. For example, one wishes to calculate the anomeric free energy for the equilibrium of α and β anomers of monosaccharides. The simplest approach one can follow is to calculate the energy difference between the lowest energy α and the lowest energy β anomer. Of course, this approach ignores entropic effects due to the fact that first of all there are multiple conformations of both the α and β anomers and, second of all, the individual conformations are not static (confined to the bottom of their energy well) but exhibit large dynamic diversity in terms of conformational changes limited to that energy well. Note that glucose, for example, possesses literally hundreds of low-energy conformations of both anomeric states.

The next level of approximation to the CFE is the inclusion of multiple conformations. With this model, two families of different conformational isomers, configurational isomers or stereoisomers, both represented by multiple conformations, are considered as two sets of discrete energy levels corresponding to the energies of the individual conformations. A simple statistical mechanics calculation can then be used to estimate the free energy difference between the two families. Note that such families, henceforth termed somewhat confusingly conformational families, can represent either multiple conformations related to each other by some geometrical criterion, such as holding axial or equatorial substituents, or multiple conformations of stereoisomers. The terms conformational free energy and conformational free energy difference are used generally throughout. However, for example, the *A* values of cyclohexane derivatives (vide infra) represent actual conformational free energy differences, but the anomeric free energy of monosaccharides refers to the free energy difference between α and β diastereomers, i.e., α, β anomers.

The ultimate approach for calculating the free energy difference ΔG between two conformational families, in the classical sense, involves the evaluation of the molecular configuration integral Q

$$Q_1 = \sum_{i=1}^{n_1} \int_{V_i^1} e^{-(E(\mathbf{r})-E_0)/RT}\, d\mathbf{r}, \quad Q_2 = \sum_{i=1}^{n_2} \int_{V_i^2} e^{-(E(\mathbf{r})-E_0)/RT}\, d\mathbf{r} \tag{1}$$

$$\Delta G_{12} = -RT \ln \frac{Q_1}{Q_2} = -RT \ln K_{12} \tag{2}$$

where the indices 1 and 2 refer to two conformational families. It is assumed throughout that the dominant part of the configuration integral comes from contributions at or near to low-energy conformations. Therefore, Q is summed over, respectively, n_1 and n_2 conformations, each encompassing different V volumes of the conformational space. $E(\mathbf{r})$ is the molecular mechanics energy with respect to the nuclear coordinates $\mathbf{r}$. E_0 is the global minimum energy which is the common reference for both conformational families. R is the gas constant, and T is the absolute temperature. K_{12} in the second term of eq 2 is the population ratio of the conformational families at equilibrium. Note that all of the symmetry-related copies of a single conformation including conformational enantiomers should be included in the sum in eq 1 to account for the statistical correction for conformational symmetry.

Direct evaluation of the configuration integral has been considered to be impossible to solve except for problems of very low dimensionality. Instead, indirect methods utilizing various simulation techniques based on free energy perturbation (FEP)[1] and, recently, a novel smart Monte Carlo approach termed jump-between-wells (JBW)[2] have been used extensively to calculate conformational free energy differences. The JBW method coupled with molecular dynamics involves directly monitoring the populations of various conformations of the two conformational families (rhs of eq 2) in a simulation in which conformational interconversions occur frequently, producing converged, Boltzmann-weighted ensembles of conformational states.

Very recently, direct methods have also emerged that involve the direct evaluation of the configuration integral as sums over conformational minima (eq 1). Most notably, a new method

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1997.

S1089-5639(97)02794-1 CCC: $14.00

termed "mining minima" has been introduced in which the configuration integral is evaluated over the "soft modes" identified as torsion angles.[3] It should be stressed, however, that the exclusion of "hard modes" such as bond lengths and bond angles is, in general, a poor approximation. Cyclic structures, for example, undergo sufficient variation of their ring bond angles and even bond lengths during conformational interconversions to contribute a significant amount to the conformational free energy. Therefore, a method is sought that can evaluate the configuration integral in all degrees of freedom in order to calculate accurate free energies.

The "mining minima" method could, in principle, include the hard modes, but in practice, it is limited to only a few degrees of freedom because it is based on simple Monte Carlo integration. The evaluation of multidimensional integrals is a daunting task, and for dimensions greater than three, Monte Carlo integration is the only method available. However, simple Monte Carlo integration requires a vast number of function evaluations to afford reasonably accurate integrals when more than about half a dozen dimensions are considered.[4] In this paper, a novel Monte Carlo integration technique is presented, termed mode integration (MINTA) that operates in all degrees of freedom. The new method allows, for the first time, for the direct calculation of conformational free energies of diverse systems, and gives results that are virtually identical with those obtained by converged JBW simulations. The MINTA method is applied here to predict the *A* value of cyclohexane derivatives, and the anomeric free energy of carbohydrates (tetrahydropyran derivatives and pyranose monosaccharides) using a continuum solvation model.

Theory

Conformational Search. Mode integration relies upon an exhaustive conformational search of the low-energy minima. Our method of choice is a particularly efficient conformational search procedure termed low-mode conformational search (LMOD).[5]

Monte Carlo Integration. Mode integration is based on a novel use of the harmonic approximation to achieve remarkable accuracy in the numerical integration of high-dimensional molecular configuration integrals. Importance sampling Monte Carlo integration utilizes a sampling function to preferentially sample the integrand in those regions where the integrand has a significant magnitude. The basic theorem of importance sampling Monte Carlo integration asserts that the optimal choice of the sampling function p for the numerical integration of a function f (in any dimensions) is $p \propto |f|$, i.e., the more one knows about f the better one can estimate $\int f$.[4c] The basic tenet of mode integration is to construct a sampling function that is very similar to the partition function (eq 1) in the vicinity of low-energy minima on the PES.

The obvious approach is to apply the harmonic approximation to each low-energy well on the PES by a local, second-order Taylor expansion of the potential energy function.[6] The resulting sampling functions are normalized multivariate Gaussians

$$p_i = \sqrt{\frac{\det \mathbf{H}_i}{(2\pi RT)^n}} \exp\left(-\frac{1}{2RT}(\mathbf{r}-\mathbf{r}_i)\mathbf{H}_i(\mathbf{r}-\mathbf{r}_i)\right) \quad (3)$$

where $\mathbf{r}_i$ and $\mathbf{H}_i$ denote, respectively, the bottom of a particular energy well and the associated local Hessian matrix. The Hessian is evaluated at the bottom of the well. The number of degrees of freedom n is equal to the number of unconstrained internal coordinates. Note that p_i also depends on the temperature via RT.

Mode integration means, in essence, that eq 3 is utilized as a sampling function to integrate eq 1. However, for the reduction of the proposed theory in practice, one needs to consider two important issues: definition of the volume that is associated with a particular conformation in conformational space and to find an optimal coordinate system that guarantees the most accurate integral estimates. The optimal coordinate system can be derived by considering a simple fact about multivariate Gaussians. In real Gaussians, **H** is essentially the inverse of the covariance matrix which accounts for the correlation between the random variables of the multivariate Gaussian distribution. Unless the covariance matrix is diagonal, the off-diagonal elements give rise to mixed terms in the Gaussian exponential. However, by suitable coordinate transformation termed principal component transformation, which is based on the diagonalization of the covariance matrix, the mixed terms can always be eliminated. The resulting Gaussian can be expressed as a product of n independent, univariate Gaussian functions. Analogously, a similar transformation can be applied to the Hessian matrix. The eigenvectors of the Hessian matrix $\mathbf{H}_i$ span a local, orthogonal coordinate system centered at $\mathbf{r}_i$, which is closely related to the normal modes of vibration. Utilizing the eigenvectors of the Hessian matrix in eq 3 results in a sampling function that can be expressed as a product of n independent, univariate Gaussian functions whose single variable represents the corresponding normal mode. Importance sampling based on the proposed, separable sampling function involves consecutively sampling n independent, one-dimensional Gaussian functions.

The normal mode coordinate system also provides a simple definition of the integration volume. The tail-off of one-dimensional Gaussians can be measured conveniently in units of standard deviation (σ). Approximately 68% of the underlying population is covered by $\pm 1\sigma$, and 99.7% by $\pm 3\sigma$. In terms of a Gaussian partition function, this means that beyond 3−4 standard deviations the Boltzmann probability vanishes. Thus, the proposed multidimensional integration volume is a n-dimensional hypercube that is aligned along the normal modes and whose dimensions depend on the harmonic vibrational frequencies as well as the temperature. Note that the integration volume is different for different conformations. MINTA therefore, utilizes the following separable sampling function for the Monte Carlo integration of the molecular configuration integral

$$p_i = \prod_j^n \sqrt{\frac{\lambda_j}{2\pi RT}} \exp\left(-\frac{\lambda_j x_j^2}{2RT}\right) \quad (4)$$

where $j = 1...n$ normal modes, $\lambda_1...\lambda_n$ are the eigenvalues of the non-mass-weighted Hessian matrix, and $x_1...x_n$ are the corresponding normal coordinates represented by the eigenvectors of the Hessian matrix. The center of the normal coordinate system is located at the bottom of the energy well where the gradient of the potential energy vanishes. $\sqrt{(RT/\lambda_j)}$ represents the quasi standard deviation of the Gaussian. Note that eq 4 refers to a single conformation i. Mode integration of the full molecular configuration integral involves integrating the individual, conformational partition functions via importance sampling using eq 4 and summing them up (eq 1)

$$Q = \sum_i^{n_{\text{conf}}} s_i \int_{\square_i} \left(\frac{e^{-(E(\mathbf{x}_i)-E_{0i})/RT}}{p_i}\right) p_i \, d\mathbf{x}_i \approx \sum_i^{n_{\text{conf}}} s_i \left\langle \frac{e^{-(E(\mathbf{x}_i)-E_{0i})/RT}}{p_i} \right\rangle_{\square_i} \quad (5)$$

where p_i is the Gaussian sampling function (eq 4), $\mathbf{x}_i$ is the vector of normal coordinates, and s_i is the symmetry number, all for conformation i. The integration volume $\square_i$ is a n-dimensional hypercube with dimensions $d_j = 2$ const $\sqrt{(RT/\lambda_j)}$ measured in units of the quasi standard deviation (eq 4) where $j = 1...n$ and const is a multiplication factor. Note that in eq 5 E_{0i}, the reference for the energy difference in the Boltzmann exponential, is the bottom of the individual energy wells not the global minimum E_0 (eq 1). Of course, this is only a matter of scaling. The right-hand side of eq 5 is utilized to estimate the configuration integral by importance sampling inside the multidimensional hypercubes.

The essence of mode integration can be captured in a single sentence: The underlying idea that makes the method work is a novel use of the harmonic approximation for the particularly effective importance sampling of the true potential energy surface in the framework of multidimensional Monte Carlo integration.

Computational Details

The low-mode conformational search procedure and the MINTA method have been integrated into the MacroModel/Batchmin computational chemistry program package.[7] The calculations have been carried out using version 5.5 of Batchmin running on Silicon Graphics workstations.

Integration Volume. Preliminary calculations were aimed at defining the size of the hypercubes in eq 5. Note that due to possibly overlapping hypercubes, double sampling can, in principle, give rise to biased integral estimates. Randomly selected conformations of several cycloalkanes and monosaccharides were subjected to test calculations with variable size hypercubes ranging from 1 to 5 standard deviations. The results clearly showed that the intuitive choice of 3σ was generally applicable to give converged free energies. Furthermore, it was shown that the overlap issue was negligible except for one notable case, the twist-boat (TB) conformation of cycloheptane which is discussed further below. For all other cases, a simple test showed that the 3σ hypercubes did not extend beyond the catchment region of their associated conformation. In this test calculation, thousands of structures generated by the Monte Carlo sampling were all saved and re-minimized to see whether sampling had escaped from the minimum well. Using 3σ, none of the trial structures crossed a conformational barrier during reminimization. Even using 4σ and 5σ afforded only a very few cases where reminimization of a trial structure carried it over to a different conformational minimum. Therefore, it was concluded that the integration volumes (eq 5) should be defined as 3σ hypercubes throughout.

Results and Discussion

Anharmonic Effects in Cycloheptane. The TB conformation of cycloheptane has proven to be an extremely useful test for MINTA. In the test aimed at determining the integration volume (vide supra), several trial conformations in the 3σ hypercube were interconverted to the lower energy twist-chair (TC) conformation. Senderowitz et al. also investigated cycloalkanes including cycloheptane to compare conformational free energy differences obtained by JBW and the harmonic approximation (rigid-rotor harmonic-oscillator (RRHO) model).[2] It was concluded in their study that JBW could properly account for anharmonic effects, but generally, the harmonic approximation was not so bad except for TB cycloheptane. The free energy difference between TB and TC cycloheptane was found to be 1.66 kcal/mol (RRHO) and 2.8 kcal/mol (JBW), respectively.[2] Such a big difference between the JBW and the RRHO calculation is indicative of significant anharmonic effects. The following study was aimed at mapping the real shape of the TB cycloheptane potential energy surface (PES) to determine what is wrong with the harmonic approximation.

The adaptive implementation of importance sampling Monte Carlo integration provides a computational tool to trace the potential energy surface. The so-called VEGAS algorithm[4b,c] operates by constructing, adaptively, a separable sampling function. The first iteration is simple Monte Carlo integration based on uniform sampling. However, the results of the function evaluations are saved and used to adjust the sampling function in subsequent iterations based on the fact that the optimal sampling function $p \propto |f|$ (vide supra). The sampling function is adjusted in subsequent iterations to preferentially sample regions where the magnitude of the integrand is large. The VEGAS algorithm has been implemented in the mode integration method with an important change. Instead of starting with uniform sampling, eq 4 is applied to initiate VEGAS. Unfortunately, VEGAS can only be used for up to 10 dimensions[4c] and therefore cannot generally be used alone for the evaluation of molecular configuration integrals. However, it can be used to trace the potential energy surface in the directions of low-frequency modes. Since a converged VEGAS sampling function is proportional to the absolute value of the integrand, it is clear that a converged VEGAS iteration initiated from eq 4 affords a sampling function which is proportional to the exact partition function. In particular, the converged pieces of the separable sampling function (eq 4) trace the exact shape of one-dimensional slices of the partition function, aligned by the corresponding normal coordinate directions. For a quadratic potential, the exact shape is a Gaussian. Therefore, the deviation of the exact shape from the corresponding Gaussian reveals the nature of the anharmonic effect.

The results of a VEGAS calculation applied to the first 10 low-frequency modes of TB cycloheptane are shown in Figure 1. The first five modes are aligned from left to right. In the top row, the thick lines delineate the exact shape of the potential surface derived from the VEGAS calculation along the corresponding normal modes and the thin lines show the corresponding harmonic potentials. In the bottom row, the thick lines depict the true one-dimensional partition functions, and the thin lines show the corresponding Gaussians. Note that for better visualization the cumulative partition function is shown instead of the bell-shaped function defined in eq 4. Also note that the whole figure is autoscaled, the actual extent of the PES slices is given below the graphs. One glance at Figure 1 immediately reveals what is wrong with the harmonic approximation. The very low MM2[8] frequency (7.0 cm^{-1}) is unrealistic. In fact, the corresponding 3σ hypercube extends to a spurious ± 25 Å in the direction of this normal mode. Several other force fields and an ab initio calculation also afforded similar, all too low frequencies: MM3,[9] 21.9 cm^{-1}; AMBER,[10] 19.3 cm^{-1}; AMBER94,[11] 29.4 cm^{-1}; OPLS,[12] 12.0 cm^{-1}; MMFF,[13] 25.1 cm^{-1}; and HF/6-31G* 22.3 cm^{-1}.

The real shape of the PES slice depicted by the thick line in Figure 1 describes a much steeper PES. The real shape can be characterized by a fairly narrow, flat-bottom well. Note that it has been already known from free energy simulations that distorting a molecule along a single, extremely flat mode often results in "bumping" in a steep (potential) wall. However, until very recently, this phenomenon has been treated simply as "anharmonic effect" without further specification. Recent quantum mechanical calculations have shown that similar effects in a small protein (BPTI) can be attributed to anharmonicity of the low-frequency vibrational wave functions.[14] In their work

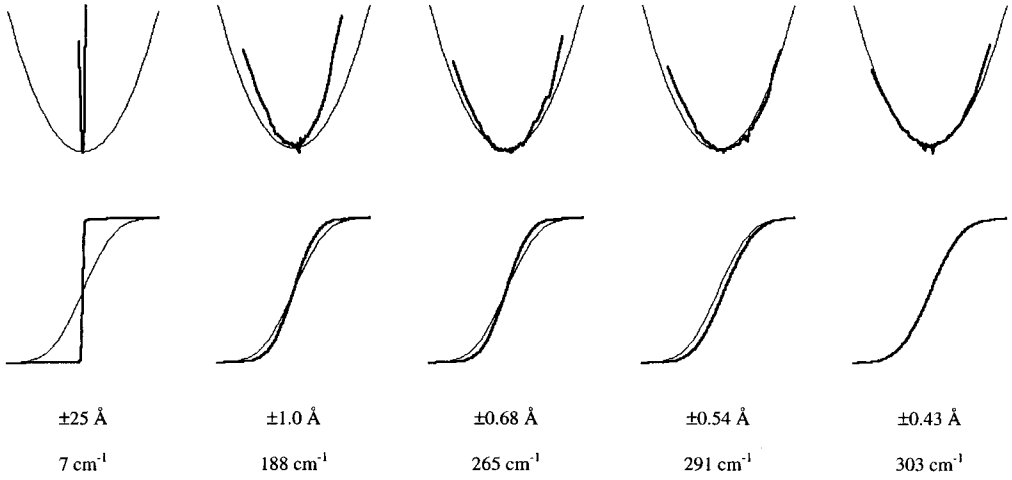

Figure 1. Anharmonic effects in TB cycloheptane: true vs harmonic one-dimensional PES slices. In the top row, the thick lines delineate the exact shape of the potential surface along the first five low-frequency normal modes and the thin lines show the corresponding harmonic potentials. In the bottom row, the thick lines depict the true one-dimensional, cumulative partition functions and the thin lines show the corresponding cumulative Gaussians. The plots are autoscaled; the actual extent of the PES slices are given below the graphs along with the MM2 harmonic frequencies.

the authors presented a PES slice very similar to those shown in Figure 1, by which they demonstrated that inclusion of the cubic and quartic diagonal terms of the Taylor expansion of the potential energy sufficed to trace the true PES along a low-frequency mode almost perfectly.[14b] They also stressed the experimental significance of the problem of "extreme flat modes" in a variety of vibrational spectroscopies.

In summary, the problem with the harmonic approximation in this case is that the very low frequency incorrectly suggests a shallow energy well based on the local curvature of the PES at the flat bottom. JBW gives the correct answer, because it samples the whole well, not only the flat bottom. For the same reason, mode integration should give the same correct answer with a suitable hypercube in eq 5. Therefore, a simple safeguard is used to prevent inadequate sampling due to flat-bottom wells. Generally, the 3σ rule is applied to the hypercube, but for extremely soft modes the corresponding dimension of the hypercube is limited to ± 3 Å. A MINTA calculation based on the $3\sigma/3$ Å rule using 10 000 energy evaluations, sampling all degrees of freedom, afforded 3.0 kcal/mol free energy difference between TB and TC cycloheptane; however, without the 3 Å limit the result was 3.2 kcal/mol.

The true shape of the PES slice associated with mode 4 (Figure 1) exemplifies another anharmonic effect, namely, an asymmetrical energy well. However, it is clear that the higher modes are, in fact, very close to be harmonic, the difference between the true PES slice and the corresponding harmonic potential completely vanishes from the fifth mode on. It has been shown recently that in a similar manner the higher frequency modes of the protein BPTI are nearly harmonic.[14b,15] This very fact suggests an astonishing simplification of MINTA: numerical integration of the soft modes and analytical integration of the hard modes based on the harmonic quantum oscillator model. Note that the "soft" modes are defined in a more general sense than just torsions,[3] representing low-mode conformational motions generally applicable to cyclic and acyclic molecules.[5,16] The combined numerical/analytical MINTA algorithm involves the numerical integration of the soft modes via eqs 4, 5 (with or without VEGAS), and the analytical integration of the hard modes using the harmonic approximation. In all of the following calculations, however, MINTA was used to integrate all degrees of freedom numerically.

***A* Value of Methyl-, Isopropyl-, and Methylphenylcyclohexane.** The so-called *A* value of methylcyclohexane is the free energy difference between axial and equatorial methylcyclohexane. *A* values of numerous cyclohexane derivatives have been determined experimentally including isopropylcyclohexane and 1-methyl-1-phenylcyclohexane.[17] Recent JBW simulations afforded converged *A* values consistent with experiment.[2] Our MINTA calculations afforded virtually identical results summarized in Table 1.

TABLE 1: Calculated and Experimental *A* Values[a] of Methylcyclohexane, Isopropylcyclohexane, and 1-Methyl-1-phenylcyclohexane at 300 K

	ΔG (kcal/mol)[b]		
	MM2, in vacuo		
	JBW[2]	MINTA	expt[2]
methyl	1.9	1.9	1.6−1.8
isopropyl	2.6	2.4	2.0−2.4
methylphenyl	−0.10	0.15	∼0
	0.23[c]	0.34[c]	0.3[c]

[a] *A* value = free energy difference between axial and equatorial substituted cyclohexane. [b] Positive values favor the equatorial methyl, equatorial isopropyl, and axial phenyl substituents, respectively. [c] 200 K.

The MINTA calculations involved 10 000 function evaluations per conformation using the MM2 force field with infinite nonbonded cutoffs, in vacuo, and required only a few minutes of CPU time. It should be noted that FEP methods based on statistical mechanics generally perform well for similar systems in vacuum,[18] however, MINTA is much faster. A recent, successful FEP simulation that involved the interconversion of chair conformers of methylcyclohexane by mutating an axial dummy atom and an equatorial united atom methyl into one another (using the AMBER force field) required 21 stages using stochastic dynamics and spanned a total of 4.2 ns to reach convergence.[2] The united atom (8 atoms) FEP simulation took 50 CPU minutes on a 250 MHz R4400 processor, while the full atom (21 atoms) MINTA calculation described above took only a little over 1 CPU minute (a speed factor of ∼43).

The MINTA integral estimates were calculated as block averages based on 10 independent 1 000 point samples.[4b] The resulting configuration integrals were all subject to less than 5% relative error (calculated as ± 1 standard deviation)[4b] which is equivalent to ± 0.03 kcal/mol in free energy at room temperature. Note that the particularly good results in Table 1 are not only indicative of predictive power, but equally importantly, the remarkable agreement between JBW and

TABLE 2: Calculated and Experimental Anomeric Free Energies[a] of Tetrahydropyran Derivatives and Pyranose Monosaccharides at 300 K

	ΔG (kcal/mol)[b]		
	AMBER*, GB/SA water		
	JBW[20e]	MINTA	expt/ab initio[20e]
2-hydroxytetrahydropyran	0.28[c]	0.40[c]	0.69[c]
	−1.02	−1.02	−0.95
2-methoxytetrahydropyran	0.65[c]	0.68[c]	0.94[c]
	0.36[d]	0.38[d]	0.64
	−0.41	−0.35	−0.7 to 0.1
glucose	−0.22	−0.11	−0.34
methyl glucoside	0.53	0.59	0.42
mannose	0.21	0.21	0.34 to 0.45
methyl mannoside	1.34	1.35	1.70
galactose	−0.03	0.10	−0.37
2-deoxyglucose	−0.45	−0.38	−0.05
N-acetylglucosamine	0.50	0.52	0.51

[a] Anomeric free energy = free energy difference between α and β anomers. [b] Positive values favor the α anomer; negative values favor the β anomer. [c] In vacuo. [d] GB/SA chloroform.

MINTA results is the best cross-validation of these fundamentally different methods.

Anomeric Free Energy Calculations. The successful test on *A* values prompted a series of significantly more difficult calculations of anomeric free energies. The modeling of carbohydrates has long been a particularly difficult problem, because conventional molecular mechanics force fields could not account for stereoelectronic effects resulting from the unique, condensed, and highly polar structure of carbohydrates.[19,20e] The anomeric free energy in particular, representing the free energy difference between the α and β stereoisomers of carbohydrates, can be measured experimentally, but has escaped calculations until very recently.[20] Especially, the introduction of carbohydrate parameters in the AMBER*[7] force field based on ab initio calculations on pyranoses afforded a suitable force field for JBW calculations.[20e] Therefore, a series of MINTA calculations have been carried out using the same AMBER* force field to calculate the anomeric free energy of tetrahydropyran derivatives and pyranose monosaccharides, glucose, methyl glucoside, mannose, methyl mannoside, galactose, 2-deoxyglucose, and *N*-acetylglucosamine. Once again, MINTA predictions were consistent with experiment and virtually identical with JBW predictions. The results are summarized in Table 2.

The MINTA calculations were carried out with the united atom AMBER* force field augmented with the new carbohydrate parameters.[20e] Solvation treatment included the GB/SA[21f] continuum model with constant ($\epsilon = 1$) dielectric Coulomb equation and extended nonbonded cutoffs (20 Å for Coulomb and 8 Å for van der Waals interactions). The GB/SA model belongs to the family of continuum models[21] that have become popular alternatives to explicit (molecular) solvent models[22] for the modeling of molecules in solution using molecular mechanics. Explicit solvent models, while exact in principle, are very expensive computationally and are subject to random noise impeding convergence in FEP simulations. Furthermore, there is no compelling evidence that using contemporary force fields explicit solvent models are more accurate than continuum models.[23] Therefore, the use of a continuum model for the calculation of anomeric free energies is justified. The GB/SA model in particular has been shown to afford free energies of solventlike dielectric polarization for small organic molecules and peptides in water comparable to those calculated by accurate but significantly slower Poisson−Boltzmann methods.[24]

The MINTA integrals were calculated as block averages based on 10 × 1000 independent energy evaluations per conformation. The resulting anomeric free energies in Table 2 are subject to less than ±0.03 kcal/mol error. The MINTA calculations required 25−45 CPU seconds per conformation on a 75 MHz R8000 processor.

Note that the pyranose monosaccharides have literally hundreds of conformations making it virtually impossible to claim a complete conformational search. Therefore, at first glance it seemed reasonable to use the exact same set of the 100 lowest energy conformations for the MINTA calculation that were originally used for the JBW simulation[20e] to ensure unbiased comparison. However, the resulting MINTA anomeric free energies were off by up to 0.5 kcal/mol with respect to the JBW results. A more careful look at the JBW algorithm explains the discrepancy. The basic tenet of JBW is to interconvert conformations found in a preceding search, using an internal coordinate transformation matrix. However, JBW also involves small random variations of the internal coordinates and it is also coupled with stochastic dynamics.[2,20e] This means that JBW, in fact, explores more of the conformational space than just the conformations used to seed the JBW simulation. Of course, this also means that MINTA, which is restricted to the conformations at hand, should utilize all of the low-energy conformations found by the preceding search. Therefore, the MINTA calculations were repeated utilizing the full set of conformations found during the conformational search. Indeed, the results shown in Table 2 are in astonishing agreement with JBW! The largest difference is only 0.13 kcal/mol (galactose), and for half of the molecules, the difference is within the ±0.03 kcal/mol error bar. It should be noted, once again, that such remarkable agreement of two entirely different methods on a difficult problem cross-validates them by any standard.

Conclusions

The calculation of conformational free energies is one of the key issues in computational chemistry. In this paper, a novel theoretical approach is proposed for the direct calculation of conformational free energies without the need for expensive free energy simulations and the use of computational alchemy.[18] The new MINTA algorithm is based on a particularly efficient implementation of importance sampling Monte Carlo integration to solve the molecular configuration integral in all degrees of freedom. The MINTA method was applied here to study anharmonic effects in cycloheptane and to predict the *A* value of cyclohexane derivatives and the anomeric free energy of carbohydrates (tetrahydropyran derivatives and pyranose monosaccharides). The MINTA calculation showed that the harmonic approximation was not applicable to TB cycloheptane because the shape of the true potential energy surface along the lowest frequency normal mode can be characterized by a narrow, flat-bottom well. The MINTA predictions on the *A* values and the anomeric free energies were virtually identical with those obtained by converged JBW free energy simulations. For the calculation of the *A* value of methylcyclohexane, MINTA was at least 43 times faster than FEP. It should also be stressed that although inaccuracies in the computational methods and force fields have much room for improvement and good agreement with experiment can always be attributed, at least to some extent, to cancellation of error, 0.5 kcal/mol or better accuracy is well within the range of serious interest for experimentalists. In recent studies MINTA has been shown to reproduce experimental binding free energies of enantioselective binding of small peptides to synthetic hosts to within the same <0.5 kcal/mol accuracy.[25] In conclusion, we believe that MINTA should find wide utility for direct conformational free energy calculations.

Acknowledgment. I am indebted to H. Senderowitz for graciously making the original raw data available for the anomeric free energy calculations and to W. C. Still, and W. C. Guida for carefully reading the manuscript.

References and Notes

(1) Kollman, P. A. *Chem. Rev. (Washington, D.C.)* **1993**, *93*, 2395−2417.

(2) Senderowitz, H.; Guarnieri, F.; Still, W. C. *J. Am. Chem. Soc.* **1995**, *117*, 8211−8219.

(3) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem. A* **1997**, *101*, 1609−1618.

(4) (a) Stroud, A. H. *Approximate Calculation of Multiple Integrals*; Prentice Hall: Englewood Cliffs, New Jersey, 1971; Chapter 6. (b) Lepage, G. P. *J. Comput. Phys.* **1978**, *27*, 192−203. (c) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C Second Edition*; Cambridge University Press: New York, 1992; pp 316−328.

(5) Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019.

(6) (a) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325−332. (b) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370−1374.

(7) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.

(8) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127−8134.

(9) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.

(10) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta Jr., S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(11) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, G. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(12) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(13) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−519.

(14) (a) Roitberg, A.; Gerber, R. B.; Elber, R.; Ratner, M. A. *Science* **1995**, *268*, 1319−1322. (b) Roitberg, A. E.; Gerber, R. B.; Ratner, M. A. *J. Phys. Chem. B* **1997**, *101*, 1700−1706.

(15) Go, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696.

(16) Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1993**, *115*, 2107−2119.

(17) Eliel, L. E.; Wilen, S. H. *Stereochemistry of Organic Compounds*; Wiley-Interscience: New York, 1994; pp 696−702.

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

(19) (a) Melberg, S.; Rasmussen, K. *J. Mol. Struct.* **1979**, *57*, 215. (b) Melberg, S.; Rasmussen, K. *Carbohyd. Res.* **1979**, *76*, 23. (c) Melberg, S.; Rasmussen, K. *Carbohyd. Res.* **1980**, *78*, 215. (d) Jeffrey, G. A.; Taylor, R. *J. Comput. Chem.* **1980**, *1*, 99. (e) Rasmussen, K. *Acta Chem. Scand.* **1982**, *A36*, 323. (f) Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohyd. Res.* **1988**, *180*, 207. (g) Mardsen, A.; Robson, B.; Thompson, J. S. *J. Chem. Soc., Faraday Trans. 1* **1988**, *84*, 2519. (h) Homans, S. W. *Biochemistry* **1990**, *29*, 9110. (i) Grootenhuis, P. D. J.; Haasnoot, C. A. G. *Mol. Simul.* **1993**, *10*, 75. (j) Glennon, T. M.; Zheng, Y. J.; Le Grand, S. M.; Shutzberg, B. A.; Merz, K. M. *J. Comput. Chem.* **1994**, *15*, 1019. (k) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832. (l) Asensio, J. L.; Jimenez-Barbero, J. *Biopolymers* **1995**, *35*, 55. (m) Reiling, S.; Schlenkrich, M.; Brickmann, J. *J. Comput. Chem.* **1996**, *17*, 450. (n) Csonka, G. I.; Éliás, K.; Csizmadia; I. G. *J. Comput. Chem.* **1997**, *18*, 330−342. (o) Csonka, G. I.; Éliás, K.; Kolossváry, I.; Sosa, C. P.; Csizmadia; I. G. *J. Phys. Chem. A*, in press.

(20) (a) Ha, S.; Gao, J.; Tidor, B.; Brady, J. W.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 1553. (b) Zheng, Y.-J.; Le Grand, S. M.; Merz, K. M.; *J. Comput. Chem*. **1992**, *13*, 772.(c) van Eijck, B. P.; Hooft, R. W. W.; Kroon, J. *J. Phys. Chem.* **1993**, *97*, 12093. (d) Schmidt, R. K.; Karplus, M.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 541.(e) Senderowitz, H.; Parish, C.; Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 2078−2086. (f) Senderowitz, H.; Still, W. C. *J. Phys. Chem. B* **1997**, *101*, 1409−1412.

(21) (a) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199. (b) Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086. (c) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4105, 4109, 4118. (d) Warshel, A.; Russel, S. T. *Rev. Biophys.* **1984**, *17*, 283. (e) Gilson, M.; Honig, B. *Proteins* **1988**, *4*, 7. (f) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129. (g) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305.

(22) (a) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913. (b) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926. (c) Jorgensen, W. L.; Ravimohan, C. J. *J. Chem. Phys.* **1985**, *83*, 3050.

(23) Marten, B.; Kim. K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775−11788.

(24) (a) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014. (b) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J. Phys. Chem. B* **1997**, *101*, 1190−1197.

(25) Kolossváry, I. *J. Am. Chem. Soc.* **1997**, *119*, 10233−10234.